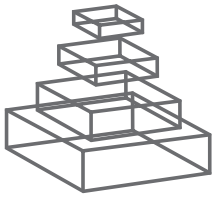
frontiers
RESEARCH TOPICS

STATISTICAL ANALYSIS OF MULTI-CELL RECORDINGS: LINKING POPULATION CODING MODELS TO EXPERIMENTAL DATA

Hosted by
Matthias Bethge, Philipp Berens, Jakob Macke

frontiers in
COMPUTATIONAL NEUROSCIENCE



frontiers

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-012-6
DOI 10.3389/978-2-88919-012-6

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

STATISTICAL ANALYSIS OF MULTI-CELL RECORDINGS: LINKING POPULATION CODING MODELS TO EXPERIMENTAL DATA

Hosted By
Matthias Bethge, Max Planck Institute for Biological Cybernetics, Germany
Philipp Berens, Max Planck Institute for Biological Cybernetics, Germany
Jakob H. Macke, University College London, United Kingdom

Modern recording techniques such as multi-electrode arrays and 2-photon imaging are capable of simultaneously monitoring the activity of large neuronal ensembles at single cell resolution. This makes it possible to study the dynamics of neural populations of considerable size, and to gain insights into their computations and functional organization. The key challenge with multi-electrode recordings is their high-dimensional nature. Understanding this kind of data requires powerful statistical techniques for capturing the structure of the neural population responses and their relation with external stimuli or behavioral observations.

Contributions to this Research Topic should advance statistical modeling of neural populations. Questions of particular interest include:

1. What classes of statistical methods are most useful for modeling population activity?
2. What are the main limitations of current approaches, and what can be done to overcome them?
3. How can statistical methods be used to empirically test existing models of (probabilistic) population coding?
4. What role can statistical methods play in formulating novel hypotheses about the principles of information processing in neural populations?

Table of Contents

frontiers in
COMPUTATIONAL NEUROSCIENCE

EDITORIAL
published: 28 July 2011
doi: 10.3389/fncom.2011.00035

Statistical analysis of multi-cell recordings: linking population coding models to experimental data

Jakob Macke[1,2,3], Philipp Berens[1,2,4,5]* and Matthias Bethge[1,2,4]

[1] *Computational Vision and Neuroscience Group, Max Planck Institute for Biological Cybernetics, Tübingen, Germany*
[2] *Werner Reichardt Centre for Integrative Neuroscience and Institute of Theoretical Physics, University of Tübingen, Tübingen, Germany*
[3] *Gatsby Unit for Computational Neuroscience, University College London, London, UK*
[4] *Bernstein Center for Computational Neuroscience Tübingen, Tübingen, Germany*
[5] *Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA*
**Correspondence: philipp.berens@tuebingen.mpg.de*

Modern recording techniques such as multi-electrode arrays and two-photon imaging methods are capable of simultaneously monitoring the activity of large neuronal ensembles at single cell resolution. These methods finally give us the means to address some of the most crucial questions in systems neuroscience: what are the dynamics of neural population activity? How do populations of neurons perform computations? What is the functional organization of neural ensembles?

While the wealth of new experimental data generated by these techniques provides exciting opportunities to test ideas about how neural ensembles operate, it also provides major challenges: multi-cell recordings necessarily yield data which is high-dimensional in nature. Understanding this kind of data requires powerful statistical techniques for capturing the structure of the neural population responses, as well as their relationship with external stimuli or behavioral observations. Furthermore, linking recorded neural population activity to the predictions of theoretical models of population coding has turned out not to be straightforward.

These challenges motivated us to organize a workshop at the 2009 Computational Neuroscience Meeting in Berlin to discuss these issues. In order to collect some of the recent progress in this field, and to foster discussion on the most important directions and most pressing questions, we issued a call for papers for this Research Topic. We asked authors to address the following four questions:

1. What classes of statistical methods are most useful for modeling population activity?
2. What are the main limitations of current approaches, and what can be done to overcome them?
3. How can statistical methods be used to empirically test existing models of (probabilistic) population coding?
4. What role can statistical methods play in formulating novel hypotheses about the principles of information processing in neural populations?

A total of 15 papers addressing questions related to these themes are now collected in this Research Topic. Three of these articles have resulted in "Focused reviews" in Frontiers in Neuroscience (Crumiller et al., 2011; Rosenbaum et al., 2011; Tchumatchenko et al., 2011), illustrating the great interest in the topic. Many of the articles are devoted to a better understanding of how correlations arise in neural circuits, and how they can be detected, modeled, and interpreted. For example, by modeling how pairwise correlations are transformed by spiking non-linearities in simple neural circuits, Tchumatchenko et al. (2010) show that pairwise correlation coefficients have to be interpreted with care, since their magnitude can depend strongly on the temporal statistics of their input-correlations. In a similar spirit, Rosenbaum et al. (2010) study how correlations can arise and accumulate in feed-forward circuits as a result of pooling of correlated inputs.

Lyamzin et al. (2010) and Krumin et al. (2010) present methods for simulating correlated population activity and extend previous work to more general settings. The method of Lyamzin et al. (2010) allows one to generate synthetic spike trains which match commonly reported statistical properties, such as time varying firing rates as well signal and noise correlations. The Hawkes framework presented by Krumin et al. (2010) allows one to fit models of recurrent population activity to the correlation-structure of experimental data. Louis et al. (2010) present a novel method for generating surrogate spike trains which can be useful when trying to assess the significance and time-scale of correlations in neural spike trains. Finally, Pipa and Munk (2011) study spike synchronization in prefrontal cortex during working memory.

A number of studies are also devoted to advancing our methodological toolkit for analyzing various aspects of population activity (Gerwinn et al., 2010; Machens, 2010; Staude et al., 2010; Yu et al., 2010). For example, Gerwinn et al. (2010) explain how full probabilistic inference can be performed in the popular model class of generalized linear models (GLMs), and study the effect of using prior distributions on the parameters of the stimulus and coupling filters. Staude et al. (2010) extend a method for detecting higher-order correlations between neurons via population spike counts to non-stationary settings. Yu et al. (2010) describe a new technique for estimating the information rate of a population of neurons using frequency-domain methods. Machens (2010) introduces a novel extension of principal component analysis for separating the variability of a neural response into different sources.

Focusing less on the spike responses of neural populations but on aggregate signals of population activity, Boatman-Reich et al. (2010) and Hoerzer et al. (2010) describe methods for a quantitative analysis of field potential recordings. While Boatman-Reich et al. (2010) discuss a number of existing techniques in a unified framework and highlight the potential pitfalls associated with such approaches, Hoerzer et al. (2010) demonstrate how multivariate autoregressive models and the concept of Granger causality can be used to infer local functional connectivity in area V4 of behaving macaques.

A final group of studies is devoted to understanding experimental data in light of computational models (Galán et al., 2010; Pandarinath et al., 2010; Shteingart et al., 2010). Pandarinath et al. (2010) present a novel mechanism that may explain how neural networks in the retina switch from one state to another by a change in gap junction coupling, and conjecture that this mechanism might also be found in other neural circuits. Galán et al. (2010) present a model of how hypoxia may change the network structure in the respiratory networks in the brainstem, and analyze neural correlations in multi-electrode recordings in light of this model. Finally, Shteingart et al. (2010) show that the spontaneous activation sequences they find in cultured networks cannot be explained by Zipf's law, but rather require a wrestling model.

The papers of this Research Topic thus span a wide range of topics in the statistical modeling of multi-cell recordings. Together with other recent advances, they provide us with a useful toolkit to tackle the challenges presented by the vast amount of data collected with modern recording techniques. The impact of novel statistical methods on the field and their potential to generate scientific progress, however, depends critically on how readily they can be adopted and applied by laboratories and researchers working with experimental data. An important step toward this goal is to also publish computer code along with the articles (Barnes, 2010) as a successful implementation of advanced methods also relies on many details which are hard to communicate in the article itself. In this way it becomes much more likely that other researchers can actually use the methods, and unnecessary re-implementations can be avoided. Some of the papers in this Research Topic already follow this goal (Gerwinn et al., 2010; Louis et al., 2010; Lyamzin et al., 2010). We hope that this practice becomes more and more common in the future and encourage authors and editors of Research Topics to make as much code available as possible, ideally in a format that can be easily integrated with existing software sharing initiatives (Herz et al., 2008; Goldberg et al., 2009).

REFERENCES

Barnes, N. (2010). Publish your computer code: it is good enough. *Nature* 467, 753.

Boatman-Reich, D., Franaszczuk, P. J., Korzeniewska, A., Caffo, B., Ritzl, E. K., Colwell, S., and Crone, N. E. (2010). Quantifying auditory event-related responses in multichannel human intracranial recordings. *Front. Comput. Neurosci.* 4:4. doi: 10.3389/fncom.2010.00004

Crumiller, M., Knight, B., Yu, Y., and Kaplan, E. (2011). Estimating the amount of information conveyed by a population of neurons. *Front. Neurosci.* 5:90. doi: 10.3389/fnins.2011.00090

Galán, R. F., Dick, T. E., and Baekey, D. M. (2010). Analysis and modeling of ensemble recordings from respiratory pre-motor neurons indicate changes in functional network architecture after acute hypoxia. *Front. Comput. Neurosci.* 4:131. doi: 10.3389/fncom.2010.00131

Gerwinn, S., Macke, J. H., and Bethge, M. (2010). Bayesian inference for generalized linear models for spiking neurons. *Front. Comput. Neurosci.* 4:12. doi: 10.3389/fncom.2010.00012

Goldberg, D. H., Victor, J. D., Gardner, E. P., and Gardner, D. (2009). Spike train analysis toolkit: enabling wider application of information-theoretic techniques to neurophysiology. *Neuroinformatics* 7, 165–178.

Herz, A. V. M., Meier, R., Nawrot, M. P., Schiegel, W., and Zito, T. (2008). G-Node: an integrated tool-sharing platform to support cellular and systems neurophysiology in the age of global neuroinformatics. *Neural Netw.* 21, 1070–1075.

Hoerzer, G. M., Liebe, S., Schloegl, A., Logothetis, N. K., and Rainer, G. (2010). Directed coupling in local field potentials of macaque V4 during visual short-term memory revealed by multivariate autoregressive models. *Front. Comput. Neurosci.* 4:14. doi: 10.3389/fncom.2010.00014

Krumin, M., Reutsky, I., and Shoham, S. (2010). Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Front. Comput. Neurosci.* 4:147. doi: 10.3389/fncom.2010.00147

Louis, S., Gerstein, G. L., Grün, S., and Diesmann, M. (2010). Surrogate spike train generation through dithering in operational time. *Front. Comput. Neurosci.* 4:127. doi: 10.3389/fncom.2010.00127

Lyamzin, D. R., Macke, J. H., and Lesica, N. A. (2010). Modeling population spike trains with specified time-varying spike rates, trial-to-trial variability, and pairwise signal and noise correlations. *Front. Comput. Neurosci.* 4:144. doi: 10.3389/fncom.2010.00144

Machens, C. K. (2010). Demixing population activity in higher cortical areas. *Front. Comput. Neurosci.* 4:126. doi: 10.3389/fncom.2010.00126

Pandarinath, C., Bomash, I., Victor, J. D., Prusky, G. T., Tschetter, W. W., and Nirenberg, S. (2010). A novel mechanism for switching a neural system from one state to another. *Front. Comput. Neurosci.* 4:2. doi: 10.3389/fncom.2010.00002

Pipa, G., and Munk, M. H. J. (2011). Higher order spike synchrony in prefrontal cortex during visual memory. *Front. Comput. Neurosci.* 5:23. doi: 10.3389/fncom.2011.00023

Rosenbaum, R., Trousdale, J., and Josi , K. (2011). The effects of pooling on spike train correlations. *Front. Neurosci.* 5:58. doi: 10.3389/fnins.2011.00058

Rosenbaum, R. J., Trousdale, J., and Josi , K. (2010). Pooling and correlated neural activity. *Front. Comput. Neurosci.* 4:9. doi: 10.3389/fncom.2010.00009

Shteingart, H., Raichman, N., Baruchi, I., and Ben-Jacob, E. (2010). Wrestling model of the repertoire of activity propagation modes in quadruple neural networks. *Front. Comput. Neurosci.* 4:25. doi: 10.3389/fncom.2010.00025

Staude, B., Grün, S., and Rotter, S. (2010). Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference. *Front. Comput. Neurosci.* 4:16. doi: 10.3389/fncom.2010.00016

Tchumatchenko, T., Geisel, T., Volgushev, M., and Wolf, F. (2010). Signatures of synchrony in pairwise count correlations. *Front. Comput. Neurosci.* 4:1. doi: 10.3389/neuro.10.001.2010

Tchumatchenko, T., Geisel, T., Volgushev, M., and Wolf, F. (2011). Spike correlations – what can they tell about synchrony? *Front. Neurosci.* 5:68. doi: 10.3389/fnins.2011.00068

Yu, Y., Crumiller, M., Knight, B., and Kaplan, E. (2010). Estimating the amount of information carried by a neuronal population. *Front. Comput. Neurosci.* 4:10. doi: 10.3389/fncom.2010.00010

Received: 15 June 2011; accepted: 14 July 2011; published online: 28 July 2011.

Citation: Macke J, Berens P and Bethge M (2011) Statistical analysis of multi-cell recordings: linking population coding models to experimental data. Front. Comput. Neurosci. ***5****:35. doi: 10.3389/fncom.2011.00035*

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 19 April 2010
doi: 10.3389/fncom.2010.00009

Pooling and correlated neural activity

Robert J. Rosenbaum*, James Trousdale and Krešimir Josić

Department of Mathematics, College of Natural Sciences and Mathematics, University of Houston, Houston, TX, USA

Edited by:
Philipp Berens,
Baylor College of Medicine, USA
MaxPlanck Institute for Biological Cybernetics, Germany

Reviewed by:
Nestor Parga,
Columbia University, USA
John A. Hertz,
Niels Bohr Institute, Denmark
Arvind Kumar,
University of Freiburg, Germany

****Correspondence:***
Robert J. Rosenbaum,
University of Houston,
Department of Mathematics, Houston,
TX 77204-3008, USA.
e-mail: robertr@math.uh.edu

Correlations between spike trains can strongly modulate neuronal activity and affect the ability of neurons to encode information. Neurons integrate inputs from thousands of afferents. Similarly, a number of experimental techniques are designed to record pooled cell activity. We review and generalize a number of previous results that show how correlations between cells in a population can be amplified and distorted in signals that reflect their collective activity. The structure of the underlying neuronal response can significantly impact correlations between such pooled signals. Therefore care needs to be taken when interpreting pooled recordings, or modeling networks of cells that receive inputs from large presynaptic populations. We also show that the frequently observed runaway synchrony in feedforward chains is primarily due to the pooling of correlated inputs.

Keywords: correlation, pooling, synchrony, feedforward networks, synfire chains

INTRODUCTION

Cortical neurons integrate inputs from thousands of afferents. Similarly, a variety of experimental techniques record the pooled activity of large populations of cells. It is therefore important to understand how the structured response of a neuronal network is reflected in the pooled activity of cell groups.

It is known that weak dependencies between the response of cell pairs in a population can have a significant impact on the variability and signal-to-noise ratio of the pooled signal (Shadlen and Newsome, 1998; Salinas and Sejnowski, 2000; Moreno-Bote et al., 2008). It has also been observed that weak correlations between cells in two populations can cause much stronger correlations between the pooled activity of the populations (Bedenbaugh and Gerstein, 1997; Chen et al., 2006; Gutnisky and Josic´, 2010; Renart et al., 2010). We give a simple example of this effect in **Figure 1C**: Weak correlations were introduced between the spiking activity of cells in two non-overlapping presynaptic pools each providing input to a postsynaptic cell (see diagram in **Figure 1B**). The activity between pairs of excitatory, and pairs of inhibitory cells was correlated, but excitatory–inhibitory pairs were uncorrelated. Even without shared inputs and with background noise, pooling resulted in strong correlations in postsynaptic membrane voltages. The connectivity in the presynaptic network was irrelevant – it only mattered that the inputs to the downstream neurons reflected the pooled activity of the afferent populations. A similar effect can cause large correlations between recordings of multiunit activity (MUA) or recordings of voltage sensitive dyes (VSD), even when correlations between cells in the recorded populations are small (Bedenbaugh and Gerstein, 1997; Chen et al., 2006; Stark et al., 2008). The effect is the same, but in this case pooling occurs at the level of a recording device rather than a downstream neuron (compare **Figures 1A,B**).

We present a systematic overview, as well as extensions and applications of a number of previous observations related to this phenomenon. Using a linear model, we start by examining the potential effects of pooling on recordings from large populations obtained using VSD or MUA recording techniques. These techniques are believed to reflect the pooled postsynaptic activity of groups of cells. We extend earlier models introduced to examine the impact of pooling on correlations (Bedenbaugh and Gerstein, 1997; Chen et al., 2006; Nunez and Srinivasan, 2006), and show that heterogeneities in the presynaptic pools can have subtle effects on correlations between pooled signals.

Since neurons respond to input from large presynaptic populations, pooling also impacts the activity of single cells and cell pairs. As observed in **Figure 1C**, pooling can inflate weak correlations between afferents. However, excitatory–inhibitory correlations (Okun and Lampl, 2008) can counteract this amplification, as shown in **Figure 1D** (Hertz, 2010; Renart et al., 2010). We examine these effects analytically by modeling the subthreshold activity of postsynaptic cells as a filtered version of the inputs received (Tetzlaff et al., 2008). The impact of correlated subthreshold activity on the output spiking statistics is a nontrivial question which we address only briefly (Moreno-Bote and Parga, 2006; de la Rocha et al., 2007; Ostojić et al., 2009).

The effects of pooling provide a simple explanation for certain aspects of the dynamics of feedforward chains. Simulations and *in vitro* experiments show that layered feedforward architectures give rise to a robust increase in synchronous spiking from layer to layer (Diesmann et al., 1999; Litvak et al., 2003; Reyes, 2003; Doiron et al., 2006; Kumar et al., 2008). We describe how output correlations in one layer impact correlations between the pooled inputs to the next layer. This approach is used to derive a mapping that describes how correlations develop across layers (Tetzlaff et al., 2003; Renart et al., 2010), and to illustrate that the pooling of correlated inputs is the primary mechanism responsible for the development of synchrony in feedforward chains. Examining how correlations are mapped between layers also helps explain why asynchronous states are rarely observed in feedforward networks in the absence of strong background noise (van Rossum et al., 2002;

Vogels and Abbott, 2005). This is in contrast to recurrent networks which can display stable asynchronous states (Hertz, 2010; Renart et al., 2010) similar to those observed *in vivo* (Ecker et al., 2010).

MATERIALS AND METHODS

CORRELATIONS BETWEEN STOCHASTIC PROCESSES

The cross-covariance of a pair of stationary stochastic processes, $x(t)$ and $y(t)$, is $C_{xy}(t) = \mathrm{cov}(x(s), y(s+t))$. The auto-covariance function, $C_{xx}(t)$, is the cross-covariance between a process and itself. The cross- and auto-covariance functions measure second order dependencies at time lag t between two processes, or a process and itself. We quantify the total magnitude of interactions over all time using the asymptotic statistics,

$$\gamma_{xy} = \int_{-\infty}^{\infty} C_{xy}(t)dt, \quad \sigma_x^2 = \gamma_{xx}, \quad \rho_{xy} = \frac{\gamma_{xy}}{\sigma_x \sigma_y}. \tag{1}$$

While the asymptotic correlation, ρ_{xy}, measures correlations between $x(t)$ and $y(t)$ over large timescales, the auto- and cross-covariance functions determine the timescale of these dependencies.

CORRELATIONS BETWEEN SUMS OF RANDOM VARIABLES

Given two collections of correlated random variables $\{x_i\}_{i=1}^{n_x}$ and $\{y_j\}_{j=1}^{n_y}$, define the pooled variables, $X = \Sigma_i x_i$ and $Y = \Sigma_i y_i$. Since covariance is bilinear ($\mathrm{cov}(\Sigma_i x_i, \Sigma_j y_j) = \Sigma_{ij}\mathrm{cov}(x_i, y_j)$) the variance and covariance of the pooled variables are

$$\sigma_X^2 = \sum_{i=1}^{n_x}\sum_{\substack{j=1 \\ j\neq i}}^{n_x} \sigma_{x_i}\sigma_{x_j}\rho_{x_i x_j} + \sum_{i=1}^{n_x}\sigma_{x_i}^2, \quad \text{and} \quad \gamma_{XY} = \sum_{i=1}^{nx}\sum_{j=1}^{ny}\sigma_{x_i}\sigma_{y_j}\rho_{x_i y_j},$$

and similarly for σ_Y^2.

Using these expressions along with some algebraic manipulation, the correlation coefficient, $\rho_{XY} = \gamma_{XY}/\sqrt{\sigma_X \sigma_Y}$, between the pooled variables can be written as

$$\rho_{XY} = \frac{\bar{\rho}xy}{\sqrt{\left[w_x\bar{\rho}_{xx} + \frac{1}{n_x}\left(\frac{\bar{v}_x}{\overline{\sigma_x\sigma_y}} - w_x\bar{\rho}_{xx}\right)\right]\left[w_y\bar{\rho}_{yy} + \frac{1}{n_y}\left(\frac{\bar{v}_y}{\overline{\sigma_x\sigma_y}} - w_y\bar{\rho}_{yy}\right)\right]}} \tag{2}$$

$$= \frac{\bar{\rho}_{xy}}{\sqrt{\bar{\rho}_{xx}\bar{\rho}_{yy}}} + \mathcal{O}\left(\frac{1}{\sqrt{n_x n_y}}\right), \tag{3}$$

where

$$w_x = \frac{\overline{\sigma_x\sigma_x}}{\overline{\sigma_x\sigma_y}}, \qquad \bar{v}_x = \frac{1}{n_x}\sum_{i=1}^{n_x}\sigma_{x_i}^2,$$

$$\overline{\sigma_x\sigma_y} = \frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sigma_{x_i}\sigma_{y_i}, \quad \overline{\sigma_x\sigma_x} = \frac{1}{n_x(n_x-1)}\sum_{i=1}^{n_x}\sum_{\substack{j=1 \\ j\neq 1}}^{n_x}\sigma_{x_i}\sigma_{x_j},$$

$$\bar{\rho}_{xy} = \frac{1}{n_x n_y \overline{\sigma_x\sigma_y}}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sigma_{x_i}\sigma_{y_j}\rho_{x_i y_j},$$

$$\bar{\rho}_{xx} = \frac{1}{n_x(n_x-1)\overline{\sigma_x\sigma_x}}\sum_{i=1}^{n_x}\sum_{\substack{j=1 \\ j\neq 1}}^{n_x}\sigma_{x_i}\sigma_{y_j}\rho_{x_i x_j}$$

and similarly for w_y, v_y, $\overline{\sigma_y\sigma_y}$, and $\bar{\rho}_{yy}$. In deriving Eq. (3) we assumed that all pairwise statistics are uniformly bounded away from zero in the asymptotic limit.

Each overlined term above is a population average. Notably, $\bar{\rho}_{xy}$ represents the average correlation between x_i and y_j pairs, weighted by the product of their standard deviations, and similarly for $\bar{\rho}_{xx}$ and $\bar{\rho}_{yy}$. Correlation between weighted sums can be obtained by substituting $x_i \to w_{x_i}x_i$ and $y_j \to w_{y_j}y_j$ for weights w_{x_i} and w_{y_j} and making the appropriate changes to the terms in the equation above (e.g., $\sigma_{x_i} \to |w_{x_i}|\sigma_{x_i}$, $\rho_{x_i y_j} \to \mathrm{sign}(w_i w_j)\rho_{x_i y_j}$). Overlap between the two populations can be modeled by taking $\rho_{x_i y_j} = 1$ for some pairs.

Assuming that variances are homogeneous within each population, that is $\sigma_{x_i} = \sigma_x$ and $\sigma_{y_j} = \sigma_y$ for $i = 1, \ldots, n_x$ and $j = 1, \ldots, n_y$, simplifies these expressions. In particular, $\bar{v}_x = \overline{\sigma_x\sigma_x} = \sigma_x^2$, $\overline{\sigma_x\sigma_y} = \sigma_x\sigma_y$, and

$$\rho_{XY} = \frac{\bar{\rho}_{xy}}{\sqrt{\left[\bar{\rho}_{xx} + \frac{1}{n_x}\left(1-\bar{\rho}_{xx}\right)\right]\left[\bar{\rho}_{yy} + \frac{1}{n_y}\left(1-\bar{\rho}_{yy}\right)\right]}}. \tag{4}$$

Assuming further that the populations are symmetric, $\sigma_x = \sigma_y = \sigma$, $n_x = n_y = n$, and $\bar{\rho}_{xx} = \bar{\rho}_{yy}$, the expression above simplifies to

$$\rho_{XY} = \frac{\rho^b}{\rho^w + \frac{1}{n}\left(1-\rho^w\right)} = \frac{\rho^b}{\rho^w} + \mathcal{O}\left(\frac{1}{n}\right), \tag{5}$$

where $\rho^b = \bar{\rho}_{xy}$ is the average pairwise correlation between the two populations and $\rho^w = \bar{\rho}_{xx} = \bar{\rho}_{yy}$ is the average pairwise correlation within each population. Eq. (5) was derived in Bedenbaugh and Gerstein (1997) in an examination of correlations between multiunit recordings. In Chen et al. (2006), a version of Eq. (5) with $\rho^w = \rho^b$ is derived in the context of correlations between two VSD signals. The asymptotic, $\rho_{xy} \to 0$, limit when $\rho^w = \rho^b$ is discussed in Renart et al. (2010).

Note that the results above hold for correlations computed over arbitrary time windows. We concentrate on infinite windows, and discuss extensions in the Appendix.

NEURON MODEL

In the second part of the presentation we consider two excitatory and two inhibitory input populations projecting to two postsynaptic cells. The j^{th} excitatory input to cell k is labeled $e_{j,k}(t)$ ($k = 1$ or 2). Similarly, $i_{j,k}(t)$ denotes the j^{th} inhibitory input to cell k. Each cell receives n_e excitatory and n_i inhibitory inputs with individual rates ν_e and ν_i respectively.

Each of the excitatory and inhibitory inputs to cell k, are stationary spike trains modeled by point processes, $e_{j,k}(t) = \Sigma_i \delta(t - t_{j,k}^i)$ and $i_{j,k}(t) = \Sigma_i \delta(t - s_{j,k}^i)$ where $\{t_{j,k}^i\}$ and $\{s_{j,k}^i\}$ are input spike times. We assume that the spike trains are stationary in a multivariate sense (Stratonovich, 1963). The pooled excitatory and inhibitory inputs to neuron k are $E_k(t) = \Sigma_{j=1}^{n_e} e_{j,k}(t)$, and $I_k(t) = \Sigma_{j=1}^{n_i} i_{j,k}(t)$.

To generate correlated inputs to cells, we used the multiple interaction process (MIP) method (Kuhn et al., 2003), then jittered each spike time independently by a random value drawn from an exponential distribution with mean 5ms. The resulting processes are

Poisson with cross-covariance functions proportional to a double exponential, $C_{xy}(t) \sim e^{-|t|/5}$. Note that since each input is Poisson, $\sigma_e^2 = \nu_e$ and $\sigma_i^2 = \nu_i$.

While the dynamics of the afferent population were not modeled explicitly, the response of the two downstream neurons was obtained using a conductance-based IF model. The membrane potentials of the neurons were described by

$$C_m \frac{dV_k}{dt} = -g_L(V_k - V_L) - g_{E_k}(t)(V_k - V_E) - g_{I_k}(t)(V_k - V_I), \qquad (6)$$

with excitatory and inhibitory conductances determined by $g_{E_k}(t) = (E_k * \alpha_e)(t)$ and $g_{I_k}(t) = (I_k * \alpha_i)(t)$ where $*$ denotes convolution. We used synaptic responses of the form $\alpha_e(t) = \mathcal{E}\tau_e^{-2} t e^{-t/\tau_e}\Theta(t)$ and $\alpha_i(t) = \mathcal{I}\tau_i^{-2} t e^{-t/\tau_i}\Theta(t)$ where $\Theta(t)$ is the Heaviside function. The *area* of a single excitatory or inhibitory postsynaptic conductance (EPSC or IPSC) is therefore equal to the synaptic weight, $\mathcal{E}$ or $\mathcal{I}$, with units nS·ms. This analysis can easily be extended to situations where each input, $e_{j,k}$ or $i_{j,k}$, has a distinct synaptic weight.

When examining spiking activity, we assume that when V_k crosses a threshold voltage, V^{th}, an output spike is produced and V_k is reset to V_L. When examining sub-threshold dynamics, we considered the free membrane potential without threshold.

As a measure of balance between excitation and inhibition we used (Troyer and Miller, 1997; Salinas and Sejnowski, 2000)

$$\beta = \frac{|V_L - V_E|\mathcal{E}\nu_e n_e}{|V_L - V_I|\mathcal{I}\nu_i n_i}.$$

When $\beta = 1$, the net excitation and inhibition are balanced and the mean free membrane potential equals V_L. In simulations, we set $V_L = -60$ mV, $V_E = 0$ mV, $V_I = -90$ mV, $\tau_e = 10$ ms, $\tau_i = 20$ ms, $C_m = 114$ pF, and $g_L = 4.086$ nS, giving a membrane time constant, $\tau_m = C_m/g_L = 27.9$ ms. In all simulations except those in **Figure 7**, the cells are balanced ($\beta = 1$).

The conductance-based IF neuron behaves as a nonlinear filter in the sense that membrane potentials cannot be written as a linear transformation of the inputs. However, following Kuhn et al. (2004) and Coombes et al. (2007), we derive a linear approximation to the conductance based model. Let $U = V_k - V_L$ so that Eq. (6) becomes

$$C_m \frac{dU}{dt} = \left(-g_L - g_E(t) - g_I(t)\right)U - g_E(t)\left(V_L - V_E\right) - g_I(t)\left(V_L - V_I\right).$$

Define the effective membrane time constant, $\tau_{eff} = C_m/(E[g_L + g_E(t) + g_I(t)]) = C_m/(g_L + n_e\nu_e\mathcal{E} + n_i\nu_i\mathcal{I}$ Substituting this average value in the previous equation yields the linear approximation to the conductance based model,

$$\frac{dU}{dt} = -\frac{1}{\tau_{eff}}U + J_k(t), \qquad (7)$$

where $J_k(t) = (-g_{E_k}(t)(V_L - V_E) - g_{I_k}(t)(V_L - V_I))/C_m$ is the total input current to cell k. Solving and reverting to the original variables gives the linear approximation $V_k(t) = (J_k * K)(t) + V_L$, where $K(t) = \Theta(t)e^{-t/\tau_{eff}}$ is the kernel of the linear filter induced by Eq. (7).

RESULTS

The pooling of signals from groups of neurons can impact both recordings of population activity and the structure of inputs to postsynaptic cells. We start by discussing correlations in pooled recordings using a simple linear model. A similar model is then used to examine the impact of pooling on the statistics of inputs to cells. For simplicity we assume that all spike trains are stationary. However, non-stationary results can be obtained using similar methods as outlined in the Section "Discussion." Though all parameters are defined in the Meterials and Methods, **Tables 1 and 2** in the Appendix contain brief descriptions of parameters for quick reference. Also, **Tables 3 and 4** summarize the values of parameters used for simulations throughout the article.

CORRELATIONS BETWEEN POOLED RECORDINGS

Pooling can impact correlations between recordings of population activity obtained from voltage sensitive dyes (VSDs), multi-unit recordings and other techniques. Such signals might each represent the summed activity of hundreds or thousands of neurons. Let two recorded signals, $X_1(t)$ and $X_2(t)$, represent the weighted activity of cells in two populations (see diagram in **Figure 1A**). If we assume homogeneity in the input variances and equal size of the recorded populations, Eq. (4) gives the correlation between the recorded signals

$$\rho_{X_1X_2} = \frac{\bar{\rho}_{12}}{\sqrt{\left[\bar{\rho}_{11} + \frac{1}{n}\left(1 - \bar{\rho}_{11}\right)\right]\left[\bar{\rho}_{22} + \frac{1}{n}\left(1 - \bar{\rho}_{22}\right)\right]}} = \frac{\bar{\rho}_{12}}{\sqrt{\bar{\rho}_{11}\bar{\rho}_{22}}} + \mathcal{O}\left(\frac{1}{n}\right). \qquad (8)$$

Here n represents the number of neurons recorded, $\bar{\rho}_{kk}$, $k = 1,2$ represents the average correlation between cells contributing to signal $X_k(t)$, and $\bar{\rho}_{12}$ represents the average correlation between cells contributing to different signals. The averages are weighted so that cells that contribute more strongly to the recording, such as those closer to the recording site, contribute more to the average correlations (see Materials and Methods). Cells common to both recorded populations can be modeled by setting the corresponding correlation coefficients to unity. A form of Eq. (8) with $\bar{\rho}_{11} = \bar{\rho}_{22}$ was derived by Bedenbaugh and Gerstein (1997).

When the two recording sites are nearby, so that $\bar{\rho}_{12} \approx \bar{\rho}_{11} \approx \bar{\rho}_{22}$, even small correlations between individual cells are amplified by pooling so that the correlations between the recorded signals can be close to 1. This effect was observed in experiments and explained in similar terms in Stark et al. (2008).

A significant stimulus-dependent change in correlations between individual cells might be reflected only weakly in the correlation between the pooled signals. This can occur, for instance, in recordings of large populations when $\bar{\rho}_{12}$, $\bar{\rho}_{11}$, and $\bar{\rho}_{22}$ are increased by the same factor when a stimulus is presented. Similarly, an increase in correlations between cells can actually lead to a *decrease* in correlations between recorded signals when $\bar{\rho}_{11}$ and $\bar{\rho}_{22}$ increase by a larger factor than $\bar{\rho}_{12}$.

To illustrate these effects, we construct a simple model of stimulus dependent correlations motivated by the experiments in Chen et al. (2006), in which VSDs were used to record the population response in visual area V1 during an attention task. In their experiments,

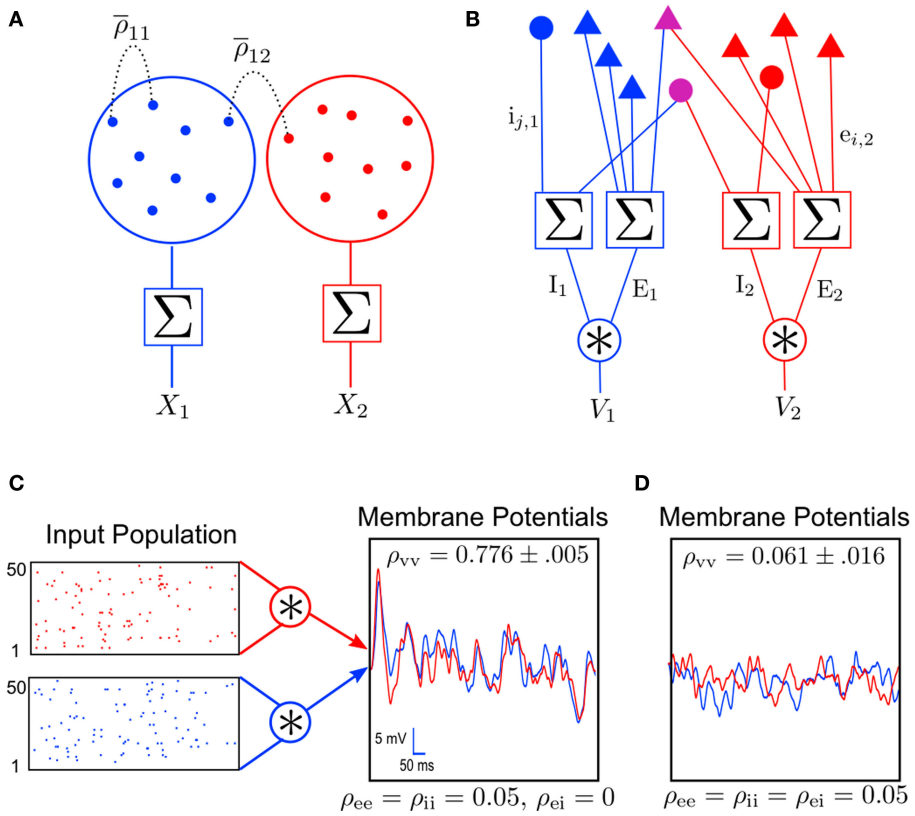


FIGURE 1 | Models of pooled recordings and the effects of pooling on correlations. (A) Pooling in experimental recordings. Cells from different populations are correlated with average correlation coefficient $\bar{\rho}_{12}$ and cells from the same population have average correlation coefficient $\bar{\rho}_{11}$ or $\bar{\rho}_{22}$. We examine correlations between the pooled signals, X_1 and X_2. **(B)** Pooled inputs to cell pairs. Individual excitatory (e) and inhibitory (i) inputs are correlated with coefficient ρ_{ee}, ρ_{ii}, and ρ_{ei}, respectively. The total input to a cell is the summed activity of its excitatory (E_k) and inhibitory (I_k) presynaptic population. The membrane potentials, V_1 and V_2, are obtained by filtering these inputs. **(C)** A simulation of the setup in **(B)** with background noise. Correlations between excitatory and between inhibitory cells were uniform ($\rho_{ee} = \rho_{ii} = 0.05$), but excitatory–inhibitory correlations were absent ($\rho_{ei} = 0$). The raster plot shows the activity in a subset of the input population. The correlation coefficient between the sub-threshold activity of the postsynaptic cells was $\rho_{vv} = 0.768 \pm 0.001$ s.e. Each cell receives 250 correlated and 250 uncorrelated excitatory Poisson inputs as well as 84 correlated and 84 uncorrelated inhibitory Poisson inputs ($n_e = 250$ and $n_i = 84$, $q_e = q_i = 1$, $\nu_e = 5$ Hz, $\nu_i = 7.5$ Hz, $\mathcal{I} = 4\mathcal{E}$, and $\mathcal{E} \approx 2.3$ nS·ms – see Materials and Methods and **Figure 3A** for notation and precise model description). **(D)** Same as **(B)**, but with $\rho_{ei} = 0.05$, $\nu_e = \nu_i = 5$ Hz, $\mathcal{I} = 6\mathcal{E}$ to maintain balance. In this case $\rho_{vv} = 0.0085 \pm 0.0024$ s.e.. The simulations were run 8000 times at 10 s each.

the imaged area is divided into 64 pixels, each 0.25 mm × .25 mm in size. The signal recorded from each pixel represents the pooled activity of $n \approx 1.25 \times 10^4$ neurons.

We model correlations between the signals, $X_1(t)$ and $X_2(t)$, recorded from two pixels in the presence or absence of a stimulus (see **Figure 2B**), using a simplified model of stimulus dependent rates and correlations. The firing rate of a cell located at distance d from the center of the retinotopic image of a stimulus is

$$r(d) = \begin{cases} B + \dfrac{(1-B)(1+\cos(d\pi))^{\lambda}}{2} & \text{stimulus present} \\ B & \text{stimulus absent.} \end{cases} \quad (9)$$

Here, $B \in [0,1]$ represents baseline activity and $\lambda \geq 1$ controls the rate at which activity decays with d. Both d and r were scaled so that their maximum value is 1 (see **Figure 2A**).

We assume that the correlation between the responses of two neurons is proportional to the geometric mean of their firing rates (de la Rocha et al., 2007; Shea-Brown et al., 2008), and that correlations decay exponentially with cell distance (Smith and Kohn, 2008; see however Poort and Roelfsema, 2009; Ecker et al., 2010). We therefore model the correlation between two cells as $\rho_{jk} = S\sqrt{r(d_j)r(d_k)}\,e^{-\alpha D_{j,k}}$ where d_j and d_k are the distances from each cell to the center of the retinotopic image of the stimulus, $D_{j,k}$ is the distance between cells j and k, α is the rate at which correlations decay with distance, and $S \leq 1$ is a constant of proportionality.

If pixels are small compared to the scales at which correlations are assumed to decay, then the average correlation between cells within the same pixel are $\bar{\rho}_{11} = Sr(d_1)$ and $\bar{\rho}_{22} = Sr(d_2)$. The average correlation between cells in different pixels is $\bar{\rho}_{12} = S\sqrt{r(d_1)r(d_2)}\,e^{-\alpha D_{1,2}}$.

In this case, whether a stimulus is present or not, the correlation between the pooled signals is of the form $\rho_{X_1X_2} = e^{-\alpha D_{1,2}} + \mathcal{O}(1/n)$. Thus, even significant stimulus dependent changes in correlations would be invisible in the recorded signals. This overall trend is consistent with the results in Chen et al. (2006) (compare **Figure 2C** to their Figure 2f). In such settings, it is difficult to conclude whether pairwise correlations are stimulus dependent or not from the pooled data.

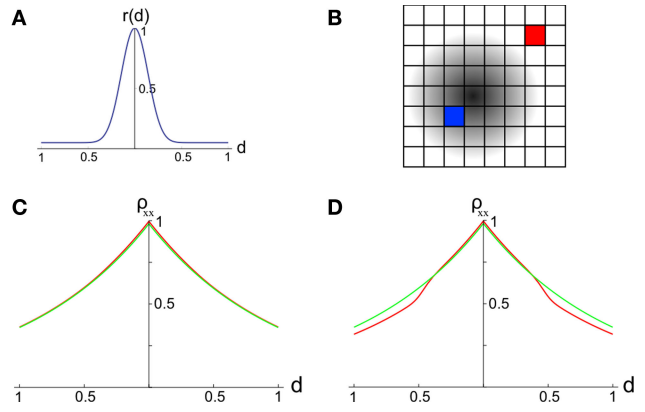


FIGURE 2 | The effect of pooling on recordings of stimulus dependent correlations. (A) The response amplitude of a model neuron as a function of its distance from the retinotopic image of a stimulus [Eq. (9)] with $B = 0.05$ and $\lambda = 10$. **(B)** A diagram of our model. Signals $X_1(t)$ and $X_2(t)$ are recorded from two pixels (red and blue squares). The activity in response to a stimulus is shown as a gradient centered at some pixel (the center of the retinotopic image of the stimulus). **(C)** The prediction of the correlation between two pixels obtained using the stimulus-dependent model considered in the text with stimulus present (red) and absent (green). We assumed that one pixel is located at the stimulus center ($d_1 = 0$). Parameters are as in **(A)** with $\alpha = 1$, $S = 0.1$, and $n = 1.25\times10^4$. A stimulus dependent change in correlations is undetectable. **(D)** Same as in **(C)**, except that baseline activity, B, was scaled by 0.5 in the presence of a stimulus. Compare to Figure 2f in Chen et al., 2006.

However, in Supplementary Figure 3 of Chen et al. (2006) the presence of a stimulus apparently results in a slight decrease in correlations between more distant pixels. In **Figure 2D** this effect was reproduced using the alternative model described above, with the additional assumption that baseline activity, B, decreases in the presence of a stimulus (Mitchell et al., 2009). The effect can also be reproduced by assuming that spatial correlation decay, α, increases when a stimulus is present.

As this example shows, care needs to be taken when inferring underlying correlation structures from pooled activity. The statistical structure of the recordings can depend on pairwise correlations between individual cells in a subtle way, and different underlying correlation structures may be difficult to distinguish from the pooled signals. However, downstream neurons may also be insensitive to the precise structure of pairwise correlations, as they are driven by the pooled input from many afferents.

CORRELATIONS BETWEEN THE POOLED INPUTS TO CELLS

We next examine the effects of pooling by relating the correlations between the activity of downstream cells to the pairwise correlations between cells in the input populations (see **Figure 1B**). The idea that pooling amplifies correlations carries over from the previous section. However, the presence of inhibition and non-instantaneous synaptic responses introduces new issues.

A homogeneous population with overlapping and independent inputs

For simplicity, we first consider a homogeneous population model (see **Figure 3A**). Each cell receives n_e inputs from a homogeneous pool of inputs with pairwise correlation coefficients ρ_{ee} and an additional $q_e n_e$ inputs from an outside pool of independent inputs. The two cells share $p_e n_e$ of the inputs drawn from the correlated pool. Processes in the independent pool are uncorrelated with all other processes. All excitatory inputs have variance σ_e^2.

The correlation between the pooled excitatory inputs is given by (see Appendix)

$$\rho_{E_1E_2} = \frac{\rho_{ee} + \frac{p_e}{n_e}(1-\rho_{ee})}{\rho_{ee} + \frac{1}{n_e}(1-\rho_{ee}+q_e)}. \tag{10}$$

A form of this equation, with $p_e = 0$ and $q_e = 0$, is derived in Chen et al. (2006). In the absence of correlations between processes in the input pools, $\rho_{ee} = 0$, the correlation between the pooled signals is just the proportion of shared inputs, $\rho_{E_1E_2} = p_e$. When $\rho_{ee} > 0$ and n_e is large, pooled excitatory inputs are highly correlated, even when pairwise correlations in the presynaptic pool, ρ_{ee}, are small, and the neurons do not share inputs ($p_e = 0$). Even when most inputs to the downstream cells are independent ($q_e > 1$), correlations between the pooled signals will be nearly 1 for sufficiently large input pools (see **Figure 4A**).

Under analogous homogeneity assumptions for the inhibitory pools, the correlation, $\rho_{I_1I_2}$, between the pooled inhibitory inputs is given by an equation identical to Eq. (10), and the correlation between the pooled excitatory and inhibitory inputs is given by

$$\rho_{E_1I_2} = \rho_{I_1E_2} = \frac{\rho_{ei}}{\sqrt{\left(\rho_{ee} + \frac{1}{n_e}(1-\rho_{ee}+q_e)\right)\left(\rho_{ii} + \frac{1}{n_i}(1-\rho_{ii}+q_i)\right)}}. \tag{11}$$

Interestingly, since $|\rho_{E_1I_2}| \le 1$, pairwise excitatory–inhibitory correlations obey the bound $|\rho_{ei}| \le \sqrt{\rho_{ee}\rho_{ii}} + \mathcal{O}(1/\sqrt{n_e n_i})$. Combining this inequality with Eq. (10) and the analogous equation for $\rho_{I_1I_2}$, it follows that $|\rho_{E_1I_2}| \le \sqrt{\rho_{E_1E_2}\rho_{I_1I_2}} + \mathcal{O}(1/\sqrt{n_e n_i})$ for homogeneous populations. These are a result of the non-negative definiteness of covariance matrices.

Heterogeneity and the effects of spatially dependent correlations

We next discuss how heterogeneity can dampen the amplification of correlations due to pooling. In the absence of any homogeneity assumptions on the excitatory input population (see the population model in the Materials and Methods), Eq. (3) gives the pooled excitatory signals, $\rho_{E_1E_2} = \bar{\rho}_{e_1e_2} / \sqrt{\bar{\rho}_{e_1e_1}\bar{\rho}_{e_2e_2}} + \mathcal{O}(1/\sqrt{n_{e_1} n_{e_2}})$. The term $\bar{\rho}_{e_1e_2}$ is a weighted average of the correlation coefficients *between* the two excitatory populations, and $\bar{\rho}_{e_1e_1}$ and $\bar{\rho}_{e_2e_2}$ are weighted averages of the correlations *within* each excitatory input population.

To illuminate this result, we assume symmetry between the populations: Let $n_{e_k} = n_e$ and $\sigma_{e_k^j} = \sigma_e$ for $k = 1,2$ and $j = 1,2, \ldots, n_e$, and assume $\bar{\rho}_{e_1e_1} = \bar{\rho}_{e_2e_2}$. The average "within" and "between" correlations, are $\rho_{ee}^w = \bar{\rho}_{e_1e_1} = \bar{\rho}_{e_2e_2}$ and $\rho_{ee}^b = \bar{\rho}_{e_1e_2}$ respectively (see **Figure 3B**). Under these assumptions, Eq. (5) can be applied to obtain (See also Bedenbaugh and Gerstein, 1997)

$$\rho_{E_1E_2} = \frac{\rho_{ee}^b}{\rho_{ee}^w + \frac{1}{n_e}(1-\rho_{ee}^w)} = \frac{\rho_{ee}^b}{\rho_{ee}^w} + \mathcal{O}\left(\frac{1}{n_e}\right) \tag{12}$$

which is plotted in **Figure 4A** (green line) and **Figure 4B**. For large n_e, the correlation between the pooled signals is the ratio of "between" and "within" correlations.

This observation has implications for a situation ubiquitous in the cortex. A neuron is likely to receive afferents from cells that are physically close. The activity of nearby cells may be more strongly correlated than the activity of more distant cells (Chen et al., 2006; Smith and Kohn, 2008). We therefore expect that pairwise correlations *within* each input pool are on average larger than correlations *between* two input pools, that is, $\rho^w_{ee} > \rho^b_{ee}$. This reduces the correlation between the inputs, regardless of the input population size.

An increase in correlations in the presynaptic pool can also decorrelate the pooled signals. If correlations *within* each input pool increase by a greater amount than correlations *between* the two pools, then the variance in the input to each cell will increased by a larger amount than the covariance between the inputs. As a consequence the correlations between the pooled inputs will be reduced. Modulations in correlation have been observed as a consequence of attention in V4 (Cohen and Maunsell, 2009; Mitchell et al., 2009; but apparently not in V1, Roelfsema et al., 2004). Such changes may be, in part, a consequence of small changes in "within" correlations between neurons in V1.

Equation 12 implies that correlations between large populations cannot be significantly larger than the correlations within each population. Since $|\rho_{E_1E_2}| \le 1$, it follows that $|\rho^b_{ee}| \le |\rho^w_{ee}| + \mathcal{O}(1/n_e)$.

The correlation, $\rho_{I_1I_2}$, between the pooled inhibitory inputs is given by an identical equation to Eq. (12) and the correlation between the pooled excitatory and inhibitory inputs is given by

$$\rho_{E_1I_2} = \rho_{I_1E_2} = \frac{\rho^b_{ei}}{\sqrt{\left(\rho^w_{ee} + \frac{1}{n_e}\left(1-\rho^w_{ee}\right)\right)\left(\rho^w_{ii} + \frac{1}{n_i}\left(1-\rho^w_{ii}\right)\right)}} = \frac{\rho^b_{ei}}{\sqrt{\rho^w_{ee}\rho^w_{ii}}} + \mathcal{O}\left(\frac{1}{\sqrt{n_e n_i}}\right).$$

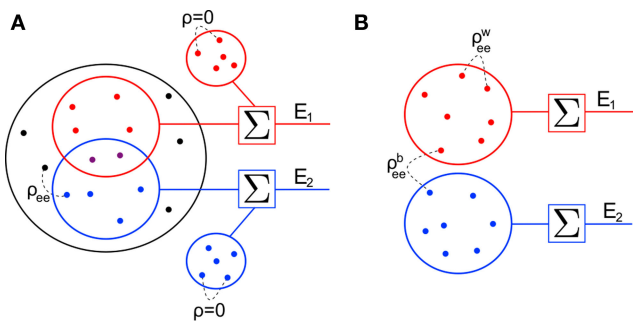


FIGURE 3 | Two population models considered in the text. (A) Homogeneous population with overlap and independent inputs: A homogeneous pool of correlated inputs (large black circle) with correlation coefficient between any pair of processes equal to ρ_{ee}. Each cell draws n_e inputs (larger red and blue circles) from this homogeneous input pool. Of these n_e correlated inputs, $p_e n_e$ are shared between the two neurons (purple dots). In addition, each cell receives $q_e n_e$ independent inputs (smaller red and blue circles), for a total of $n_e + q_e n_e$ inputs. All inputs have variance σ^2_e. **(B)** A population model with distinct "within" and "between" correlations: Each cell receives n_e inputs. The average correlation between two inputs to the same cell is ρ^w_{ee}, and between inputs to different cells is ρ^b_{ee}.

Correlations between the free membrane potentials

We now look at the correlation between the free membrane potentials of two downstream neurons. The *free* membrane potentials are obtained by assuming an absence of threshold or spiking activity. For simplicity we assume symmetry in the statistics of the inputs to the postsynaptic cells: $\sigma_{E_k} = \sigma_E$, $\sigma_{I_k} = \sigma_I$, $\rho_{E_1I_2} = \rho_{E_2I_1}$, $\rho_{E_1I_1} = \rho_{E_2I_2}$, $\rho_{E_1E_1} = \rho_{E_2E_2}$ and $\rho_{I_1I_1} = \rho_{I_2I_2}$. The analysis is similar in the asymmetric case.

In the Section "Materials and Methods", we derive a linear approximation of the free membrane potentials,

$$V_k(t) = \left(J_k * K\right)(t) + V_L,$$

where $J_k(t) = -(g_{E_k}(t)(V_L - V_E) + g_{I_k}(t)(V_L - V_I))/C_m$ are the total input currents and $K(t) = \Theta(t)e^{-t/\tau_{eff}}$ for $k = 1,2$. Under this approximation, the correlation, $\rho_{V_1V_2}$, between the membrane potentials is

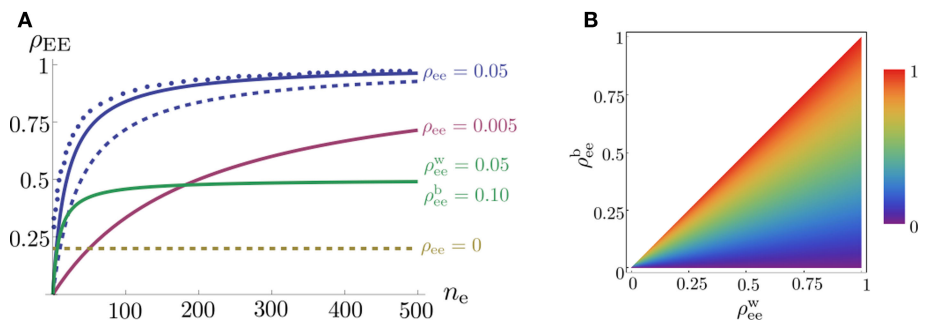


FIGURE 4 | The effect of pooling on correlations between summed input spike trains. (A) The correlation coefficient between the pooled excitatory spike trains (ρ_{EE}) is shown as a function of the size of the correlated excitatory input pool (n_e) for various parameter settings. The solid blue line was obtained by setting $\rho_{ee} = 0.05$ for the population model in **Figure 3A** in the absence of shared or independent inputs ($p_e = q_e = 0$). The dashed line illustrates the decorrelating effects of the addition of n_e independent inputs ($q_e = 1$, $q_e = 0$, $\rho_{ee} = 0.05$). The dotted blue line shows that shared inputs increase correlations, but have a diminishing effect on ρ_{EE} with increasing input population size ($p_e = 0.2$, $q_e = 0$, $\rho_{ee} = 0.05$). The solid pink line shows the effect of reducing the pairwise input correlations ($\rho_{ee} = 0.005$, $p_e = q_e = 0$). The dashed tan line was obtained with uncorrelated inputs so that correlations reflected shared inputs alone ($p_e = 0.2$, $\rho_{ee} = q_e = 0$). The green line was obtained with disparity in the "within" and "between" correlations ($\rho^b_e = 0.05$ and $\rho^w_e = 0.1$) using the model in **Figure 3B. (B)** The correlations coefficient, ρ_{EE}, between the pooled inputs as a function of the within and between correlations (ρ^b_e and ρ^w_e) for $n_e = 50$. Note that the pooled correlation is relatively constant along lines through the origin. Thus, changing ρ^b_e and ρ^w_e by the same proportion does not affect the pooled correlation.

equal to the correlation, $\rho_{in} = \rho_{J_1 J_2}$, between the total input currents and can be written as a weighted average of the pooled excitatory and inhibitory spike train correlations (see Appendix),

$$\rho_{V_1 V_2} \approx \rho_{in} = \frac{W_E^2 \rho_{E_1 E_2} + W_I^2 \rho_{I_1 I_2} - 2 W_E W_I \rho_{E_1 I_2}}{W_E^2 + W_I^2 - 2 W_E W_I \rho_{E_1 I_1}} \quad (13)$$

where $\rho_{E_1 E_2}, \rho_{E_1 I_2}, \rho_{I_1 I_2}$ and $\rho_{E_1 I_1}$ are derived above, and $W_E = \mathcal{E}|V_E - V_L|\sigma_E$ and $W_I = \mathcal{I}|V_I - V_L|\sigma_I$ are weights for the excitatory and inhibitory contributions to the correlation. In **Figure 5**, we compare this approximation with simulations.

The correlation between the membrane potentials has *positive* contributions from the correlation between the excitatory inputs ($\rho_{E_1 E_2}$), and between the inhibitory inputs ($\rho_{I_1 I_2}$). Contributions coming from excitatory–inhibitory correlations ($\rho_{E_1 I_2}$ and $\rho_{E_2 I_1}$) are negative, and can thus decorrelate the activity of downstream cells. This "cancellation" of correlations is observed in **Figures 1D and 5**, and can lead to asynchrony in recurrent networks (Hertz, 2010; Renart et al., 2010).

IMPLICATIONS FOR SYNCHRONIZATION IN FEEDFORWARD CHAINS

Feedforward chains, like that depicted in **Figure 6A**, have been studied extensively (Diesmann et al., 1999; van Rossum et al., 2002; Litvak et al., 2003; Reyes, 2003; Tetzlaff et al., 2003; Câteau and Reyes, 2006; Doiron et al., 2006; Kumar et al., 2008). In such networks, cells in a layer necessarily share some of their inputs, leading to correlations in their spiking activity (Shadlen and Newsome, 1998). Frequently, spiking in deeper layers is highly synchronous (Reyes, 2003; Tetzlaff et al., 2003). However, in the presence of background noise, correlations can remain negligible (van Rossum et al., 2002; Vogels and Abbott, 2005).

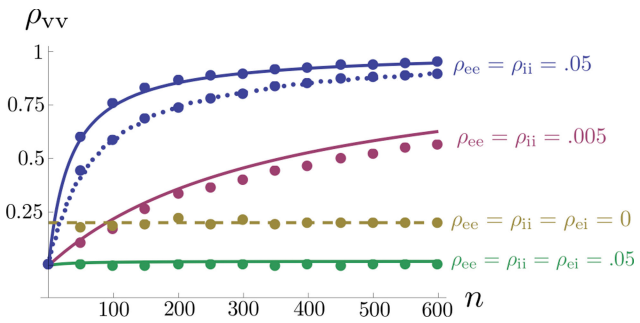


FIGURE 5 | The effects of pooling on correlations between postsynaptic membrane potentials. Results of the linear approximation (solid, dotted, and dashed lines) match simulations (points). For the solid blue line, $\rho_{ee} = \rho_{ii} = 0.05$, and $\rho_{ei} = p_e = p_i = q_e = q_i = 0$. The total number of excitatory and inhibitory inputs to each cell was $n = n_e + q_e n_e$, and $n_i + q_i n_i$ respectively. Here $n_i = n_e/3$, with other parameters given in the Section "Materials and Methods." The dotted blue line was obtained by including independent inputs, $q_e = q_i = 1$. The pink line was obtained by decreasing input correlations to $\rho_{ee} = \rho_{ii} = 0.005$. The solid green line was obtained by including excitatory–inhibitory correlations, $\rho_{ei} = 0.05$, so that total input correlations canceled. The dashed tan line was obtained by setting $\rho_{ee} = \rho_{ii} = \rho_{ei} = q_e = q_i = 0$ and $p_e = p_i = 0.2$ so that correlations are due to input overlap alone. In all cases, $\mathcal{E} = \frac{590}{n}$ nS·ms, and $\mathcal{I} = 4\mathcal{E}$. Standard errors are smaller than twice the radii of the points.

Feedforward chains amplify correlations as follows: When inputs to the network are independent, small correlations are introduced in the second layer by overlapping inputs. The inputs to each subsequent layer are pooled from the previous layer. The amplification of correlations by pooling is the primary mechanism for the development of synchrony (Compare solid and dotted blue lines in **Figure 4A**). Overlapping inputs serve primarily to "seed" synchrony in early layers. The internal dynamics of the neurons and background noise can decorrelate the output of a layer, and compete with the correlation amplification due to pooling.

We develop this explanation by considering a feedforward network with each layer containing N_e excitatory and N_i inhibitory cells. Each cell in layer $k+1$ receives n_e excitatory and n_i inhibitory inputs selected randomly from layer k. For simplicity we assume that all excitatory and inhibitory cells are dynamically identical and $\mathcal{E}|V_E - V_L| = \mathcal{I}|V_E - V_L|$. Spike trains driving the first layer are statistically homogeneous with pairwise correlations ρ_0.

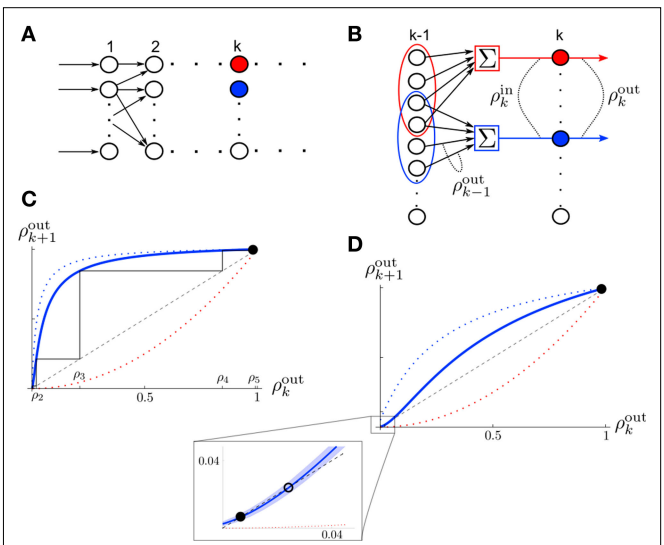


FIGURE 6 | The development of synchrony in a feedforward chain can be understood using a model dynamical system (Tetzlaff et al., 2003). **(A)** Schematic diagram of the network. Each layer consists of N_e excitatory and N_i inhibitory cells. Each cell in layer k receives precisely n_e excitatory and n_i inhibitory, randomly selected inputs from layer $k-1$. **(B)** Stages of processing in the feedforward network. Inputs from layer $k-1$ are pooled with overlap, and drive the cells in layer k. **(C)** The correlation transfer map described by the pooling function, $P(\rho)$ (blue dotted line), is composed with the decorrelating transfer function, $S(\rho) = \rho^2$ (red dotted line), to obtain the mapping, $T = S \circ P$ (solid blue line). Cobwebs show the development of correlations in the discrete dynamical system defined by $\rho_{k+1}^{out} = T(\rho_k^{out})$ with $\rho_0 = 0$. Nearly perfect correlations develop by the fifth layer. The identity is shown as a dashed line. **(D)** Closer to balance ($\beta \approx 1$), the correlating effects of pooling are weakened, and the model develops a stable fixed point close to $\rho = 0$. However, cells may no longer decorrelate their inputs in the balanced regime, and fluctuations in the input statistics due to random connectivity can destabilize the fixed point and lead to synchrony. The shaded region in the inset represents the region two standard deviations away from the mean (blue line) when randomness in the overlap is taken into account (see Appendix). The standard deviations were calculated using Monte Carlo simulations. In C and D, $N_e = 12000$ and $n_e = 600$. In C, $N_i = 8000$ and $n_i = 400$. In D, $N_i = 10500$ and $n_i = 525$ to obtain approximate balance ($\beta = 600/525$). Filled black circles represent stable fixed points and open black circles represent unstable fixed points.

To explain the development of correlations, we consider a simplified model of correlation propagation (See also Renart et al., 2010 for a recurrent version). In the model, any two cells in a layer share the expected proportion $p_e = n_e/N_e$ of their excitatory inputs and $p_i = n_i/N_i$ of their inhibitory inputs (the expected proportions are taken with respect to random connectivity). We also assume that inputs are statistically identical across a layer.

For a pair of cells in layer $k \geq 1$, let ρ_k^{in} and ρ_k^{out} represent the correlation coefficient between the total input currents and output spike trains respectively. The outputs from layer k are pooled (with overlap) to obtain the inputs to layer $k+1$. Using the results developed above, $\rho_{in}^1 = P(\rho_0)$ and $\rho_{k+1}^{in} = P(\rho_k^{out})$, for $k \geq 1$, where (see Appendix and Tetzlaff et al., 2003 for a similar derivation)

$$P(\rho) = \frac{\rho(\beta-1)^2 + \frac{1}{n_i}(1-\rho)\left(\beta p_e + p_i\right)}{\rho(\beta-1)^2 + \frac{1}{n_i}(1-\rho)(1+\beta)}. \quad (14)$$

Here β measures the balance between excitation and inhibition (see Materials and Methods). From our assumptions, $\beta = n_e/n_i$. With imbalance ($\beta \neq 1$) and a large number of cells in a layer, pooling amplifies small correlations, $P(\rho) > \rho$, as discussed earlier.

To complete the description of correlation transfer from layer to layer, we relate the correlations between inputs to a pair of cells, ρ_k^{in}, to correlations in their output spike trains, ρ_k^{out}. We assume that there is a transfer function, S, so that $\rho_k^{out} = S(\rho_k^{in})$ at each layer k. We additionally assume that $S(0) = 0$ and $S(1) = 1$, that is uncorrelated (perfectly correlated) inputs result in uncorrelated (perfectly correlated) outputs. We also assume that the cells are decorrelating, $|\rho| > |S(\rho)| > 0$ for $\rho \neq 0,1$ (Shea-Brown et al., 2008). This is an idealized model of correlation transfer, as output correlations depend on cell dynamics and higher order statistics of the inputs (Moreno-Bote and Parga, 2006; de la Rocha et al., 2007; Barreiro et al., 2009; Ostojić et al., 2009).

Correlations between the spiking activity of cells in layers $k+1$ are related to correlations in layer k by the layer-to-layer transfer function, $T = S \circ P$. The development of correlations across layers is modeled by the dynamical system, $\rho_{k+1}^{out} = T(\rho_k^{out})$, with $\rho_1^{out} = S(\rho_0)$.

When the network is not balanced ($\beta \neq 1$), pooling amplifies correlations at each layer and the activity between cells in deeper layers can become highly correlated (see **Figure 6C**). The output of the first layer is uncorrelated if the individual inputs are independent ($\rho_0 = 0$). In this case all of the correlations between the total inputs to the second layer come from shared inputs,

$$\rho_2^{in} = P(0) = \frac{n_e p_e + n_i p_i}{n_e + n_i}.$$

These correlations are then reduced by the second layer of cells, $\rho_2^{out} = S(\rho_2^{in}) = T(0) > 0$, and subsequently amplified by pooling and input sharing before being received by layer 3, $\rho_3^{in} = P(\rho_2^{out})$. This process continues in subsequent layers. If the correlating effects of pooling and input sharing dominate the decorrelating effects of internal cell dynamics, correlations will increase from layer to layer (see **Figure 6C**).

When $\rho_0 = 0$, overlapping inputs increase the input correlation to layer 2, but have a negligible effect on the mapping once correlations have developed since the effects of pooling dominate [see Eq. (14) and the dashed blue line in **Figure 4A** which shows that the effects of input overlaps are small when n_e is large, $\rho > 0$ and $\beta \neq 1$]. Therefore, shared inputs seed correlated activity at the first layer, and pooling drives the development of larger correlations. When $\rho_0 = 0$, we cannot expect large correlations before layer 3, but when $\rho_0 > 0$ large correlations can develop by layer 2.

To verify this conclusion, we constructed a two-layer feedforward network with no overlap between inputs ($P_e = P_i = 0$). In **Figure 7A**, the inputs to layer 1 were independent ($\rho_0 = 0$), and the firing of cells in layer 2 was uncorrelated. In **Figure 7B**, we introduced small correlations ($\rho_0 = 0.05$) between inputs to layer 1. These correlations were amplified by pooling so that strong synchrony is observed between cells in layer 2. We compared these results with a standard feedforward network with overlap in cell inputs (**Figure 7C**, where $P_e = P_i = 0.05$). Inputs to layer 1 were independent ($\rho_0 = 0$), and hence outputs from layer 1 uncorrelated. Dependencies between inputs to layer 2 were weak and due to overlap alone, $\rho_2^{in} = P(0) = 0.05$. Cells in layer 3 received pooled inputs from layer 2, and their output was highly correlated.

These results predict that correlations between spike trains develop in deeper layers, but they do not not directly address the timescale of the correlated behavior. In simulations, spiking becomes tightly synchronized in deeper layers (see for instance Litvak et al., 2003; Reyes, 2003; and **Figure 7**). This can be understood using results in Maršálek et al. (1997) and Diesmann et al. (1999) where it is shown that the response of cells to volleys of spikes is tighter than the volley itself. The firing of individual cells in the network becomes bursty in deeper layers and large correlations are manifested in tightly synchronized spiking events. Alternatively, one can predict the emergence of synchrony by observing that pooling increases correlations over finite time windows (see next section and Appendix) and therefore the analysis developed above can be adapted to correlations over small windows.

Balanced feedforward networks

In the simplified feedforward model above, when excitation balances inhibition, that is $\beta \approx 1$, correlations between the pooled inputs to a layer are due to overlap alone, $\rho_k^{in} = P(\rho_{k-1}^{out}) \approx (p_e + p_i)/2$ for all k. The correlating effects of this map are weak, and this would seem to imply that cells in balanced feedforward chains remain asynchronous. Indeed, our model of correlation propagation displays a stable fixed point at low values of ρ when $\beta \approx 1$ (see **Figure 6D**). However, in practice, synchrony is difficult to avoid without careful fine-tuning (Tetzlaff et al., 2003), and almost always develops in feedforward chains (Litvak et al., 2003). We provide some reasons for this discrepancy.

Our focus so far has been on correlations over infinitely large time windows (see Materials and Methods where we define ρ_{xy}). Even when the membrane potentials are nearly uncorrelated over large time windows, differences between the excitatory and inhibitory synaptic time constants can cause larger correlations over smaller time windows (Renart et al., 2010). This can, in turn, lead to

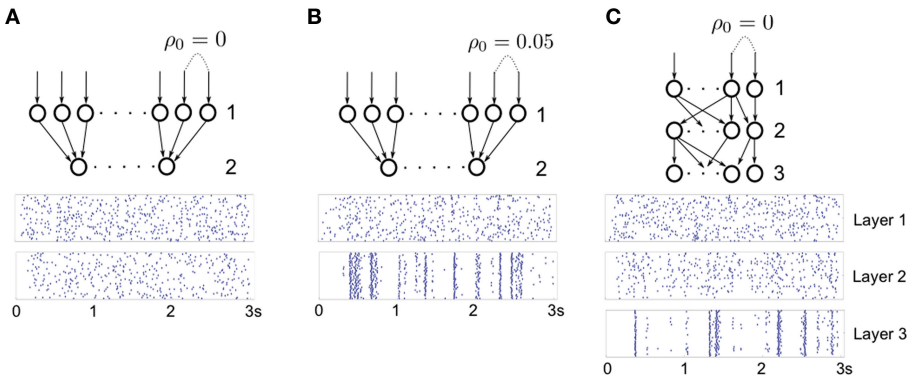


FIGURE 7 | Development of synchrony in feedforward networks. (A) A feedforward network with no overlap and independent, Poisson input. For excitatory cells, we set $\mathcal{E} \approx 1.55$nS·ms, and $\mathcal{I} \approx 4.67$nS·ms. For inhibitory cells, $\mathcal{E} \approx 3.61$nS·ms, and $\mathcal{I} \approx 10.82$nS·ms. **(B)** Same as A, except inputs to layer 1 are correlated with coefficient $\rho_0 = 0.05$. The network is highly synchronized in the second layer, even though inputs do not overlap. **(C)** Same as A, except for the presence of overlapping inputs ($\rho_e = p_i = 0.05$). Correlations due to overlap in the input to layer 2 result in average correlations of 0.05 between input currents. Layer 3 cells in C synchronize (Compare with layer 2 in B). In all three figures, each cell in the first layer was driven by excitatory Poisson inputs with rate $\nu_0 = 100$ Hz.

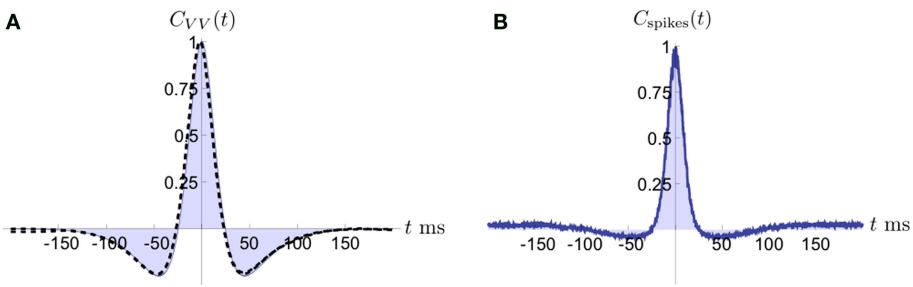


FIGURE 8 | Cross-covariance functions between membrane potentials and output spike trains. (A) The cross-covariance function between membrane potentials, scaled so that its maximum is 1. The linear approximation in Eq. (16) (blue, shaded) agrees with simulations of the full conductance-based model (black dashed line). Differences between simulations with and without threshold are too small to be observable (8000 simulations 10s each; simulations with and without threshold are shown). Parameters are as in **Figure 1D.** The cells are balanced with $\rho_{ee} = \rho_{ii} = \rho_{ei} = 0.05$ so that the correlation between the membrane potentials over long time windows is essentially zero ($\rho_{vv} = 0.0085 \pm 0.0024$ s.e. unthresholded, and $\rho_{vv} = 0.0174 \pm 0.0024$ s.e. thresholded). However, correlations over shorter time windows are positive as indicated by the central peak in the cross-covariance function. **(B)** The cross-covariance between the output spike trains is mostly positive. The correlation between the output spike trains was $\rho_{\text{spikes}} = 0.1570 \pm 0.0033$ s.e. (500 simulations of 100s each with same parameters as in A.

significant correlations between the output spike trains. We discuss this effect further in the Appendix and give an example in **Figure 8**. In this example, the correlations between the membrane potentials over long windows are nearly zero due to cancellation (see **Figure 8A** where $\rho_{VV} = 0.0174 \pm 0.0024$ s.e. with threshold present), but positive over shorter timescales. The cross-covariance function between the output spike trains is primarily positive, yielding significant spike train correlations ($\rho_{\text{spikes}} = 0.1570 \pm 0.0033$ s.e.). Therefore, the assumption that pairs of cells decorrelate their inputs may not be valid in the balanced case.

Another source of discrepancies between the idealized model and simulations of feedforward networks are inhomogeneities, which become important when balance is exact. Note that Eq. (14) is an approximation obtained by ignoring fluctuations in connectivity from layer to layer. In a random network, inhomogeneities will be introduced by variability in input population overlaps. To fully describe the development of correlations in a feedforward network, it is necessary to include such fluctuations in a model of correlation propagation. The asynchronous fixed point that appears in the balanced case has a small basin of attraction and fluctuations induced by input inhomogeneities could destroy its stability (see **Figure 6D**). Other sources of heterogeneity can further destabilize the asynchronous state (see Appendix).

It has been shown that asynchronous states can be stabilized through the decorrelating effects of background noise (van Rossum et al., 2002; Vogels and Abbott, 2005). To emulate these effects, a third transfer function, N, can be added to our model. The correlation transfer map then becomes $T(\rho) = S \circ N \circ P(\rho)$. Sufficiently strong background noise can increase decorrelation from input to output of a layer, and stabilize the asynchronous fixed point.

DISCUSSION

We have illustrated how pooling and shared inputs can impact correlations between the inputs and free membrane voltages of postsynaptic cells in a feedforward setting. The increase in correlation due to pooling was discussed in a simpler setting in (Bedenbaugh

and Gerstein, 1997; Super and Roelfsema, 2005; Chen et al., 2006; Stark et al., 2008), and similar ideas were also developed for the variance alone in (Salinas and Sejnowski, 2000; Moreno-Bote et al., 2008). The saturation of the signal-to-noise ratio with increasing population size observed in (Zohary et al., 1994) has a similar origin. Our aim was to present a unified discussion of these results, with several generalizations.

Other mechanisms, such as recurrent connectivity between cells receiving the inputs, can modulate correlated activity (Schneider et al., 2006; Ostojić et al., 2009). Importantly, the cancellation of correlations may be a dynamic phenomenon in recurrent networks, as observed in (Hertz, 2010; Renart et al., 2010). On the other hand, neurons may become entrained to network oscillations, resulting in more synchronous firing (Womelsdorf et al., 2007). A full understanding of the statistics of population activity in neuronal networks will require an understanding of how these mechanisms interact to shape the spatiotemporal properties of the neural response.

The results we presented relied on the assumption of linearity at the different levels of input integration. These assumptions can be expected to hold at least approximately. For instance, there is evidence that membrane conductances are tuned to produce a linear response in the subthreshold regime (Morel and Levy, 2009). The assumptions we make are likely to break down at the level of single dendrites where nonlinear effects may be much stronger (Johnston and Narayanan, 2008). The effects of correlated inputs to a single dendritic branch deserve further theoretical study (Gasparini and Magee, 2006; Li and Ascoli, 2006).

We demonstrated that the structure of correlations in a population may be difficult to infer from pooled activity. For instance, a change in pairwise correlations between individual cells in two populations causes a much smaller change in the correlation between the pooled signals. With a large number of inputs, the change in correlations between the pooled signals might not be detectable even when the change in the pairwise correlations is significant.

While we discussed the growth of second order correlations only, higher order correlations also saturate with increasing population size. For example, in a 3-variable generalization of the homogeneous model from **Figure 3A**, it can be shown that $\rho_{E_1E_2E_3} = 1 - \mathcal{O}(1/n_e)$ where n_e is the size of each population and $\rho_{E_1E_2E_3}$ is the triple correlation coefficient (Stratonovich, 1963) between the pooled signals E_1, E_2, and E_3. The reason that higher order correlations also saturate follows from the generalization of the following observation at second order: Pooling amplifies correlations because the variance and covariance grow asymptotically with the same rate in n_e. In particular σ^2_E and $\gamma_{E_1E_2}$ both behave asymptotically like $n_e^2\rho_{ee}\sigma_e^2 + \mathcal{O}(n_e)$, and their ratio, $\rho_{E_1E_2} = \gamma_{E_1E_2}/\sigma^2_E$, approaches unity (Bedenbaugh and Gerstein, 1997; Salinas and Sejnowski, 2000; Moreno-Bote et al., 2008).

We concentrated on correlations over infinitely long time windows (see Materials and Methods where we define ρ_{xy}). However, pooling amplifies correlations over finite time windows in exactly the same way as correlations over large time windows. Due to the filtering properties of the cells, the timescale of correlations between downstream membrane potentials may not reflect that of the inputs. We discuss this further in the Appendix where the auto- and cross-covariance functions between the membrane potentials are derived.

To simplify the presentation, we have so far assumed stationary. However, since Eq. (2) applies to the Pearson correlation between any pooled data, all of the results on pooling can easily be extended to the non-stationary case. In the non-stationary setting, the cross-covariance function has the form $R_{xy}(s, t) = \mathrm{cov}(x(s), y(s+t))$, but there is no natural generalization of the asymptotic statistics defined in Eq. (1).

Correlated neural activity has been observed in a variety of neural populations (Gawne and Richmond, 1993; Zohary et al., 1994; Vaadia et al., 1995), and has been implicated in the propagation and processing of information (Oram et al., 1998; Maynard et al., 1999; Romo et al., 2003; Tiesinga et al., 2004; Womelsdorf et al., 2007; Stark et al., 2008), and attention (Steinmetz et al., 2000; Mitchell et al., 2009). However, correlations can also introduce redundancy and decrease the efficiency with which networks of neurons represent information (Zohary et al., 1994; Gutnisky and Dragoi, 2008; Goard and Dan, 2009). Since the joint response of cells and recorded signals can reflect the activity of large neuronal populations, it will be important to understand the effects of pooling to understand the neural code (Chen et al., 2006).

APPENDIX

DERIVATION OF EQ. (10)

Equation (10) can be derived from Eq. (2). However, we find that it is more easily derived directly. We will calculate the variance, $\sigma^2_{E_1} = \sigma^2_{E_2}$, and covariance $\gamma_{E_1E_2}$ between the pooled signals.

The covariance is given by the sum of all pairwise covariances between the populations, $\gamma_{E_1E_2} = \Sigma_{e_1\in E_1, e_2\in E_2}\sigma_{e_1}\sigma_{e_2}\rho_{e_1e_2}$. Each cell receives $n_e + q_e n_e$ inputs so that there are $(n_e + q_e n_e)^2$ terms that appear in this sum. However, the $q_e n_e$ "independent" inputs from each pool are uncorrelated with all other inputs and therefore don't contribute to the sum. Of the remaining n_e^2 pairs, $n_e p_e$ are shared and therefore have correlation $\rho_{e_1e_2} = 1$. These shared processes therefore collectively contribute $n_e p_e \sigma_e^2$ to $\gamma_{E_1E_2}$. The remaining $n_e^2 - n_e p_e$ processes are correlated with coefficient ρ_{ee} and collectively contribute $(n_e^2 - n_e p_e)\rho_{ee}\sigma_e^2$. The pooled covariance is thus

$$\gamma_{E_1E_2} = \underbrace{(n_e^2 - p_e n_e)\rho_{ee}\sigma_e^2}_{\text{Correlated}} + \underbrace{n_e p_e \sigma_e^2}_{\text{Shared}}.$$

The variance is given by the sum of all pairwise covariances within a population, $\sigma^2_{E_1} = \Sigma_{e_1\in E_1, e_2\in E_1}\sigma_{e_1}\sigma_{e_2}\rho_{e_1e_2}$. As above, there are $n_e + q_e n_e$ neurons in the population, so that the sum has $(n_e + q_e n_e)^2$ terms. Of these, $n_e + q_e n_e$ are "diagonal" terms ($e_1 = e_2$), each contributing σ_e^2, for a total contribution of $(n_e + q_e n_e)\sigma_e^2$ to $\sigma^2_{E_1}$. The processes from the independent pool do not contribute any additional terms. This leaves $n_e(n_e - 1)$ correlated pairs which each contribute $\sigma_e^2\rho_{ee}$ for a collective contribution of $n_e(n_e - 1)\sigma_e^2\rho_{ee}$, giving

$$\sigma_{E_1} = \sigma_{E_2} = \underbrace{n_e(n_e - 1)\sigma_e^2\rho_{ee}}_{\text{Correlated}} + \underbrace{(n_e + n_e q_e)\sigma_e^2}_{\text{Diagonal}}.$$

Now, $\rho_{E_1E_2} = \gamma_{E_1E_2}/\sigma_{E_1}$ can be simplified to give Eq. (10). Equations for $\rho_{I_1I_2}$ and $\rho_{E_1I_2} = \rho_{I_1E_2}$ can be derived identically.

FINITE-TIME CORRELATIONS AND CROSS-COVARIANCES

Throughout the text, we concentrated on correlations over large time windows. However, the effects of pooling described by Eq. (2) apply to the correlation, $\rho_{xy}(t)$, between spike counts over any time window of

size t, defined by $\rho_{xy}(t) = \mathrm{cov}(N_x(t), N_y(t))/\sqrt{\mathrm{var}(N_x(t))\,\mathrm{var}(N_y(t))}$ where $N_x(t) = \int_0^t x(s)ds$ is the spike count over $[0,t]$ for the spike train $x(t)$. The equation also applies to the instantaneous correlation at time t, defined by $R_{xy}(t) = C_{xy}(t)/\sqrt{C_{xx}(t)C_{yy}(t)}$. Thus pooling increases correlations over all timescales equally.

However, the cell filters the pooled inputs to obtain the membrane potentials and, as a result, the correlations between membrane potentials is "spread out" in time (Tetzlaff et al., 2008). To quantify this effect, we derive an approximation to the auto- and cross-covariance functions between the membrane potentials.

The pooled input spike trains are obtained from from a weighted sum of the individual excitatory and inhibitory spike trains (see Materials and Methods). As a result cross-covariance functions between the pooled spike trains are just sums of the individual cross-covariance functions, $C_{XY}(t) = \Sigma_{x\in X, y\in Y} C_{xy}(t)$ for X, $Y = \mathrm{E}_1, \mathrm{E}_2, \mathrm{I}_1, \mathrm{I}_2$ and x, $y = \mathrm{e}$, i accordingly. Thus only the magnitude of the cross-covariance functions is affected by pooling. The change in magnitude is quadratic in n_e or n_i. This is consistent with the observation that pooling amplifies correlations equally over all timescales.

The conductances are obtained by convolving the total inputs with the synaptic filter kernels,

$$g_{\mathrm{E}_k}(t) = (\mathrm{E}_k * \alpha_\mathrm{e})(t), \quad \text{and} \quad g_{\mathrm{I}_k}(t) = (\mathrm{I}_k * \alpha_i)(t); \quad k = 1,2.$$

The cross-covariance between the conductances can therefore be written as a convolution of the cross-covariance function between the input signals and the deterministic cross-covariance between the synaptic kernels (Tetzlaff et al., 2008). In particular,

$$C_{g_x g_Y}(t) = (C_{XY} * (\alpha_x \star \alpha_y))(t) \tag{15}$$

for X, $Y = \mathrm{E}_1, \mathrm{E}_2, \mathrm{I}_1, \mathrm{I}_2$ and x, $y = \mathrm{e}$, i accordingly, where $(\alpha_x \star \alpha_y)(t) = \int_{-\infty}^{\infty} \alpha_x(s)\alpha_y(t+s)\mathrm{d}s$ is the deterministic cross-covariance between the synaptic filters, α_x and α_y. Note that total correlations remain unchanged by convolution of the input spike trains with the synaptic filters, since the integral of a convolution will be equal to the product of the integrals (Tetzlaff et al., 2008).

The total input currents, $J_K(t) = -(g_{E_k}(t)(V_L - V_I))/C_m$, obtained from the linearization of the conductance-based model described in the Section "Materials and Methods" are simply linear combinations of the individual conductances. The cross-covariance function between the input currents is therefore a linear combination of those between the conductances,

$$\begin{aligned} C_{J_h J_k}(t) = {} & |V_\mathrm{E} - V_L|^2 C_{g_{\mathrm{E}_h} g_{\mathrm{E}_k}}(t) + |V_\mathrm{I} - V_L|^2 C_{g_{\mathrm{I}_h} g_{\mathrm{I}_k}}(t) \\ & - 2|V_\mathrm{E} - V_L||V_\mathrm{I} - V_L| C_{g_{\mathrm{I}_h} g_{\mathrm{E}_k}}(t). \end{aligned}$$

Combining this result with Eq. (1), yields the correlation, $\rho_{\mathrm{in}} = \rho_{J_1 J_2}$, between the total input currents given in Eq. (13).

Using the solution of the linearized equations described in the Section "Materials and Methods", we obtain a linear approximation to the cross-covariance functions,

$$C_{V_h V_k}(t) \approx \frac{1}{2\tau_\mathrm{eff}} (C_{J_h J_k} * (K \star K))(t) \tag{16}$$

for $h,k = 1,2$ where $(K \star K)(t) = \frac{\tau_\mathrm{eff}}{2} e^{-|t|/\tau_\mathrm{eff}}$ is the cross-covariance between the linear kernel, K, and itself. The convolution with $(K \star K)(t)$ scales the area of both the auto- and cross-covariance functions by a factor of τ_eff^2, and therefore leaves the ratio of the areas, $\rho_{V_1 V_2}$ unchanged. Thus, the linear approximation predicts that $\rho_{V_1 V_2} \approx \rho_{\mathrm{in}}$.

When the total inputs are strong, τ_eff is small and we can simplify Eq. (16) by approximating $(K \star K)(t)$ with a delta function with mass τ_eff^2 so that $C_{V_1 V_2}(t) \approx \tau_\mathrm{eff} C_{J_1 J_2}(t)/2$ and similarly for $C_{V_1 V_1}(t)$. This approximation is valid when the synaptic time constants are significantly larger than τ_eff, which is likely to hold in high conductance states. We compare this approximation to cross-covariance functions obtained from simulations in **Figure 8**.

In all examples considered, the cross-covariance functions have exponentially decaying tails. We define the correlation time constant, $\tau_{xy} = \lim_{t\to\infty} -t/\ln(C_{xy}(t))$, as a measure of the decay rate of the exponential tail. If $t \gg \tau_{xy}$, then $x(s)$ and $y(s+t)$ can be regarded as approximately uncorrelated and $\gamma_{xy} \approx \int_{-t}^{t} (t - |s|/t) C_{xy}(s)\mathrm{d}s$ (Stratonovich, 1963).

The time constant of a convolution between two exponentially decaying functions is just the maximum time constant of the two functions. Thus, from the results above, the correlation time constant between the membrane potentials is the maximum of the correlation time constants between the inputs, the synaptic time constants, and the effective membrane time constant $\tau_{V_1 V_2} = \max\{\tau_{\mathrm{E}_1\mathrm{E}_2}, \tau_{\mathrm{I}_1\mathrm{I}_2}, \tau_{\mathrm{E}_1\mathrm{I}_2}, \tau_{\mathrm{E}_2\mathrm{I}_1}, \tau_\mathrm{e}, \tau_\mathrm{i}, \tau_\mathrm{eff}\}$ where $\tau_{\mathrm{E}_1\mathrm{E}_2}$, $\tau_{\mathrm{I}_1\mathrm{I}_2}$, $\tau_{\mathrm{E}_1\mathrm{I}_2}$, and $\tau_{\mathrm{E}_2\mathrm{I}_1}$ are the time constants of the input spike trains and τ_e and τ_i are synaptic time constants. Thus the cross-covariances functions between the membrane potentials are generally broader than the cross-covariance functions between the spike train inputs.

DERIVATION OF EQ. (14)

Consider a feedforward network where each layer consists of N_e excitatory cells and N_i inhibitory cells; each cell in layer k receives n_e excitatory and n_i inhibitory inputs from layer $(k-1)$, and these connections are chosen randomly and independently across neurons in layer k. Then the degree of overlap in the excitatory and inhibitory inputs to a pair of cells in layer k is a random variable. Following the derivation in Derivation of Eq. (10) in Appendix,

$$\rho_{\mathrm{E}_1\mathrm{E}_2} = \frac{\gamma_{\mathrm{E}_1\mathrm{E}_2}}{\sigma_\mathrm{E}^2} = \frac{s_\mathrm{e}\sigma_\mathrm{e}^2 + (n_\mathrm{e}^2 - s_\mathrm{e})\rho_{\mathrm{ee}}\sigma_\mathrm{e}^2}{n_\mathrm{e}\sigma_\mathrm{e}^2 + (n_\mathrm{e}^2 - n_\mathrm{e})\rho_{\mathrm{ee}}\sigma_\mathrm{e}^2},$$

where s_e denotes the number of common excitatory inputs between the two cells. To understand the origin of s_e, suppose the n_e excitatory inputs to cell 1 have been selected. Then the selection of the n_e excitatory inputs to cell 2 involves choosing, without replacement, from two pools: the first, of size n_e, projects to cell 1, and the second, of size $(N_\mathrm{e} - n_\mathrm{e})$, does not. Therefore, s_e is follows a hyper-geometric distribution with parameters $(N_\mathrm{e}, n_\mathrm{e}, n_\mathrm{e})$, and has mean $n_\mathrm{e}^2/N_\mathrm{e} = n_\mathrm{e} p_\mathrm{e}$. In addition, this random variable is independently selected amongst each pair in layer k. Using the mean value of s_e, we obtain Eq. (10).

For simplicity, we assume that $\mathcal{E}|V_\mathrm{E} - V_L| = \mathcal{I}|V_\mathrm{I} - V_L|$, so that $\beta = n_\mathrm{i}/n_\mathrm{e}$. If we assume that the statistics in the $(k-1)^{\mathrm{st}}$ layer are uniform across all cells and cell types (i.e., $\rho = \rho_{\mathrm{ee}} = \rho_{\mathrm{ii}} = \rho_{\mathrm{ei}} = \rho_{k-1}^{\mathrm{out}}$ and $\sigma_\mathrm{e} = \sigma_\mathrm{i}$) then by substituting Eq. (10) and the equivalent forms for ρ_{II}, ρ_{EI} in to Eq. (13), we may write the input correlations to the k^{th} layer as

$$\rho_{\mathrm{in}} = \frac{|V_\mathrm{E} - V_L|^2 \gamma_{\mathrm{E}_1\mathrm{E}_2} + |V_\mathrm{I} - V_L|^2 \gamma_{\mathrm{I}_1\mathrm{I}_2} - 2|V_\mathrm{E} - V_L||V_\mathrm{I} - V_L| \gamma_{\mathrm{E}_1\mathrm{I}_2}}{|V_\mathrm{E} - V_L|^2 \sigma_\mathrm{E}^2 + |V_\mathrm{I} - V_L|^2 \sigma_\mathrm{I}^2 - 2|V_\mathrm{E} - V_L||V_\mathrm{I} - V_L| \gamma_{\mathrm{E}_1\mathrm{I}_1}}.$$

Substituting the values of the covariances and variances, and dividing the numerator and denominator by $(\mathcal{E}|V_E - V_L|\sigma_e)^2$, we get

$$\rho_{in} = \frac{s_e + (n_e^2 - s_e)\rho + s_i + (n_i^2 - s_i)\rho - 2n_e n_i \rho}{n_e + (n_e^2 - n_e)\rho + n_i + (n_i^2 - n_i)\rho - 2n_e n_i \rho}.$$

Rearranging terms and dividing numerator and denominator by n_i^2, along with the substitution $\beta = n_e/n_i$, we have

$$\rho_{in} = \frac{\rho(1-\beta)^2 + \frac{1}{n_i^2}(s_e + s_i)(1-\rho)}{\rho(1-\beta)^2 + \frac{1}{n_i^2}(n_e + n_i)(1-\rho)} \quad (17)$$

This equation takes into account the variations in overlap due to finite size effects since s_e and s_i are random variables. Eq. (14) in the text represents the expected value $P(\rho) = \langle \rho_{in} \rangle$ which can be obtained by replacing the variables s_e and s_i in Eq. (17) with their respective means, $\langle s_e \rangle = n_e p_e$ and $\langle s_i \rangle = n_i p_i$. The expectation above is taken over realizations of the random connectivity of the feedforward network.

To calculate the standard deviation for the inset in **Figure 6D**, we ran Monte Carlo simulations, drawing s_e and s_i from a hypergeometric distribution and calculating the resulting transfer, $S(\rho_{in}) = \rho_{in}^2$ using Eq. (17). Note, however, that Eq. (17) and the inset in **Figure 6D**, do not account for all of the effects of randomness which may destabilize the balanced network. In deriving Eq. (17), we assumed that the statistics in the second layer were uniform. However, variations in the degree of overlap in one layer will cause inhomogeneities in the variances and rates at the next layer. In a feedforward setting, these inhomogeneities are compounded at each layer to destabilize the asynchronous fixed point.

DEFINITIONS AND VALUES OF VARIABLES USED IN THE TEXT

Table 1 | Definitions of variables pertaining to recordings.

$X_1(t)$, $X_2(t)$	Signals from two populations.
$\rho_{X_1X_2}$	Correlation between the signals.
$\bar{\rho}_{jk}$	Average pairwise correlation between a cell in population *j* and a cell in population *k*.
$r(d)$	Firing rate of a cell at distance d from the center of a stimulus.

Table 2 | Definitions of variables pertaining to downstream cells.
Subscripts e and E (i and I) denote excitation (inhibition).

n_e, n_i	Number of correlated inputs to a cell.
ν_e, ν_i	Input rates.
$\mathcal{E}$, $\mathcal{I}$	Synaptic weights.
q_e, q_i	Neurons received $n_e q_e (n_i q_i)$ independent excitatory (inhibitory) inputs.
ρ_{ee}, ρ_{ii}, ρ_{ei}	Correlations between pairs of afferents.
ρ_{xy}^w, ρ_{xy}^b	Correlations within or between two non-overlapping populations (x,y = e,i).
p_e, p_i	Proportion of shared input to the post-synaptic pair.
N_e, N_i	Number of cells per layer in feed-forward network model.
E_j, I_j	Pooled input spike trains to cell *j*.
σ_E, σ_I	Standard deviation of pooled excitatory or inhibitory spike trains.
$\gamma_{E_jE_k}$, $\gamma_{E_jI_k}$, $\gamma_{I_jI_k}$	Covariance between pooled spike trains.
$\rho_{E_jE_k}$, $\rho_{E_jI_k}$, $\rho_{I_jI_k}$	Correlations between pooled spike trains.
β	Measure of balance between excitation and inhibition in the inputs.
ρ_k^{in}, ρ_k^{out}	Correlations between inputs to or outputs from cell pairs in a feedforward network.
$C_{XY}(t)$	Cross-covariance function between processes *X* and *Y*.
$P(\rho)$	Correlations between the pooled inputs to cells in the feedforward model.
$S(\rho)$	Correlation between output spike trains in terms of input current correlations between cell pairs in the feedforward model.

Table 3 | Parameter values for simulations of two downstream cells. For fields with "var," various values of the indicated parameters were used and are described in the captions. For all simulations, V_L = –60 mV, V_E = 0 mV, V_I = –90 mV, C_m = 114 pF, g_L = 4.086 nS, τ_e = 10 ms, τ_i = 20 ms.

	ρ_{ee}, ρ_{ii}	ρ_{ei}	n_e, n_i	p_e, p_i	q_e, q_i	ν_e, ν_i (Hz)	$\mathcal{E},\mathcal{I}$ (nS·ms)
Figure 1C	0.05	0	250, 84	0	1	5, 7.5	2.3, 9.2
Figure 1D	0.05	0.05	250, 84	0	1	5	2.3, 13.8
Figure 5	var	var	var, n_e/3	var	var	5, 7.5	590/n_e, 4$\mathcal{E}$
Figure 8	0.05	0.05	250, 84	0	1	5	2.3, 13.8

Table 4 | Parameter values for simulations of feedforward networks. The parameter ν_0 is the input rate to the first layer, $(\mathcal{E}, \mathcal{I})_e$ indicates synaptic weights for excitatory cells, and $(\mathcal{E}, \mathcal{I})_i$ for inhibitory cells. For all simulations, V_L = –60 mV, V_E = 0 mV, V_I = –90 mV, C_m = 114 pF, g_L = 4.086 nS, τ_e = 10 ms, τ_i = 20 ms. For **Figure 6,** theoretical values were obtained under the assumption that $\nu_e = \nu_i$ and $\mathcal{E}|V_E - V_L| = \mathcal{I}|V_I - V_L|$.

	n_e, n_i	N_e, N_i	p_e, p_i	ν_0 (Hz)	ρ_0	$(\mathcal{E},\mathcal{I})_e$, $(\mathcal{E},\mathcal{I})_i$ (nS·ms)
Figure 6C	600, 400	12000, 8000	0.05	NA	0	NA
Figure 6D	600, 525	12000, 10500	0.05	NA	0	NA
Figure 7A	225, 75	NA	0	100	0	(1.55,5.67),(3.61,10.82)
Figure 7B	225, 75	NA	0	100	0.05	(1.55,5.67),(3.61,10.82)
Figure 7C	225, 75	4500, 1500	0.05	100	0	(1.55,5.67),(3.61,10.82)

ACKNOWLEDGMENTS

We thank Jaime de la Rocha, Brent Doiron and Eric Shea-Brown for helpful discussions. We also thank the reviewers and the handling editor for numerous useful suggestions. This work was supported by NSF Grants DMS-0604429 and DMS-0817649 and a Texas ARP/ATP award.

REFERENCES

Barreiro, A., Shea-Brown, E., and Thilo, E. (2009). Timescales of spike-train correlation for neural oscillators with common drive. *Phys. Rev. E* 81, Arxiv preprint arXiv:0907.3924.

Bedenbaugh, P., and Gerstein, G. (1997). Multiunit normalized cross correlation differs from the average single-unit normalized correlation. *Neural. Comput.* 9, 1265–1275.

Câteau, H., and Reyes, A. (2006). Relation between single neuron and population spiking statistics and effects on network activity. *Phys. Rev. Lett.* 96, 58101.

Chen, Y., Geisler, W. S., and Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nat. Neurosci.* 9, 1412–1420.

Cohen, M., and Maunsell, J. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 12, 1594–1600.

Coombes, S., Timofeeva, Y., Svensson, C., Lord, G., Josić, K., Cox, S., and Colbert, C. (2007). Branching dendrites with resonant membrane: A Òsum-over-tripsÓ approach. *Biol. Cybern.* 97, 137–149.

de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* 448, 802–806.

Diesmann, M., Gewaltig, M., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402, 529–533.

Doiron, B., Rinzel, J., and Reyes, A. (2006). Stochastic synchronization in finite size spiking networks. *Phys. Rev. E* 74, 30903.

Ecker, A., Berens, P., Keliris, G., Bethge, M., Logothetis, N., and Tolias, A. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science* 327, 584–587.

Gasparini, S., and Magee, J. (2006). State-dependent dendritic computation in hippocampal CA1 pyramidal neurons. *J. Neurosci.* 26, 2088.

Gawne, T., and Richmond, B. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? *J. Neurosci.* 13, 2758.

Goard, M., and Dan, Y. (2009). Basal forebrain activation enhances cortical coding of natural scenes. *Nat. Neurosci.* 12, 1444–1449.

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

Gutnisky, D. A., and Josić, K. (2010). Generation of spatio-temporally correlated spike-trains and local-field potentials using a multivariate autoregressive process. *J. Neurophys.* doi:10.1152/jn.00518.2009.

Hertz, J. (2010). Cross-correlations in high-conductance states of a model cortical network. *Neural. Comput.* 22, 427–447.

Johnston, D., and Narayanan, R. (2008). Active dendrites: colorful wings of the mysterious butterflies. *Trends Neurosci.* 31, 309–316.

Kuhn, A., Aertsen, A., and Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Comput.* 15, 67–101.

Kuhn, A., Aertsen, A., and Rotter, S. (2004). Neuronal integration of synaptic input in the fluctuation-driven regime. *J. Neurosci.* 24, 2345.

Kumar, A., Rotter, S., and Aertsen, A. (2008). Conditions for propagating synchronous spiking and asynchronous firing rates in a cortical network model. *J. Neurosci.* 28, 5268.

Li, X., and Ascoli, G. A. (2006). Computational simulation of the input–output relationship in hippocampal pyramidal cells. *J. Comput. Neurosci.* 21, 191–209.

Litvak, V., Sompolinsky, H., Segev, I., and Abeles, M. (2003). On the transmission of rate code in long feedforward networks with excitatory-inhibitory balance. *J. Neurosci.* 23, 3006.

Maršálek, P., Koch, C., and Maunsell, J. (1997). On the relationship between synaptic input and spike output jitter in individual neurons. *Proc. Natl. Acad. Sci. U.S.A* 94, 735.

Maynard, E. M., Hatsopoulos, N. G., Ojakangas, C. L., Acuna, B. D., Sanes, J. N., Normann, R. A., and Donoghue, J. P. (1999). Neuronal interactions improve cortical population coding of movement direction. *J. Neurosci.* 19, 8083–8093.

Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area v4. *Neuron* 63, 879–888.

Morel, D., and Levy, W. (2009). The cost of linearization. *J. Comput. Neurosci.* 27, 259–275.

Moreno-Bote, R., and Parga, N. (2006). Auto-and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Phys. Rev. Lett.* 96, 28101.

Moreno-Bote, R., Renart, A., and Parga, N. (2008). Theory of input spike auto- and cross-correlations and their effect on the response of spiking neurons. *Neural Comput.* 20, 1651–1705.

Nunez, P., and Srinivasan, R. (2006). Electric Fields of the Brain: The Neurophysics of EEG. New York, NY: Oxford University Press.

Okun, M., and Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat. Neurosci.* 11, 535–537.

Oram, M. W., Földiák, P., Perrett, D. I., and Sengpiel, F. (1998). The 'ideal homunculus': decoding neural population signals. *Trends Neurosci.* 21, 259–265.

Ostojić, S., Brunel, N., and Hakim, V. (2009). How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J. Neurosci.* 29, 10234–10253.

Poort, J., and Roelfsema, P. (2009). Noise correlations have little influence on the coding of selective attention in area v1. *Cereb. Cortex* 19, 543.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590.

Reyes, A. (2003). Synchrony-dependent propagation of firing rate in iteratively constructed networks in vitro. *Nat. Neurosci.* 6, 593–599.

Roelfsema, P., Lamme, V., and Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nat. Neurosci.* 7, 982–991.

Romo, R., Hernández, A., Zainos, A., and Salinas, E. (2003). Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* 38, 649–657.

Salinas, E., and Sejnowski, T. J. (2000). Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J. Neurosci.* 20, 6193–209.

Schneider, A., Lewis, T., and Rinzel, J. (2006). Effects of correlated input and electrical coupling on synchrony in fast-spiking cell networks. *Neurocomputing* 69, 1125–1129.

Shadlen, M. N., and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* 18, 3870–96.

Shea-Brown, E., Josic´, K., de La ocha, J., and Doiron, B. (2008). Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Phys. Rev. Lett.* 100, 108102.

Smith, M. A., and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* 28, 12591–12603.

Stark, E., Globerson, A., Asher, I., and Abeles, M. (2008). Correlations between groups of premotor neurons carry information about prehension. *J. Neurosci.* 28, 10618–10630.

Steinmetz, P., Roy, A., Fitzgerald, P., Hsiao, S., Johnson, K., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* 404, 187–190.

Stratonovich, R. (1963). Topics in the Theory of Random Noise: General Theory of Random Processes. Nonlinear Transformations of Signals and Noise. New York, NY: Gordon and Breach.

Super, H., and Roelfsema, P. (2005). Chronic Multiunit Recordings in Behaving Animals: Advantages and Limitations. Development, dynamics and pathology of neuronal networks: from molecules to functional circuits: proceedings of the 23rd International Summer School of Brain Research, held at the Royal Netherlands Academy of Arts and Sciences, Amsterdam, from 25–29 August 2003, 263.

Tetzlaff, T., Buschermöhle, M., Geisel, T., and Diesmann, M. (2003). The spread of rate and correlation in stationary cortical networks. *Neurocomputing* 52, 949–954.

Tetzlaff, T., Rotter, S., Stark, E., Abeles, M., Aertsen, A., and Diesmann, M., (2008). Dependence of neuronal correlations on filter characteristics and marginal spike train statistics. *Neural Comput.* 20, 2133–2184.

Tiesinga, P. H., Fellous, J.-M., Salinas, E., José, J. V., and Sejnowski, T. J. (2004). Inhibitory synchrony as a mechanism for attentional gain modulation. *J. Physiol.* (Paris) 98, 296–314.

Troyer, T., and Miller, K. (1997). Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell. *Neural. Comput.* 9, 971–983.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., and Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* 373, 515–518.

van Rossum, M., Turrigiano, G., and Nelson, S. (2002). Fast propagation of firing rates through layered networks of noisy neurons. *J. Neurosci.* 22, 1956–1966.

Vogels, T. P., and Abbott, L. F., (2005). Signal propagation and logic gating in networks of integrate-and-fire neurons. *J. Neurosci.* 25, 10786–10795.

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., Fries, P., and Jun. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science* 316, 1609–1612.

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

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 November 2009; paper pending published: 21 December 2009; accepted: 24 March 2010; published online: 19 April 2010.

Citation: Rosenbaum RJ, Trousdale J and Josić K (2010) Pooling and correlated neural activity. Front. Comput. Neurosci. ***4****:9. doi: 10.3389/fncom.2010.00009*

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 08 April 2010
doi: 10.3389/neuro.10.001.2010

Signatures of synchrony in pairwise count correlations

***Tatjana Tchumatchenko*[1,2,3]*, *Theo Geisel*[1,2], *Maxim Volgushev*[4,5,6] *and Fred Wolf*[1,2]**

[1] *Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany*
[2] *Bernstein Center for Computational Neuroscience Göttingen, Göttingen, Germany*
[3] *Göttingen Graduate School for Neurosciences and Molecular Biosciences, Göttingen, Germany*
[4] *Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Sciences, Moscow, Russia*
[5] *Department of Neurophysiology, Ruhr-University Bochum, Bochum, Germany*
[6] *Department of Psychology, University of Connecticut, Storrs, CT, USA*

Edited by:
Matthias Bethge, Max Planck Institute for Biological Cybernetics, Germany

Reviewed by:
Eric Shea-Brown, University of Washington, USA
Benjamin Lindner, Max Planck Institute, Germany

***Correspondence:**
Tatjana Tchumatchenko, Bernstein Center for Computational Neuroscience Göttingen, Bunsenstr. 10, 37073 Göttingen, Germany.
e-mail: tatjana@nld.ds.mpg.de

Concerted neural activity can reflect specific features of sensory stimuli or behavioral tasks. Correlation coefficients and count correlations are frequently used to measure correlations between neurons, design synthetic spike trains and build population models. But are correlation coefficients always a reliable measure of input correlations? Here, we consider a stochastic model for the generation of correlated spike sequences which replicate neuronal pairwise correlations in many important aspects. We investigate under which conditions the correlation coefficients reflect the degree of input synchrony and when they can be used to build population models. We find that correlation coefficients can be a poor indicator of input synchrony for some cases of input correlations. In particular, count correlations computed for large time bins can vanish despite the presence of input correlations. These findings suggest that network models or potential coding schemes of neural population activity need to incorporate temporal properties of correlated inputs and take into consideration the regimes of firing rates and correlation strengths to ensure that their building blocks are an unambiguous measures of synchrony.

Keywords: spike correlations, count correlations, population models, synchrony, correlation coefficient

INTRODUCTION

Coordinated activity of neural ensembles contributes a multitude of cognitive functions, e.g., attention (Steinmetz et al., 2000), encoding of sensory information (Stopfer et al., 1997; Galan et al., 2006), stimulus anticipation and discrimination (Zohary et al., 1994; Vaadia et al., 1995). Novel experimental techniques allow simultaneous recording of activity from a large number of neurons (Greenberg et al., 2008) and offer new possibilities to relate the activity of neuronal populations to sensory processing and behavior. Yet, understanding the function of neural assembles requires reliable tools for quantification, analysis and interpretation of multiple simultaneously recorded spike trains in terms of underlying connectivity and interactions between neurons.

As a first step beyond the analysis of single neurons in isolation, much attention has focused on the pairwise spike correlations (Schneidman et al., 2006; Macke et al., 2009; Roudi et al., 2009), their temporal structure and the influence of topology (Kass and Ventura, 2006; Kriener et al., 2009; Ostojic et al., 2009; Tchumatchenko et al., 2010). Pairwise neuronal correlations are traditionally quantified using count correlations, e.g., correlation coefficients (Perkel et al., 1967). However, it remains largely elusive how correlations present in the input to pairs of neurons are reflected in the count correlations of their spike trains. What are the signatures of input correlations in the count correlations? And vice versa, what conclusions about input correlations and interactions can be drawn on the basis of count correlations and their changes?

Here we address these questions using a framework of Gaussian random functions. We find that correlation coefficients can be a poor indicator of input synchrony for some cases of input correlations. In particular, count correlations computed for time bins larger than the intrinsic temporal scale of correlations can vanish for some functional forms of input correlations. These potential ambiguities were not reported in previous studies of leaky integrate and fire models which focused on the analytically accessible choice of white noise input currents (de la Rocha et al., 2007; Shea-Brown et al., 2008).

The paper is organized as follows: we first introduce several common spike count measures (Section "Materials and Methods") and the statistical framework (Section "Results"). Then we study the zero time lag correlations (Section "Spike Correlations with Zero Time Lag") and the influence of the temporal structure of input correlations on measures of spike correlations (Section "Temporal Scale of Spike Correlations"). We show that spike count correlations can vanish despite the presence of input cross correlations (Section "Vanishing Count Covariance in the Presence of Cross Correlations"). Finally, we discuss potential consequences of our findings for the design of population models and the experimentally measured spike correlations.

MATERIALS AND METHODS

MEASURES OF CORRELATION

The spike train $s_i(t)$ of a neuron i is completely described by the sequence of spike times t_i. This description is often simplified using discrete bins of size T (**Figure 1**). To describe pairwise spike correlations, several competing measures are used (Perkel et al., 1967; Svirskis and Hounsgaard, 2003; Schneidman et al., 2006; de la Rocha et al., 2007; Shea-Brown et al., 2008; Roudi et al., 2009). Here, we focus on the most commonly used measures of spike correlations: conditional

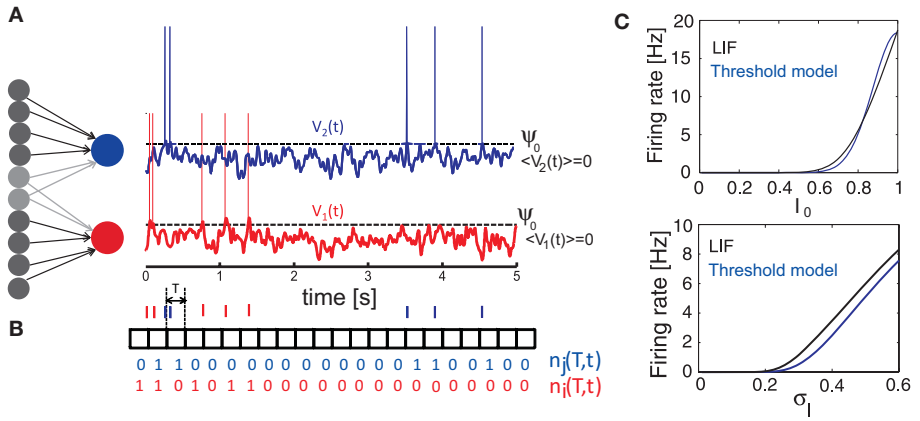


FIGURE 1 | Generation of spike trains and transformation to spike counts. **(A)** Generation of spike trains from correlated voltage traces of two neurons with common presynaptic partners. **(B)** Red and blue vertical bars indicate the spike trains of two neurons. Squares show the boundaries of bins with duration *T*. $n_i(T,t)$ and $n_j(T,t)$ illustrate corresponding binned spike trains. **(C)** Firing rate vs. input current in the LIF model (first order solution) and the threshold model (Eq. 11) computed for $\sigma_I = 0.25$ (top), $I_0 = 0.6$ (bottom) and $\psi_0 = 1$, $V_r = 0$, $\tau_M = 15$ ms and $\tau_I = 5$ ms.

firing rate, correlation coefficient, normalized correlation coefficient and count covariance. We will consider the relation between these measures and their dependence on (1) the underlying input correlation strength, (2) firing rate, (3) temporal structure of spike trains, and (4) size of the time bin used to compute count correlations.

The spike timing correlations of two spike trains $s_i(t)$ and $s_j(t)$ are often quantified using the conditional firing rate function $\nu_{\mathrm{cond},ij}(\tau)$ (Binder and Powers, 2001; Tchumatchenko et al., 2010):

$$\nu_{\mathrm{cond},ij}(\tau) = \langle s_i(t) s_j(t+\tau)\rangle / \sqrt{\nu_i \nu_j} \tag{1}$$

$$\nu_{\mathrm{cond}}(\tau) = \nu_{\mathrm{cond},ii}(\tau) = \langle s_i(t) s_i(t+\tau)\rangle / \nu_i. \tag{2}$$

Here ν_i and ν_j are the mean firing rates of neurons *i* and *j*, respectively. Correlations within a spike train are described by the auto conditional firing rate $\nu_{\mathrm{cond}}(\tau)$.

An alternative measure based on count correlations is the correlation coefficient ρ_{ij} (Perkel et al., 1967; de la Rocha et al., 2007; Greenberg et al., 2008; Shea-Brown et al., 2008; Tetzlaff et al., 2008):

$$\rho_{ij} = \frac{\mathrm{Cov}\left(n_i(T), n_j(T)\right)}{\sqrt{\mathrm{Var}\left(n_i(T), n_i(T)\right)\cdot \mathrm{Var}\left(n_j(T), n_j(T)\right)}}. \tag{3}$$

where $n_i(T)$ and $n_j(T)$ are spike counts of neuron *i* and *j* measured in synchronous time bins of width *T*, see **Figure 1**. A related measure of pairwise correlations is the normalized correlation coefficient c_{ij} (Roudi et al., 2009). It determines pairwise interactions J_{ij} in maximum entropy models of networks of *N* neurons with average firing rate $\bar{\nu}$ (Schneidman et al., 2006; Roudi et al., 2009):

$$c_{ij} = \frac{\mathrm{Cov}\left(n_i(T), n_j(T)\right)}{\langle n_i(T)\rangle \langle n_j(T)\rangle} = \frac{\mathrm{Cov}\left(n_i(T), n_j(T)\right)}{\nu_i \nu_j T^2} \tag{4}$$

$$J_{ij} = \log\left(1 + c_{ij}\right) + O(N\bar{\nu}T). \tag{5}$$

Covariance can be obtained via the integration of cross conditional firing rate $\nu_{\mathrm{cond},ij}(\tau)$ over the time bin *T*:

$$\begin{aligned}\mathrm{Cov}\left(n_i(T), n_j(T)\right) &= \langle n_i(T), n_j(T)\rangle - \langle n_i(T)\rangle \langle n_j(T)\rangle \\ &= \langle \int_0^T s_i(x_1)dx_1 \int_0^T s_j(x_2)dx_2\rangle - \nu_i \nu_j T^2 \qquad (6)\\ &= \int_{-T}^{T} \sqrt{\nu_i \nu_j}\left(\nu_{\mathrm{cond},ij}(t) - \sqrt{\nu_i \nu_j}\right)(T - |t|)dt. \qquad (7)\end{aligned}$$

The count variance can be obtained from the auto conditional firing rate $\nu_{\mathrm{cond}}(\tau)$:

$$\mathrm{Var}\left(n_i(T), n_i(T)\right) = \nu_i \cdot T + 2\cdot \int_0^T \nu_i\left(\nu_{\mathrm{cond}}(t) - \nu_i\right)(T - |t|)dt. \tag{8}$$

For bin sizes smaller than the intrinsic time constant ($T < \tau_s$, see Eq. 14), we can directly relate conditional firing rate $\nu_{\mathrm{cond},ij}(\tau)$ and the correlation coefficient ρ_{ij}

$$\rho_{ij,T<\tau_s} \approx \frac{\sqrt{\nu_i \nu_j}\cdot\left(\nu_{\mathrm{cond},ij}(0) - \sqrt{\nu_i \nu_j}\right)T^2}{\sqrt{\nu_i \nu_j}T\sqrt{\left(1-\nu_i\cdot T\right)\left(1-\nu_i\cdot T\right)}} = \left(\nu_{\mathrm{cond},ij}(0) - \sqrt{\nu_i \nu_j}\right)T \tag{9}$$

$$c_{ij,T<\tau_s} \approx \frac{\sqrt{\nu_i \nu_j}\cdot\left(\nu_{\mathrm{cond},ij}(0) - \sqrt{\nu_i \nu_j}\right)T^2}{\nu_i \nu_j T^2} = \frac{\nu_{\mathrm{cond},ij}(0) - \sqrt{\nu_i \nu_j}}{\sqrt{\nu_i \nu_j}}. \tag{10}$$

In this limit, the properties of ρ_{ij}, c_{ij} are largely determined by $\nu_{\mathrm{cond},ij}(0)$. Several experimental studies used bin sizes ranging from $T = 0.1$ to 1 ms, which are compatible with this *T*-regime of correlation coefficients (e.g., Lampl et al., 1999; Takahashi and Sakurai, 2006).

The quantities presented here all measure different aspects of spike correlations and can potentially have different computational properties. Furthermore, each of the quantities can exhibit a nonlinear dependence on firing rate, input statistics or bin size. Below, we consider these measures of spike correlations, as well as their dependence on firing rate, input statistics and bin size.

RESULTS

To access spike correlations in a pair of neurons, we use the framework of correlated, stationary Gaussian processes to model the voltage potential $V(t)$ of each neuron. This approach generates voltage traces with statistical properties consistent with cortical neurons (Azouz and Gray, 1999; Destexhe et al., 2003). The simplest conceivable model of spike generation from a fluctuating voltage $V(t)$ identifies the spike times t_j with upward crossings of a threshold voltage (Rice, 1954; Jung, 1995; Burak et al., 2009). The times t_j determine the spike train:

$$s(t) = \sum_j \delta\left(t - t_j\right) = \delta\left(V(t) - \psi_0\right) |\dot{V}(t)| \theta\left(\dot{V}(t)\right), \tag{11}$$

where ψ_0 is the threshold voltage, and $\delta(\cdot)$ and $\theta(\cdot)$ are the Dirac delta and Heaviside theta functions, respectively. Each neuron has a stationary firing rate $\nu = \langle s(t)\rangle$. We model $V(t)$ by a random realization of a stationary continuous correlated Gaussian process $V(t)$ (Azouz and Gray, 1999; Destexhe et al., 2003) with zero mean and a temporal correlation function $C(\tau)$, which decays for larger time lags τ.

$$C(\tau) = \left\langle V(t)V(t+\tau)\right\rangle = \left\langle V(0)V(\tau)\right\rangle, \left\langle V(t)\right\rangle = 0 \tag{12}$$

$\langle\cdot\rangle$ denotes the ensemble average. We assume a smooth $C(\tau)$ such that $C^n(0)$ exist for $n \leq 6$ and the rate of threshold crossings is finite (Stratonovich, 1964). All other properties of $C(\tau)$ can be freely chosen. This makes our formal description applicable to a large class of models, each of which is characterized by a particular choice $C(\tau)$. For simulations using digitally synthesized Gaussian processes (Prichard and Theiler, 1994) and numerical integration of Gaussian integrals (e.g., Wolfram Research, 2009) we used a correlation function compatible with power spectra of cortical neurons (Destexhe et al., 2003):

$$C(\tau) = \sigma_V^2 \cosh\left(\tau/\tau_s\right)^{-1}. \tag{13}$$

In cortical neurons *in vivo* the temporal width of $C(\tau)$ can from 10 to 100 ms (Azouz and Gray, 1999; Lampl et al., 1999). We characterize the temporal width of $C(\tau)$ using the correlation time constant τ_s:

$$\tau_s = \sqrt{C(0)/|C''(0)|}. \tag{14}$$

Note, that the correlation time τ_s as defined in Eq. 14 is close to a commonly used definition of autocorrelation time $\tau_a = \int_0^\infty C(\tau)/\sigma_V^2$. For $C(\tau)$ as in Eq. 13 $\tau_a = \pi\tau_s/2$. The correlation time τ_s and the threshold ψ_0 determine the firing rate ν:

$$\nu = \frac{\exp\left[-\psi_0^2/\left(2\sigma_V^2\right)\right]}{2\pi\tau_s}. \tag{15}$$

The firing rate ν is the rate of positive threshold crossings, which is equivalent to half of the Rice rate of a Gaussian process (Rice, 1954). For non-Gaussian processes the rate of threshold crossings can deviate from Eq. 15 and there is no general approach for obtaining ν in this case (Leadbetter et al., 1983). We note, that the firing rate ν of a neuron depends only on two parameters: the correlation time and the threshold-to-variance ratio, but not on the specific functional choice of the correlation function. Hence, processes with the same correlation time but with a different functional form of $C(\tau)$ will have the same mean rate of spikes, though their spike auto and cross correlations can differ significantly. Our framework can be expected to capture neural activity in the regime where the mean time between the subsequent spikes is much longer than the decay time of the spike triggered currents. This occurs if the spikes are sufficiently far apart and the spike decision is primarily determined by the stationary voltage statistics rather than spike evoked currents. Therefore, this model should only be used in the fluctuation driven, low firing rate $\nu < 1/(2\pi\tau_s)$ regime, which is important for cortical neurons (Greenberg et al., 2008).

The leaky integrate and fire (LIF) model (Brunel and Sergi, 1998; Fourcaud and Brunel, 2002) has a similar spike generation mechanism. To compare both models, we study the transformation of input current to spikes. The LIF neuron driven by Ornstein–Uhlenbeck current $I(t)$ with time constant τ_I can be described by

$$\tau_M \dot{V}(\tau) = -V + I_0 + I(t), \tag{16}$$

where τ_M is the membrane time constant and I_0 is the mean input current. When $V(t)$ reaches the threshold ψ_0, the neuron emits a spike, and $V(t)$ is reset to V_r. The LIF model mainly differs from our framework by the presence of reset after each spike. For low firing rates, where the reset has little influence on the following spike, the threshold model and the LIF model can be expected to yield equivalent results. In **Figure 1C** we compare the first order firing rate approximation (first order in $\sqrt{\tau_I/\tau_M}$) of a LIF neuron driven by colored noise, which can be obtained via involved Fokker–Planck calculations (Brunel and Sergi, 1998; Fourcaud and Brunel, 2002) and the firing rate of the corresponding threshold neuron $\nu = (2\pi\sqrt{\tau_I\tau_M})^{-1}\exp(-(I_0-\psi_0)^2(\tau_I+\tau_M)/(2\sigma_I^2\tau_I))$. In general, the details of the spike generating model can have a strong effect on current susceptibility and spike correlations (Vilela and Lindner, 2009). However, we find that both models have a very similar current susceptibility for a range of input currents and spike correlations derived in the forthcoming sections are consistent with the corresponding correlations in the LIF model, e.g., firing rate dependence of weak cross correlations (de la Rocha et al., 2007; Shea-Brown et al., 2008), the influence of noise mean and variance on the firing rates and spike correlations (Brunel and Sergi, 1998; de la Rocha et al., 2007; Ostojic et al., 2009), sublinear dependence of correlation coefficients on input strength (Moreno-Bote and Parga, 2006; de la Rocha et al., 2007).

We include cross correlation between two spike trains i and j via a common component in $V_i(t)$ and $V_j(t)$, $r > 0$:

$$\begin{aligned} V_i(t) &= \sqrt{1-r}\xi_i(t) + \sqrt{r}\xi_c(t) \\ V_j(t) &= \sqrt{1-r}\xi_j(t) + \sqrt{r}\xi_c(t). \end{aligned} \tag{17}$$

where ξ_c denotes the common component and ξ_i, ξ_j are the individual noise components. In a Gaussian ensemble any expectation value is determined by pairwise covariances only. Thus

all pairwise correlations are determined by the joint Gaussian probability density $p(\vec{k}) = \exp(-\vec{k}^T C^{-1}\vec{k}/2)/(4\pi^2\sqrt{\mathrm{Det}C})$ of $\vec{k} = (V_i(0), \dot{V}_i(0), V_j(\tau), \dot{V}_j(\tau))$, where

$$C = \begin{pmatrix} \sigma^2_{V_i} & 0 & C_{ij}(\tau) & C'_{ij}(\tau) \\ 0 & \sigma^2_{\dot{V}_i} & -C'_{ij}(\tau) & -C''_{ij}(\tau) \\ C_{ij}(\tau) & -C'_{ij}(\tau) & \sigma^2_{V_i} & 0 \\ C'_{ij}(\tau) & -C''_{ij}(\tau) & 0 & \sigma^2_{\dot{V}_j} \end{pmatrix}. \quad (18)$$

Matrix entries are covariances $C_{xy} = \langle k_x k_y \rangle$ with $C_{ij} = rC(\tau)$. Below, we calculate the conditional firing rate $\nu_{\mathrm{cond},ij}(\tau)$ (Eqs 1 and 11) for several important limits.

SPIKE CORRELATIONS WITH ZERO TIME LAG

The above framework allows one to derive an analytical expression for the cross conditional firing rate with zero time lag, $\nu_{\mathrm{cond},ij}(0)$. Via Eqs 5, 9 and 10 $\nu_{\mathrm{cond},ij}(0)$ can be related to c_{ij}, ρ_{ij} and J_{ij}. For a pair of statistically identical neurons with $(\nu = \nu_1 = \nu_2)$. $\nu_{\mathrm{cond},ij}(0)$ in Eq. 1 can be solved by transforming the correlation matrix C (Eq. 18) into a block diagonal form via a variable transformation:

$$\Sigma = \frac{V_1(0) + V_2(\tau)}{\sqrt{2}\sqrt{\sigma^2_V + rC(\tau)}}, \dot{\Sigma} = \frac{\dot{V}_1(0) + \dot{V}_2(\tau)}{\sqrt{2}\sqrt{\sigma^2_{\dot{V}} - rC''(\tau)}},$$

$$\Delta = \frac{V_1(0) - V_2(\tau)}{\sqrt{2}\sqrt{\sigma^2_V - rC(\tau)}}, \dot{\Delta} = \frac{\dot{V}_1(0) - \dot{V}_2(\tau)}{\sqrt{2}\sqrt{\sigma^2_{\dot{V}} + rC''(\tau)}}.$$

The matrix C is then the identity matrix for $\tau = 0$, and $\Sigma = \sqrt{2}\psi_0/\sqrt{\sigma^2_V + r\sigma^2_V}, \Delta = 0$. We obtain:

$$\nu_{\mathrm{cond},ij}(0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\dot{\Sigma}, d\dot{\Delta} \exp\left(-\left(\frac{\psi_0^2}{\sigma^2_V(1+r)} + \frac{\dot{\Delta}^2 + \dot{\Sigma}^2}{2}\right)\right)$$
$$\times \frac{\sigma^4_{\dot{V}}\sqrt{(1-r^2)}}{\nu 8\pi^2\sqrt{\mathrm{Det}C}}\left(\dot{\Sigma}^2(1+r) - \dot{\Delta}^2(1-r)\right)$$
$$\times \theta\left(\frac{\sigma_{\dot{V}}}{\sqrt{2}}\left(\dot{\Sigma}\sqrt{(1+r)} + \dot{\Delta}\sqrt{(1-r)}\right)\right) \times \theta\left(\frac{\sigma_{\dot{V}}}{\sqrt{2}}\left(\dot{\Sigma}\sqrt{(1+r)} - \dot{\Delta}\sqrt{(1-r)}\right)\right)$$
$$= \frac{1}{4\pi^2\nu\tau_s^2}\exp\left(\frac{-\psi_0^2}{\sigma^2_V(1+r)}\right)\left[1 + \frac{2r\cdot\arctan\left(\sqrt{\frac{1+r}{1-r}}\right)}{\sqrt{1-r^2}}\right]. \quad (19)$$

Equation 19 (**Figure 3A**) shows, as expected, that $\nu_{\mathrm{cond},ij}(0)$ increases with increasing strength of input correlations r. Since both correlation coefficients ρ_{ij}, and normalized correlation coefficient c_{ij} are proportional to $\nu_{\mathrm{cond},ij}(0)$ (Eqs 9 and 10), both measures also increase with increasing r, which is consistent with experimental findings (Binder and Powers, 2001; de la Rocha et al., 2007). However, the functional form of r-dependence and the sensitivity to the firing rate ν of c_{ij} and ρ_{ij} are different (**Figure 2**). The normalized correlation coefficient c_{ij} and pairwise coupling J_{ij} are both inversely proportional to ν, and thus decrease with increasing ν for any value of r (Eqs 4 and 5; **Figure 2B**). Notably, we find that c_{ij} can be normalized to $c_{ij} \to c_{ij}\cdot(\nu T)$ to yield a less ambiguous measure of the input correlation strength (Eqs 4 and 10; **Figures 3C,D**). Additionally, we find that the firing rate dependence of ρ_{ij} is different for the weak and strong correlations.

Equation 19 further exposes one important feature of $\nu_{\mathrm{cond},ij}(0)$, and thus of c_{ij} and ρ_{ij} for small time bins: all three measures depend on the temporal scale of the input correlations (τ_s), but not on the functional form of input correlation $C(\tau)$. Thus, changes in $\nu_{\mathrm{cond},ij}(0)$ and correlation coefficient ρ_{ij} can be interpreted as a change of the strength of underlying input correlation strength, if a firing rate modification can be excluded.

In the linear r-regime, the analytical expression for $\nu_{\mathrm{cond},ij}(0)$ can be further simplified:

$$\nu_{\mathrm{cond},ij}(0) \approx \nu\left(1 + \frac{r}{2}\left(\pi + 4\left|\log(\nu 2\pi\tau_s)\right|\right)\right). \quad (20)$$

In this limit, $\nu_{\mathrm{cond},ij}(0)$ shows a strong dependence on the firing rate ν (**Figure 3A**, right, **Figure 2A**, top). This dependence is remarkably similar to the firing rate dependence found previously *in vitro* and *in vivo* in cortical neurons and LIF models (de la Rocha et al., 2007; Greenberg et al., 2008; Shea-Brown et al., 2008).

In the limit of strong input correlations, Eq. 19 can be simplified to:

$$\nu_{\mathrm{cond},ij}(0) \approx \frac{1}{2\sqrt{2}\sqrt{1-r}\tau_s}. \quad (21)$$

In this regime, $\nu_{\mathrm{cond},ij}(0)$ does not depend on the firing rate ν (Amari, 2009). Furthermore, for strong input correlations and small bin sizes T the correlation coefficient ρ_{ij} also changes only

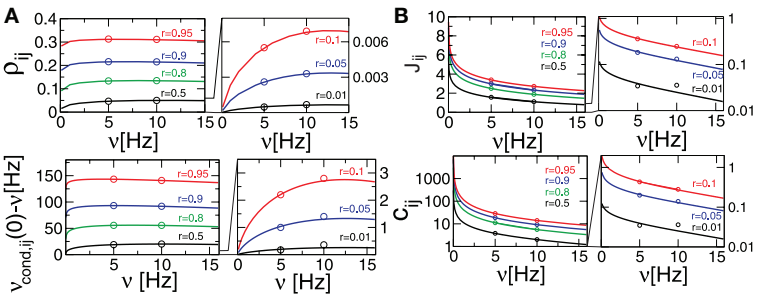


FIGURE 2 | Dependence of correlation coefficient ρ_{ij} and conditional rate $\nu_{\mathrm{cond},ij}(0)$ on firing rate and correlation strength. (**A**, top) ρ_{ij} vs. ν, (**A**, bottom) $\nu_{\mathrm{cond},ij}(0)$ vs. ν, as in Eq. 19. (**B**, top) Pairwise couplings J_{ij} vs. ν, as in Eq. 5. (**B**, bottom) c_{ij} vs. ν. All quantities are computed for τ_s = 10 ms, $C(\tau)$ as in Eq. 13 and $\nu = \nu_1 = \nu_2$; circles denote the corresponding simulation results. ρ_{ij}, c_{ij} and J_{ij} are computed for $T = \tau_s/4$.

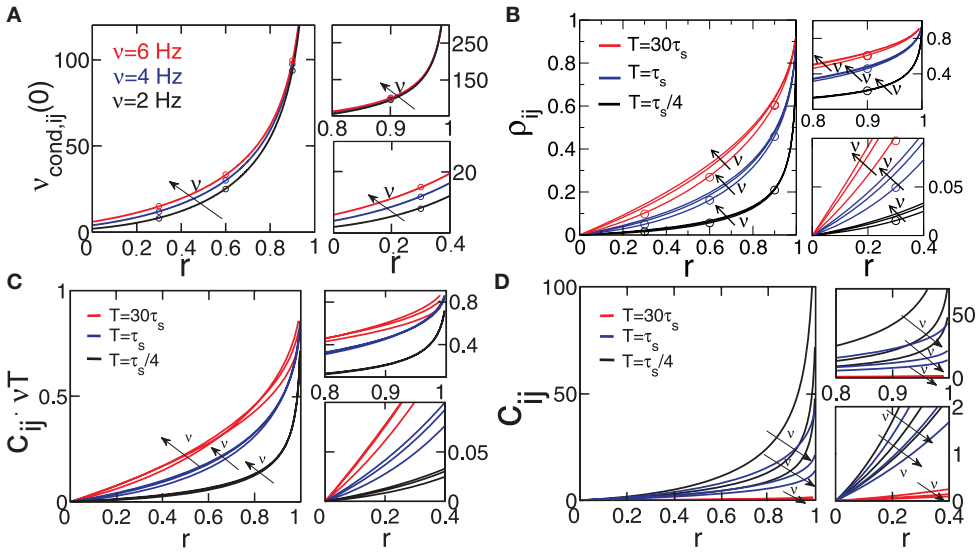


FIGURE 3 | Dependence of spike correlation measures on firing rate ν and correlation strength *r*. **(A)** $\nu_{cond,ij}(0)$ vs. *r* **(B)** ρ_{ij} vs. *r* for bin widths $T = 30\tau_s$ (red), $T = \tau_s$ (blue), $T = \tau_s/4$ (black). **(C)** $c_{ij}\,\nu T$ vs. *r*. **(D)** c_{ij} vs. *r*. All quantities are computed for $C(\tau)$ as in Eq. 13, correlation time τ_s = 10 ms and three firing rates ν = 2, 4, 6 Hz, $\nu = \nu_1 = \nu_2$; circles denote simulation results for the corresponding parameters.

marginally over a range of firing rates ($0 < \nu < 15$ Hz, **Figure 2A**), since it depends linearly on $\nu_{cond,ij}(0)$. Note, as *r* is approaching 1 the temporal width of $\nu_{cond,ij}(\tau)$ is approaching 0 and the peak $\nu_{cond,ij}(0)$ diverges, corresponding to the delta peak in the autoconditional firing rate $\nu_{cond}(\tau)$ which results from the self-reference of a spike. For $r \approx 1$, almost every spike in one train has a corresponding spike in the other spike train, however these two are jittered. The temporal jitter of the spikes can be characterized by the peak of the conditional firing rate $\nu_{cond,12}(\tau) = 1/(2\sqrt{2}\sqrt{1-r}\tau_s) - 3\tau^2/(8\sqrt{2}(1-r)^{3/2}\tau_s^3) + O[(\tau/(\sqrt{1-r}\tau_s))^4]$ and its temporal width $\propto \sqrt{2}\sqrt{1-r}\tau_s$, both of which are threshold and firing rate independent in this limit. Notably, the threshold independence and the dependence on temporal scale of input correlations are consistent with previous experimental findings on spike reliability (Mainen and Sejnowski, 1995).

TEMPORAL SCALE OF SPIKE CORRELATIONS

So far we considered only spike correlations occurring with zero time lag. However, spike correlations can also span across significant time intervals (Azouz and Gray, 1999; Destexhe et al., 2003). The temporal structure of spike correlations, as reflected in the conditional firing rate $\nu_{cond,ij}(\tau)$, can induce temporal correlations within and across time bins and could potentially alter count correlations. To capture correlations with a non-zero time lag, spike correlation measures are calculated for time bins *T* spanning tens to hundreds of milliseconds, e.g., 20 ms (Schneidman et al., 2006), 30–70 ms (Vaadia et al., 1995), 192 ms (Greenberg et al., 2008) and 2 s (Zohary et al., 1994). For time bins longer than the time constant of the input correlations, measures of correlations become sensitive to the temporal structure of $\nu_{cond,ij}(\tau)$. Moreover, the values of ρ_{ij} and c_{ij} depend on the bin size *T* used for their calculation. **Figure 3** shows how dependence of ρ_{ij} and c_{ij} on the firing rate is altered by a change in bin size. Increasing the bin size leads to the increase of the calculated correlation coefficient ρ_{ij}, and also increases the sensitivity of ρ_{ij} to the firing rate. The fact that increasing *T* brings the calculated correlation coefficient closer to the underlying input correlation *r* could justify the use of long time bins in the above studies. But do correlation coefficients always increase with increasing time bins? To further clarify how the temporal structure of input correlations influences the temporal correlations within and across spike trains, we investigate the covariance of spike counts recorded at different times

$$\mathrm{Cov}\left(n_i(T,t), n_j(T,t+\tau)\right) = \left\langle n_i(T,0) n_j(T,\tau)\right\rangle - \nu_i\nu_j T^2$$
$$= \int_{-T}^{T} \sqrt{\nu_i\nu_j}\left(\nu_{cond,ij}(\tau+t) - \sqrt{\nu_i\nu_j}\right)(T-|t|)dt, \qquad (22)$$

where $n_i(T, t)$ and $n_i(T, t+\tau)$ are the spike counts of neurons *i*, *j* measured in time bins of the same duration *T*, but shifted by the time lag τ. For each time lag τ, covariance of the spike counts can be calculated using $\nu_{cond,ij}(\tau)$ (Eq. 1). Below, we will first address the temporal structure of auto correlations in a spike train, and then consider the cross correlations between spike trains.

The auto conditional firing rate $\nu_{cond}(\tau)$

For large time lags τ we expect the auto conditional firing rate to approach the stationary rate but to deviate from it significantly for small time lags. Of particular importance for population models is the limit of small but finite τ, which determines the time scale on which adjacent time bins are correlated. At $\tau = 0$, the auto conditional firing rate has a δ-peak reflecting the trivial auto correlation of each spike with itself. In the limit of small but finite time lag ($0 < \tau < \tau_s$) we find a period of intrinsic silence, where the leading order $\propto\tau^4$ is independent of a particular functional

choice of $C(\tau)$. We solve $\nu_{cond}(\tau)$ (Eq. 2) by transforming the correlation matrix in Eq. 18 into a block diagonal form using new variables

$$\Sigma = \frac{V(0)+V(\tau)}{\sqrt{2}\sqrt{\sigma_V^2 + C(\tau)}}, \dot{\Sigma} = \frac{\dot{V}(0)+\dot{V}(\tau)}{\sqrt{2}\sqrt{\sigma_{\dot{V}}^2 - C''(\tau)}},$$

$$\Delta = \frac{V(0)-V(\tau)}{\sqrt{2}\sqrt{\sigma_V^2 - C(\tau)}}, \dot{\Delta} = \frac{\dot{V}(0)-\dot{V}(\tau)}{\sqrt{2}\sqrt{\sigma_{\dot{V}}^2 + C''(\tau)}}.$$

Then only few elements of the corresponding symmetric density matrix $C_{\Sigma,\dot{\Delta},\dot{\Sigma},\Delta}$ remain non-zero: the diagonal elements $C_{\Sigma,\dot{\Delta},\dot{\Sigma},\Delta_{ii}} = 1$, $i \in \{1, 2, 3, 4\}$ and the non-diagonal elements

$$C_{\Sigma,\dot{\Delta},\dot{\Sigma},\Delta_{12}} = \frac{-C'(\tau)}{\sqrt{\sigma_V^2 + C(\tau)}\sqrt{\sigma_{\dot{V}}^2 + C''(\tau)}},$$

$$C_{\Sigma,\dot{\Delta},\dot{\Sigma},\Delta_{34}} = \frac{C'(\tau)}{\sqrt{\sigma_V^2 - C(\tau)}\sqrt{\sigma_{\dot{V}}^2 - C''(\tau)}}.$$

For $C(\tau)$ as in Eq. 13 we obtain a simple analytical expression in the limit of $0 < \tau < \tau_s$:

$$\nu_{cond}(\tau) = \frac{\nu^{1/4}}{3 \cdot \left(2\pi\tau_s\right)^{\frac{3}{4}}} \cdot (\tau/\tau_s)^4. \quad (23)$$

This equation shows that $\nu_{cond}(\tau)$ depends on the temporal structure of a neuron's input and firing rate, **Figure 4B**. Respectively, the silence period after each spike depends on the functional form and time constant of the voltage correlation function $C(\tau)$ and firing rate (**Figures 4B and 5A**). **Figure 4B** illustrates $\nu_{cond}(\tau)$ obtained using numerical integration of Gaussian probability densities (e.g., Wolfram Research, 2009), $\nu_{cond}(\tau)$ obtained from simulations of digitally synthesized Gaussian processes (Prichard and Theiler, 1994) and the $\tau < \tau_s$ approximation in Eq. 23. In this framework, the silence period after each spike mimics the refractoriness present in real neurons (Dayan and Abbott, 2001).

Count correlations within a spike train

Here we study how the input correlations shape the temporal structure of spike autocorrelations. In particular, we focus on how the input correlations and spike autocorrelations are reflected in count correlations within a spike train. The silence period after a spike is reflected in vanishing $\nu_{cond}(\tau)$ for $0 < \tau < \tau_s$ and results in negative covariation of spike counts in adjacent time bins. We find that the relation between $\nu_{cond}(\tau)$ and spike count covariance is most salient for higher firing rates (**Figure 4C**, 10 Hz). For small time bins, the covariance mimics the functional form of $\nu_{cond}(\tau)$ for time bins covering several time constants. Plots of spike count covariance calculated for increasing bin sizes T reveal an important feature of count correlations: covariance of adjacent bins persists even when the bin size is increased well over the time scale of intrinsic correlations ($T >> \tau_s$), **Figure 4**. This suggests that avoiding statistical dependencies associated with neuronal refractoriness by choosing longer time bins (Shlens et al., 2006) might not be possible, particularly for higher firing rate neurons. We conclude that temporal count correlations within a spike train generally need to be considered in the design of population models.

Cross conditional firing rate $\nu_{cond,ij}(\tau)$

We explore the temporal structure of spike correlations in a weakly correlated pair of statistically identical neurons ($\nu = \nu_1 = \nu_2$). This is an important regime for cortical neurons *in vivo* (Greenberg et al., 2008; Smith and Kohn, 2008). To solve $\nu_{cond,ij}(\tau)$ (Eq. 1), we expand the probability density $p(V_1(t), \dot{V}_1(t), V_2(t+\tau), \dot{V}_2(t+\tau))$ using a von Neumann series of the correlation matrix C in Eq. 18. We obtain $\nu_{cond,ij}(\tau)$ up in linear order

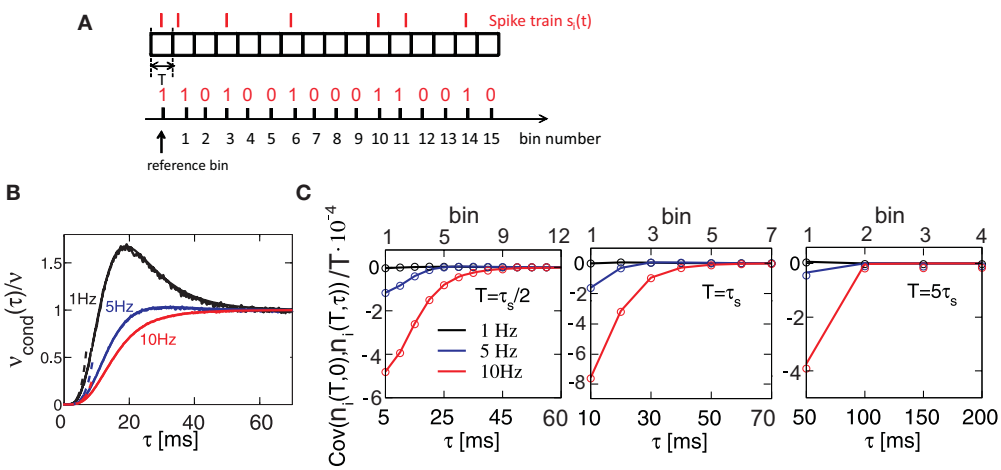


FIGURE 4 | Spike correlations and count correlations within a spike train. **(A)** Example of a binned spike train $s_i(t)$, bins numbered with respect to a reference time bin. **(B)** $\nu_{cond}(\tau)$ vs. τ for τ = 10 ms, numerical solution and simulations for the firing rates ν = 1 Hz (black), 5 Hz (blue) and ν = 10 Hz (red) are superimposed. Dotted lines denote the corresponding solutions for small τ (Eq. 23). **(C)** $\mathrm{Cov}(n_i(T,0),n_j(T,\tau))/T$ vs. τ for τ_s = 10 ms, time bin $T = \tau_s/2$ = 5 ms (left), T = 10 ms = τ_s (middle), $T = 5\tau_s$ = 50 ms (right). Circles denote the corresponding simulation points, adjacent time bins are denoted by the first points on the time axis. All spike correlations are computed for $C(\tau)$ as in Eq. 13.

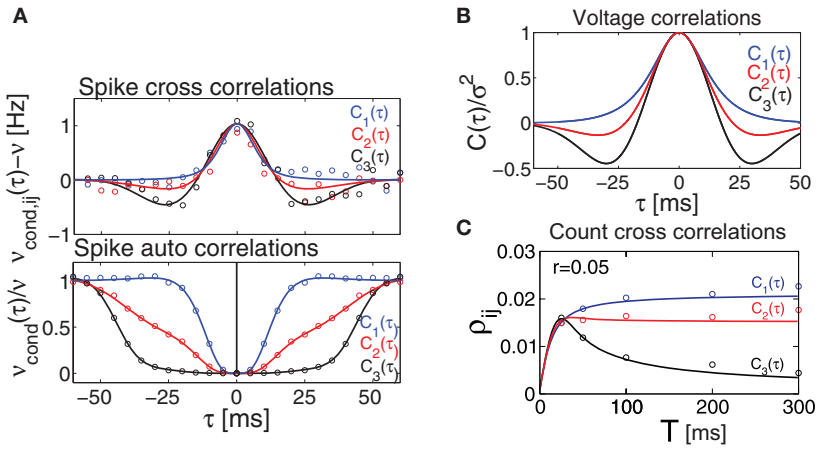


FIGURE 5 | Influence of temporal structure on pairwise spike correlations. **(A)** Spike cross correlations $\nu_{cond,ij}(\tau)$ and auto correlations $\nu_{cond}(\tau)$ for three voltage correlation functions $C_i(\tau)$. **(B)** Voltage correlations $C_1(\tau) = \sigma_V^2 \cosh(\tau/\tau_s)^{-1}$ (blue), $C_2(\tau) = \sigma_V^2 \cosh(\tau/(\sqrt{2\tau_s}))^{-1} \cos(\tau/\sqrt{2\tau_s})$ (red), $C_3(\tau) = \sigma_V^2[\exp(-\tau^2/(6\tau_s^2)) - \tau^2/(3\tau_s^2)\exp(-\tau^2/(6\tau_s^2))]$(black). Note, all voltage correlations $C_i(\tau)$ share the same correlation time τ_s but have a different functional form. **(C)** ρ_{ij} vs. T for voltage correlation functions $C_i(\tau)$. For all figures the correlation time τ_s = 10 ms, ν = 5 Hz, $\nu = \nu_1 = \nu_2$; circles denote the corresponding simulation points.

$$\nu_{cond,ij}(\tau) = \nu\left(1 + r\left(\tilde{c}(\tau)k^2 - \pi\tau_s^2\tilde{c}''(\tau)/2\right)\right), \text{ or}$$
$$\nu_{cond,ij}(\tau) = \nu\left(1 + r\left(\tilde{c}(\tau)2|\log(2\pi\nu\tau_s)| - \pi\tau_s^2\tilde{c}''(\tau)/2\right)\right), \quad (24)$$

where $\tilde{c}(\tau) = C(\tau)/\sigma_V^2$ and $k = \psi_0/\sigma_V$. Equation 24 shows that weak spike correlations are generally firing rate dependent and directly reflect the structure of input correlations $C(\tau)$. **Figure 5A** shows three examples of voltage correlations which have the same τ_s, but different functional form. All three functional dependencies are reflected in the cross conditional firing rate $\nu_{cond,ij}$, but result in markedly different shapes of auto conditional rate $\nu_{cond}(\tau)$ (**Figures 5A,B**). In the next section we study how the functional choice of $C(\tau)$ affects the correlation coefficient.

Count correlations across spike trains

We now use the spike correlation function obtained above to study the pairwise count covariance.

$$\mathrm{Cov}\left(n_i(T), n_j(T)\right) = \int_{-T}^{T} \nu^2 r\left(\tilde{c}(t)2\,|\log(2\pi\nu_s)| - \pi\tau_s^2\tilde{c}''(t)/2\right)(T - |t|)dt, \quad (25)$$

which allows to obtain the correlation coefficient for a weakly correlated pair of neurons:

$$\rho_{ij} = \frac{\mathrm{Cov}\left(n_i(T), n_j(T)\right)}{\sqrt{\mathrm{Var}\left(n_i(T), n_i(T)\right)\mathrm{Var}\left(n_j(T), n_j(T)\right)}} = \frac{\int_{-T}^{T} \nu r\left(2\tilde{c}(t)\,|\log(2\pi\nu\tau_s)| - \pi/2\tau_s^2\tilde{c}''(t)\right)(T - |t|)/T dt}{\sqrt{\left(1 + 2\cdot\int_0^T \left(\nu_{cond}(t) - \nu_i\right)(T - |t|)/T dt\right)\left(1 + 2\cdot\int_0^T \left(\nu_{cond}(t) - \nu_j\right)(T - |t|)/T dt\right)}} \quad (26)$$

This offers the opportunity to study how changes in the input structure affect spike count correlations. **Figure 5** shows that correlation coefficient ρ_{ij} depends on both bin size T and the functional form of input correlation function $C(\tau)$. **Figure 5C** illustrates that different functional form of underlying membrane potential correlations can lead to a strikingly different dependence of ρ_{ij} on the bin size. After an initial increase for all three voltage correlation functions, correlation coefficient continues increasing slowly for C_1, remains at the same level for C_2, but decreases dramatically for C_3. This latter type of behavior was not observed in previous studies of LIF models (de la Rocha et al. (2007), Suppl.), which focused on the analytically accessible choice of white noise currents and reported a monotonously increasing correlation coefficient in the limit of large T. Below we will further consider how dependence of ρ_{ij} on T is influenced by the choice of the form of voltage correlations $C(\tau)$. We will show that some voltage correlation functions can lead to vanishing correlation coefficients in the limit of large bin size T.

Vanishing count covariance in the presence of cross correlations

Count covariances and correlation coefficients rely on the integral of the spike correlation function (Eqs 3 and 7). In cortical neurons, the spike correlation functions can exhibit oscillations and significant undershoots in addition to a correlation peak (Lampl et al., 1999; Galan et al., 2006), this may alter the correlation coefficients and their dependence on bin size T. In the weak correlation regime we obtained an analytic expression for $\nu_{cond,ij}(\tau)$ (Eqs 24 and 26). This allows us to explore analytically how a change in the functional choice of voltage correlations will influence count correlations. To qualify as a reliable measure of synchrony, count cross correlations between two neurons should reflect primarily correlation strength and be independent of the functional form of input correlations. Our framework offers the possibility to test this hypothesis and explore whether previously reported finite correlation coefficients obtained for LIF model using white noise approximation (Shea-Brown et al., 2008) can be generalized to a larger class of input correlations.

Here we consider spike correlations generated by a voltage correlation function with a substantial undershoot (e.g., as in Figure 1E in Lampl et al., 1999). For illustration, we could use any voltage correlation function with a large undershoot and vanishing long-timescale variability ($\int_{-\infty}^{\infty} C(\tau)d\tau = 0$). Besides variance and correlation time, the variability as quantified by $\int_{-\infty}^{\infty} C(\tau)d\tau$ is an important characteristic of every noise process. For analytical tractability we chose the voltage correlation function $C_3(\tau)$ as the normalized second derivative of the function $\tilde{C}_3(\tau) = -3\tau_s^2 \exp(-\tau^2/(6\tau_s^2))$:

$$C_3(\tau) = \sigma_V^2 \left(\exp\left(\frac{-\tau^2}{6\tau_s^2}\right) - \frac{\tau^2}{3\tau_s^2}\exp\left(\frac{-\tau^2}{6\tau_s^2}\right)\right). \quad (27)$$

Defined this way, the correlation time of $C_3(\tau)$ is τ_s and $\int_{-\infty}^{\infty} C_3(\tau)d\tau = 0$, which is equivalent to vanishing spectral power for zero frequency. **Figure 5** illustrates functional form of $C_3(\tau)$ and the corresponding spike cross and auto correlations. The functional form of $C_3(\tau)$ fulfills $\lim_{T\to\infty} \int_{-T}^{T} C_3(t)(T-|t|)/T dt = 0$. This leads to a vanishing count covariance and spike correlation coefficient for $T \gg \tau_s$ (Eq. 26):

$$\mathrm{Cov}\left(n_i(T), n_j(T)\right)/T = \frac{\nu^2 r \tau_s^2 \left[12\left|\log\left(2\pi\nu\tau_s\right)\right|\left(1-\exp\left(\frac{-T^2}{6\tau_s^2}\right)\right) + \pi\left(1+\exp\left(\frac{-T^2}{6\tau_s^2}\right)\left(\frac{T^2}{3\tau_s^2}-1\right)\right)\right]}{T} \quad (28)$$

$$\Rightarrow \lim_{T/\tau_s\to\infty} \frac{\mathrm{Cov}\left(n_i(T), n_j(T)\right)}{T} \to 0, \lim_{T/\tau_s\to\infty} \rho_{ij} \to 0 \quad (29)$$

We note that the correlation coefficients and count covariances calculated for this functional form of input correlations can be arbitrarily small if $T \gg \tau_s$. This means that the absence of long-timescale variability in the inputs ($\int_{-\infty}^{\infty} C_3(\tau)d\tau = 0$) is equivalent to an absence of long-timescale co-variability in the spike counts. Notably, despite vanishing cross covariance, the variability of the single spike train is maintained and count variance of the single spike train (Eq. 8) is finite for $C_3(\tau)$ in infinite time bins. Equation 28 implies that experimental correlation coefficients calculated for large time bins are most susceptible to the influence of temporal structure of correlations, and experimental studies focusing on large bin sizes [e.g., T = 192 ms (Greenberg et al., 2008) or T = 2 s (Zohary et al., 1994)] could potentially underestimate the correlation strength. For the important regime of low firing rates (Greenberg et al., 2008), where the reset has little influence on the following spike, the threshold model and the LIF model can be expected to yield equivalent results. In this case, Eq. 28 and **Figure 5** suggest that finite correlation coefficients, which are increasing with bin size T as reported for the LIF model (de la Rocha et al., 2007) might be limited to the subset of input correlation functions without sizable undershoots. To obtain finite count cross correlations, the voltage correlation functions need to fulfill $\int_{-\infty}^{\infty} C(\tau)d\tau > 0$, as $C_1(\tau), C_2(\tau)$ in **Figure 5** do.

Notably, spike count correlations of cortical neurons *in vivo* can decrease or increase as the length of the time bin increases (Averbeck and Lee, 2003; Smith and Kohn, 2008). These results are consistent with our findings (**Figure 5C**). Thus, in contrast to the correlation coefficients computed for small T which are independent of $C(\tau)$ (Eqs 9 and 19), the count correlations computed for $T \geq \tau_s$ are a potentially unreliable measure of synchrony.

DISCUSSION

Unambiguous and concise measures of spike correlations are needed to quantify and decode neuronal activity (Abbott and Dayan, 1999; Greenberg et al., 2008; Krumin and Shoham, 2009). Pairwise spike count correlations are frequently used to describe interneuronal correlations (Averbeck and Lee, 2003; Kass and Ventura, 2006; Greenberg et al., 2008) and many population models are based on these measures (Schneidman et al., 2006; Shlens et al., 2006; Roudi et al., 2009). However, quantitative determinants of count correlations so far remained largely elusive. Here, we used a simple statistical model framework based on the threshold crossings and the flexible choice of temporal input structure to study the signatures of input correlations in count correlations. In general, the details of the spike generating model can have a strong effect on spike correlations, f.e. depending on the dynamical regime, two quadratic integrate and fire neurons or two LIF neurons can be more strongly correlated (Vilela and Lindner, 2009). Notably, we found that our statistical framework can replicate many important aspects of neuronal correlations, e.g., nonlinear dependence of spike correlations on the input correlation strength (Binder and Powers, 2001) (Eq. 19), firing rate dependence of weak spike correlations (Svirskis and Hounsgaard, 2003; de la Rocha et al., 2007) (Eq. 20), and independence of spike reliability of the threshold (Mainen and Sejnowski, 1995) (Eq. 21). Furthermore, spike correlations derived here are consistent with many recent results in the commonly used LIF model, e.g., firing rate dependence of weak cross correlations (de la Rocha et al., 2007; Shea-Brown et al., 2008) (Eqs 20 and 24), the influence of noise mean and variance on the firing rates and weak spike correlations (Brunel and Sergi, 1998; de la Rocha et al., 2007; Ostojic et al., 2009) (Eqs 15, 20 and 24), or sublinear dependence of correlation coefficients on input strength (Moreno-Bote and Parga, 2006; de la Rocha et al., 2007) (Eq. 19, **Figure 3**). While the analytical accessibility of the LIF model is limited by the technically demanding multi dimensional Fokker–Planck equations and provides solutions only in special limiting cases (Brunel and Sergi, 1998; de la Rocha et al., 2007; Shea-Brown et al., 2008), the framework presented here allows for an analytical description of spike correlations.

Measurements of correlation coefficients under different experimental conditions often aim to compare the input correlation strength in pairs of neurons (Greenberg et al., 2008; Mitchell et al., 2009). But is a change in count correlations always indicative of a change in input correlations? The tractability of our framework revealed that spike count correlations can be a poor indicator of input synchrony for some cases of input correlations. Count correlations computed for time bins smaller than

the intrinsic scale of temporal correlations could be independent of the functional form of input correlations but depend on the firing rate and input correlation strength. This suggests that a change in the correlation coefficient can be related to a change in the input correlation strength, if a firing rate change and a change of intrinsic time scale can be excluded. On the other hand, a change in correlation coefficients computed for large time bins is indicative of a change in input correlation strength only if a change in firing rate, time scale and functional form of input correlations can be excluded. Furthermore, count correlations computed for large time bins can either increase or decrease with increasing time bin or even vanish in a correlated pair. This seemingly contradictory behavior is consistent with the functional dependence of spike count correlations observed in cortical neurons (Averbeck and Lee, 2003; Kass and Ventura, 2006; Smith and Kohn, 2008).

Our results suggest that emulating neuronal spike trains, building efficient population models or determining potential decoding algorithms requires the analysis of full spike correlation functions in order to compute unambiguous spike count correlations. In particular, spike count coefficients computed for time bins larger than intrinsic timescale of correlations can be an ambiguous estimate of input cross correlations in a neuronal population with potentially heterogeneous distribution of input structures. Furthermore, the details of the spike generation model can be very influential for the transfer of current correlations to spike correlations, and the analytical results obtained here could facilitate quantitative comparisons between different types of models and between models and real neurons, by providing a maximally tractable limiting case for future studies.

ACKNOWLEDGMENTS

We wish to thank M. Gutnick, I. Fleidervich, S. Ostojic and A. Malyshev for fruitful discussions and the Bundesministerium für Bildung und Forschung (#01GQ0430,#01GQ07113), Goettingen Graduate School for Neurosciences and Molecular Biosciences, German-Israeli Foundation (#I-906-17.1/2006), University of Connecticut and the Max Planck Society for financial support.

REFERENCES

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

Amari, S. (2009). Measure of correlation orthogonal to change in firing rate. *Neural Comput.* 21, 960–972.

Averbeck, B., and Lee, D. (2003). Neural noise and movement-related codes in the macaque supplementary motor area. *J. Neurosci.* 23, 7630–7641.

Azouz, R., and Gray, C. M. (1999). Cellular mechanisms contributing to response variability of cortical neurons in vivo. *J. Neurosci.* 19, 2209–2223.

Binder, M. D., and Powers, R. K. (2001). Relationship between simulated common synaptic input and discharge synchrony in cat spinal motoneurons. *J. Neurophysiol.* 86, 2266–2275.

Brunel, N., and Sergi, S. (1998). Firing frequency of leaky integrate-and-fire neurons with synaptic current dynamics. *J. Theor. Biol.* 195, 87–95.

Burak, Y., Lewallen, S., and Sompolinsky, H. (2009). Stimulus-dependent correlations in threshold-crossing spiking neurons. *Neural Comput.* 21, 2269–2308.

Dayan, P., and Abbott, L. (2001). Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. Cambridge, MA, The MIT Press.

de la Rocha, J., Doiron, B., Shea-Brown, E., Josic, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* 448, 802–806.

Destexhe, A., Rudolph, M., and Pare, D. (2003). The high-conductance state of neocortical neurons *in vivo*. *Nat. Rev. Neurosci.* 4, 739–751.

Fourcaud, N., and Brunel, N. (2002). Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput.* 14, 2057–2110.

Galan, R., Fourcaud-Trocme, N., Ermentrout, G., and Urban, N. (2006). Correlation-induced synchronization of oscillations in olfactory bulb neurons. *J. Neurosci.* 26, 3646–3655.

Greenberg, D., Houweling, A., and Kerr, J. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat. Neurosci.* 11, 749–751.

Jung, P. (1995). Stochastic resonance and optimal design of threshold detectors. *Phys. Lett. A* 207, 93–104.

Kass, R., and Ventura, V. (2006). Spike count correlation increases with length of time interval in the presence of trial-to-trial variation. *Neural Comput.* 18, 2583–2591.

Kriener, B., Helias, M., Aertsen, A., and Rotter, S. (2009). Correlations in spiking neuronal networks with distance dependent connections. *J. Comput. Neurosci.* 27, 177–200.

Krumin, M., and Shoham, S. (2009). Generation of spike trains with controlled auto- and cross-correlation functions. *Neural Comput.* 21, 1642–1664.

Lampl, I., Reichova, I., and Ferster, D. (1999). Synchronous membrane potential fluctuations in neurons of the cat visual cortex. *Neuron* 22, 361–374.

Leadbetter, M., Lindgren, G., and Rootzen, H. (1983). Extremes and Related Properties of Random Sequences and Processes. Springer, New York (Springer Series in Statistics Edition).

Macke, J., Berens, P., Ecker, A., Tolias, A., and Bethge, M. (2009). Generating spike trains with specified correlation coefficients. *Neural Comput.* 2, 397–423.

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

Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area v4. *Neuron* 63, 879–888.

Moreno-Bote, R., and Parga, N. (2006). Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Phys. Rev. Lett.* 96, 028101.

Ostojic, S., Brunel, N., and Hakim, V. (2009). How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J. Neurosci.* 29, 10234–10253.

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967). Neuronal spike trains and stochastic point processes. ii. simultaneous spike trains. *Biophys. J.* 7, 419–440.

Prichard, D., and Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett.* 73, 951–954.

Rice, S. O. (1954). Mathematical analysis of random noise. In Selected Papers on Noise and Stochastic Processes, N. Wax, ed. (New York, Dover).

Roudi, Y., Nirenberg, S., and Latham, P. (2009). Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS Comput. Biol.*, 5, e1000380. doi:10.1371/journal.pcbi.1000380.

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012.

Shea-Brown, E., Josic, K., de la Rocha, J., and Doiron, B. (2008). Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Phys. Rev. Lett.* 100, 108102.1–108102.4.

Shlens, J., Field, G., Gauthier, J., Grivich, M., Petrusca, D., Sher, A., Litke, A., and Chichilnisky, E. (2006). The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* 26, 8254–8266.

Smith, M. A., and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* 28, 12591–12603.

Steinmetz, P. N., Roy, A., Fitzgerald, P. J., Hsiao, S. S., Johnson, K. O., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* 404, 187–190.

Stopfer, M., Bhagavan, S., Smith, B. H., and Laurent, G. (1997). Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* 390, 70–74.

Stratonovich, R. (1964). Topics in the Theory of Random Noise, Vols I–II. New York, Gordon and Breach.

Svirskis, G., and Hounsgaard, J. (2003). Influence of membrane properties on spike synchronization in neurons: theory and experiments. *Network* 14, 747–763.

Takahashi, S., and Sakurai, Y. (2006). Dynamic synchrony of firing in the

monkey prefrontal cortex during working-memory tasks. *J. Neurosci.* 26, 10141–10153.

Tchumatchenko, T., Malyshev, A., Geisel, T., Volgushev, M., and Wolf, F. (2010). Correlations and synchrony in threshold neuron models. *Phys. Rev. Lett.* 104, 058102.

Tetzlaff, T., Rotter, S., Stark, E., Abeles, M., Aertsen, A., and Diesmann, M. (2008). Dependence of neuronal correlations on filter characteristics and marginal spike train statistics. *Neural Comput.* 20, 2133–2184.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., and Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* 373, 515–518.

Vilela, R. D., and Lindner, B. (2009). Comparative study of different integrate-and-fire neurons: spontaneous activity, dynamical response, and stimulus-induced correlation. *Phys. Rev. E* 80, 031909.

Wolfram Research (2009). *As Implemented in MATHEMATICA 5.2.* Wolfram Research.

Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 November 2009; paper pending published: 07 December 2009; accepted: 05 February 2010; published online: 08 April 2010.

Citation: Tchumatchenko T, Geisel T, Volgushev M and Wolf F (2010) Signatures of synchrony in pairwise count correlations. Front. Comput. Neurosci. ***4****:1. doi: 10.3389/neuro.10.001.2010*

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 15 November 2010
doi: 10.3389/fncom.2010.00144

Modeling population spike trains with specified time-varying spike rates, trial-to-trial variability, and pairwise signal and noise correlations

Dmitry R. Lyamzin[1†], Jakob H. Macke[2,3†] and Nicholas A. Lesica[1*†]

[1] *Division of Neurobiology, Department of Biology II, Ludwig-Maximilians-University Munich, Martinsried, Germany*
[2] *Computational Vision and Neuroscience Group, Max Planck Institute for Biological Cybernetics, Tübingen, Germany*
[3] *Werner Reichhardt Centre for Integrative Neuroscience, University of Tübingen, Tübingen, Germany*

Edited by:
Klaus R. Pawelzik, University of Bremen, Germany

Reviewed by:
Udo Ernst, University of Bremen, Germany
Kresimir Josic, University of Houston, USA

****Correspondence:***
Nicholas A. Lesica, Ear Institute, University College London, 332 Gray's Inn Rd., London WC1X 8EE, UK.
e-mail: n.lesica@ucl.ac.uk

†Current address:
Dmitry R. Lyamzin and Nicholas A. Lesica, Ear Institute, University College London, London, UK.
Jakob H. Macke, Gatsby Computational Neuroscience Unit, University College London, London, UK.

As multi-electrode and imaging technology begin to provide us with simultaneous recordings of large neuronal populations, new methods for modeling such data must also be developed. Here, we present a model for the type of data commonly recorded in early sensory pathways: responses to repeated trials of a sensory stimulus in which each neuron has it own time-varying spike rate (as described by its PSTH) and the dependencies between cells are characterized by both signal and noise correlations. This model is an extension of previous attempts to model population spike trains designed to control only the total correlation between cells. In our model, the response of each cell is represented as a binary vector given by the dichotomized sum of a deterministic "signal" that is repeated on each trial and a Gaussian random "noise" that is different on each trial. This model allows the simulation of population spike trains with PSTHs, trial-to-trial variability, and pairwise correlations that match those measured experimentally. Furthermore, the model also allows the noise correlations in the spike trains to be manipulated independently of the signal correlations and single-cell properties. To demonstrate the utility of the model, we use it to simulate and manipulate experimental responses from the mammalian auditory and visual systems. We also present a general form of the model in which both the signal and noise are Gaussian random processes, allowing the mean spike rate, trial-to-trial variability, and pairwise signal and noise correlations to be specified independently. Together, these methods for modeling spike trains comprise a potentially powerful set of tools for both theorists and experimentalists studying population responses in sensory systems.

Keywords: population, correlation, noise correlation, simulation, model

INTRODUCTION

Correlated spiking activity in neuronal populations has been a subject of intense theoretical and experimental research over the past several decades, and the importance of correlations has been demonstrated in a number of contexts, including plasticity and information processing (for a recent review, see Averbeck et al., 2006). Recent advances in experimental technology have finally made it possible to observe the activity of large neuronal populations simultaneously. In order to take full advantage of these advances, new methods for the analysis and modeling of population activity must also be developed.

A number of methods exist for modeling correlated population spike trains in which some fraction of the input driving the activity of each neuron is shared with other neurons, including integrate-and-fire models and other spiking models with correlated input currents or synaptic conductances (Destexhe and Pare, 1999; Feng and Brown, 2000; Song and Abbott, 2001; Stroeve and Gilen, 2001; Salinas and Sejnowski, 2002; Dorn and Ringach, 2003; Gutig et al., 2003; Galan et al., 2006; De La Rocha et al., 2007; Shea-Brown et al., 2008; Tchumatchenko et al., 2008), stochastic spiking models with correlated rate functions (Galan et al., 2006; Brette, 2009; Krumin and Shoham, 2009), and models based on a dichotomized Gaussian (DG) framework (Macke et al., 2009; Gutnisky and Josic, 2010). There are also a variety of methods for capturing precise synchrony between neurons through explicit sharing of spikes (Kuhn et al., 2003; Galan et al., 2006; Niebur, 2007; Brette, 2009) and several models based on statistical frameworks such as maximum entropy (Schneidman et al., 2006; Shlens et al., 2006; Roudi et al., 2009).

All of the approaches described above are designed to capture and/or control the total correlation between spike trains and, as such, are of limited utility in the context of early sensory systems where it is important to separate internal network correlations from those due to the external stimulus. In this paper, we propose a framework designed specifically to model spike trains in which the total correlation can be separated into signal and noise components. If responses to repeated trials of an identical sensory stimulus are observed, the signal correlation, which reflects both correlation in the stimulus itself and similarities in neurons' preferred stimulus features, will be evident in the fraction of the response that is repeatable from trial-to-trial. Noise correlation, which results from the activity of network and intrinsic cellular mechanisms, will be evident in the fraction of the response that is variable from trial-to-trial (note that the term noise correlation is not meant to imply that the activity underlying this correlation is unimportant, but simply that is not directly dependent on the stimulus).

For modeling the population spike trains of early sensory neurons, another class of methods based on generalized linear models (GLMs) has been developed (Chornoboy et al., 1988; Paninski, 2004; Kulkarni and Paninski, 2007; Paninski et al., 2007; Pillow et al., 2008). In its typical formulation, the GLM is parameterized by a series of filters that relate the time-varying spike rate in one neuron to the sensory stimulus and the responses of other neurons. This formulation has the great strength that once the filter parameters have been estimated, the model can be used not only to simulate responses that match those measured experimentally, but also to simulate responses to novel stimuli. However, this generality comes at a cost: specifying the filters requires the estimation of a large number of parameters and, thus, a large amount of experimental data – much more than is necessary for a model designed only to simulate responses to the same stimuli that have been tested experimentally. It is possible to formulate alternatives to the typical GLM that require less experimental data by forgoing the ability to predict responses to novel stimuli and parameterizing the time-varying firing rate in response to a particular stimulus directly. However, even in this formulation, the GLM lacks a critical property: it does not enable straightforward specification or manipulation of one response property independent of the others (Krumin and Shoham, 2009; Toyoizumi et al., 2009).

In the model we present below, the time-varying spike rate, trial-to-trial variability, and pairwise signal and noise correlations can be matched to those measured experimentally, and the noise correlations can be manipulated without changes in the signal correlations or the single-cell properties. The model is an extension of previous attempts to model population spike trains as DGs (Emrich and Piedmonte, 1991; Cox and Wermuth, 2002; Macke et al., 2009; Gutnisky and Josic, 2010). In our model, the response of each cell is a binary vector determined by the thresholded sum of two inputs: a signal, which is the same for each trial of a given stimulus, and a noise, which is different for each trial, both of which can be correlated across neurons. In the first part of the paper, we show how the model parameters can be estimated from experimental data and used to simulate spike trains with properties that match those measured experimentally. We also demonstrate how the model parameters can be manipulated to obtain spike trains with arbitrary pairwise noise correlations without changes in single-cell properties. In the second part of the paper, we describe a general form of the model that can be used model spike trains with arbitrary single-cell properties and pairwise correlations.

All of the Matlab code required to perform the analyses described in this paper is available for download at http://www.ucl.ac.uk/ear/research/lesicalab.

A MODEL FOR SIMULATING AND MANIPULATING EXPERIMENTALLY RECORDED POPULATION SPIKE TRAINS

SINGLE-CELL RESPONSES

To represent a set of spike times from a single cell on a single trial $i \in \{1,2,\ldots,I\}$ of a particular stimulus, we discretize time into $n \in \{1,2,\ldots,N\}$ bins of length Δ and set $r_i[n] = 1$ if a spike occurs in bin n on trial i, and $r_i[n] = 0$ otherwise. In general, we assume that Δ is small enough that no more than one spike occurs in any given bin. Based on the responses to all trials r (an $N \times I$ binary matrix), we can define several quantities of interest:

Mean spike rate	$r_0 = \langle r \rangle_{n,i}$	(scalar)
Time-varying spike rate (PSTH)	$\bar{r} = \langle r \rangle_i$	(N-dimensional vector)
Spike train signal to noise ratio	$SNR = \dfrac{\mathrm{var}(\bar{r})}{\langle \mathrm{var}(\xi_i) \rangle_i}$	(scalar)

where $\xi_i = \bar{r} - r_i$ is the residual on trial i

Note that we use the notation $\langle \cdot \rangle_x$ to represent the expectation over all possible values of x, $\langle \cdot \rangle_{x,y}$ to represent the expectation over all possible values of x followed by the expectation of all possible values of y, and $\langle \cdot \rangle_{x \neq y}$ to represent the expectation over all possible combinations of x and y in which their values are not equal. We chose to use the above definition of SNR as the measure of trial-to-trial variability because is it commonly used in early sensory systems (Borst and Theunissen, 1999). One important property of this measure that should be noted is that its value is dependent on the bin size Δ. Thus, all of the computations described below for fitting model parameters must be repeated if the bin size is changed.

We model the response as a dichotomized sum of a deterministic "signal" and Gaussian "noise"

$$r_i[n] = \begin{cases} 1, & (s[n] + z_i[n]) > 0 \\ 0, & (s[n] + z_i[n]) \leq 0 \end{cases} \quad (1)$$

Where $r_i[n]$ is the response in time bin N on trial i, s (an N-dimensional vector) is the same on every trial and $z \sim \mathcal{N}(0, 1)$ (an N-dimensional vector) is different on every trial [note that neither s nor z are intended to correspond directly to any intracellular quantities]. Given the experimentally recorded responses of a cell, we wish to simulate responses with the same PSTH $\bar{r}$. This can be done by solving

$$\bar{r}[n] = \langle r_i[n] \rangle_i = \Phi(s[n], 1) \quad (2)$$

for $s[n]$ in each bin, where $\Phi(x, \sigma^2)$ is the CDF for a Gaussian with zero mean and variance σ^2 evaluated at x. Equation 2 is easily solved numerically, as the function is monotonic and has unique level crossings. It is clear from Eq. 2 that the choice of one for the variance of z is somewhat arbitrary; for any finite value of the variance of z, an $s[n]$ can be found to achieve any desired value of $\bar{r}[n]$. Note that if $\bar{r}[n] = 0$ or 1, then $s[n]$ must be either $-\infty$ or $+\infty$. If finite values of $s[n]$ are desired, then $\bar{r}[n]$ can be constrained to the interval $[1/I, 1 - 1/I]$ before solving Eq. 2.

Importantly, since this approach matches $\bar{r}$ exactly, it will also match the mean spike probability r_0 and the spike train signal to noise ratio SNR, as both can be uniquely defined in terms of the PSTH $\bar{r}$:

$$r_0 = \langle \bar{r} \rangle_n$$

$$\mathrm{SNR} = \frac{\mathrm{var}(\bar{r})}{\langle \mathrm{var}(\xi_i) \rangle_i} = \frac{\mathrm{var}(\bar{r})}{\langle \mathrm{var}(\bar{r} - r_i) \rangle_i}$$

$$= \frac{\mathrm{var}(\bar{r})}{\mathrm{var}(\bar{r}) + \langle \mathrm{var}(r_i) \rangle_i - 2\langle \mathrm{cov}(\bar{r}, r_i) \rangle_i}$$

where, because r_i is binary,

$$\langle \mathrm{var}(r_i) \rangle_i = r_0(1 - r_0) \quad \text{and} \quad \langle \mathrm{cov}(\bar{r}, r_i) \rangle_i = \langle \bar{r}^2 \rangle_n - r_0^2.$$

Matching $\bar{r}$ exactly will also match the mutual information transmitted by single spikes (Brenner et al., 2000). Note that if it is not necessary to match the bin to bin spike probabilities of the experimental response, but only the distribution of overall spike counts, a reduced model can be used (Macke et al., 2009).

To demonstrate the utility of this model, we first generated responses using Eq. 1 with a variety of different signals, and then attempted to reproduce the model responses after estimating s using Eq. 2. Typical results are shown in **Figure 1A**. For uniform random, sine wave, and square wave signals, the PSTH and, consequently, r_0 and SNR of the responses simulated with the estimated s closely match those of the original model generated data.

Next, we tested the model's ability to reproduce the single-cell properties of experimentally recorded responses. **Figure 1B** shows the responses of neurons in the gerbil inferior colliculus to repeated presentations of a variety of sounds. In each case, we estimated s from the experimental data using Eq. 2 and were able to simulate new responses with PSTH, r_0, and SNR that match those measured experimentally.

POPULATION RESPONSES

As described in the Introduction, correlations between cells in early sensory systems can have both signal and noise components: signal correlations arise from correlations in the stimulus itself and/or similarity in preferred stimulus features (frequency, orientation, etc.), while noise correlations arise from shared inputs that contribute to the trial-to-trial variability in responses. In our model, we adopt the most common definition of noise correlation, where ρ^{pq}_{noise}, the noise correlation between cells p and q, is given by the difference between the total correlation and the signal correlation, $\rho^{pq}_{\text{noise}} = \rho^{pq}_{\text{total}} - \rho^{pq}_{\text{signal}}$, and ρ^{pq}_{total} and $\rho^{pq}_{\text{signal}}$ are the correlation coefficients between the responses of cells p and q before and after the trial order has been shuffled. The model described above for a single cell is easily extended to capture the pairwise signal and noise correlations in a population, where the response of cell $p \in \{1, 2,\ldots,P\}$ is given by

$$r_i^p[n] = \begin{cases} 1, & \left(s^p[n] + z_i^p[n]\right) > 0 \\ 0, & \left(s^p[n] + z_i^p[n]\right) \le 0 \end{cases} \tag{3}$$

where each cell has its own s^p that is the same on every trial and z^p that is different on every trial. In this population model, $z \sim \mathcal{N}(0, \Sigma_z)$ is a multivariate (P-dimensional) Gaussian random process with covariance matrix

$$\Sigma_z = \begin{bmatrix} 1 & \rho_z^{12} & \cdots & \rho_z^{1P} \\ \rho_z^{21} & 1 & & \\ \vdots & & \ddots & \\ \rho_z^{P1} & & & 1 \end{bmatrix},$$

where ρ_z^{pq}, which is assumed to be constant across time bins and trials, is the pairwise correlation coefficient between z^p and z^q and $\rho_z^{pq} = \rho_z^{qp}$. Assuming we have the responses of a population to repeated trials of a particular stimulus, we can estimate each s^p separately to match the single-cell properties as described above. Because the response is binary and this approach matches $\bar{r}$ exactly for each cell, it will also match the signal correlation between cells. To match the noise correlation, it is necessary to find the appropriate covariance matrix Σ_z. This can be done by solving the equation that relates ρ_z^{pq} to the spike train noise correlation ρ^{pq}_{noise} numerically for each pair of cells (again, the function is monotonic and, because z is Gaussian, each ρ_z^{pq} can be solved for independently).

Thus, ρ^{pq}_{noise} can be written as

$$\begin{aligned} \rho^{pq}_{\text{noise}} &= \rho^{pq}_{\text{total}} - \rho^{pq}_{\text{signal}} \\ &= \frac{\left\langle \text{cov}\left(r_i^p, r_i^q\right)\right\rangle_i}{\sqrt{\left\langle \text{var}\left(r_i^p\right)\right\rangle_i \left\langle \text{var}\left(r_i^q\right)\right\rangle_i}} - \frac{\left\langle \text{cov}\left(r_i^p, r_j^q\right)\right\rangle_{i\neq j}}{\sqrt{\left\langle \text{var}\left(r_i^p\right)\right\rangle_i \left\langle \text{var}\left(r_i^q\right)\right\rangle_i}} \end{aligned} \tag{4}$$

where, because r_i is binary,

$$\left\langle \text{var}\left(r_i^p\right)\right\rangle_i = r_0{}^p\left(1 - r_0{}^p\right)$$

and, because z is Gaussian,

$$\left\langle \text{cov}\left(r_i^p, r_i^q\right)\right\rangle_i = \left\langle \Phi_2\left(\begin{bmatrix} s^p \\ s^q \end{bmatrix}, \begin{bmatrix} 1 & \rho_z^{pq} \\ \rho_z^{pq} & 1 \end{bmatrix}\right)\right\rangle_n - r_0{}^p r_0{}^q \quad \text{and}$$

$$\left\langle \text{cov}\left(r_i^p, r_j^q\right)\right\rangle_{i\neq j} = \left\langle \Phi\left(s^p, 1\right)\Phi\left(s^q, 1\right)\right\rangle_n - r_0{}^p r_0{}^q$$

where $\Phi_2(\vec{x}, \Sigma)$ is the CDF for a two-dimensional Gaussian with zero mean and covariance Σ evaluated at $\vec{x}$.

To demonstrate the utility of this approach, we first attempted to reproduce the single-cell properties and pairwise correlations recorded experimentally from a population of 10 cells in the cat lateral geniculate nucleus in response to repeated presentations of a natural scene movie. **Figure 2A** shows the experimental and simulated responses for two cells. As expected, the PSTH, r_0, and SNR of the experimental and simulated responses are closely matched. As shown in **Figure 2B**, the measured and simulated pairwise noise correlations in the population are also closely matched.

In addition to matching the experimentally observed responses, this approach can also be used to manipulate pairwise correlations without disturbing single-cell properties by changing the value of ρ^{pq}_{noise} on the left side of Eq. 4 before solving for ρ_z^{pq} (note that there are a number of constraints on the realizable values of ρ^{pq}_{noise} – for example, because the covariance matrix Σ_z must be positive semi-definite, it may be difficult to obtain strong negative correlations; see Macke et al., 2009 for a detailed discussion). As a demonstration, we attempted to simulate population spike trains in which the noise correlations were twice as large as those observed experimentally. As shown in **Figure 2C**, the noise correlations in the simulated data match those desired.

EVALUATING GOODNESS OF FIT

Our model is not fit directly to observed spike trains, but rather to the PSTHs and pairwise noise correlations that are extracted from them. In our framework, any set of PSTHs and noise correlations can be fit with a unique set of model parameters, but that does not mean, of course, that the model is a good description of the original spike trains. The actual goodness of fit of the model is determined by two factors: the measurement noise in the PSTHs and noise correlations and the validity of the assumption that the spike trains can be described by our model

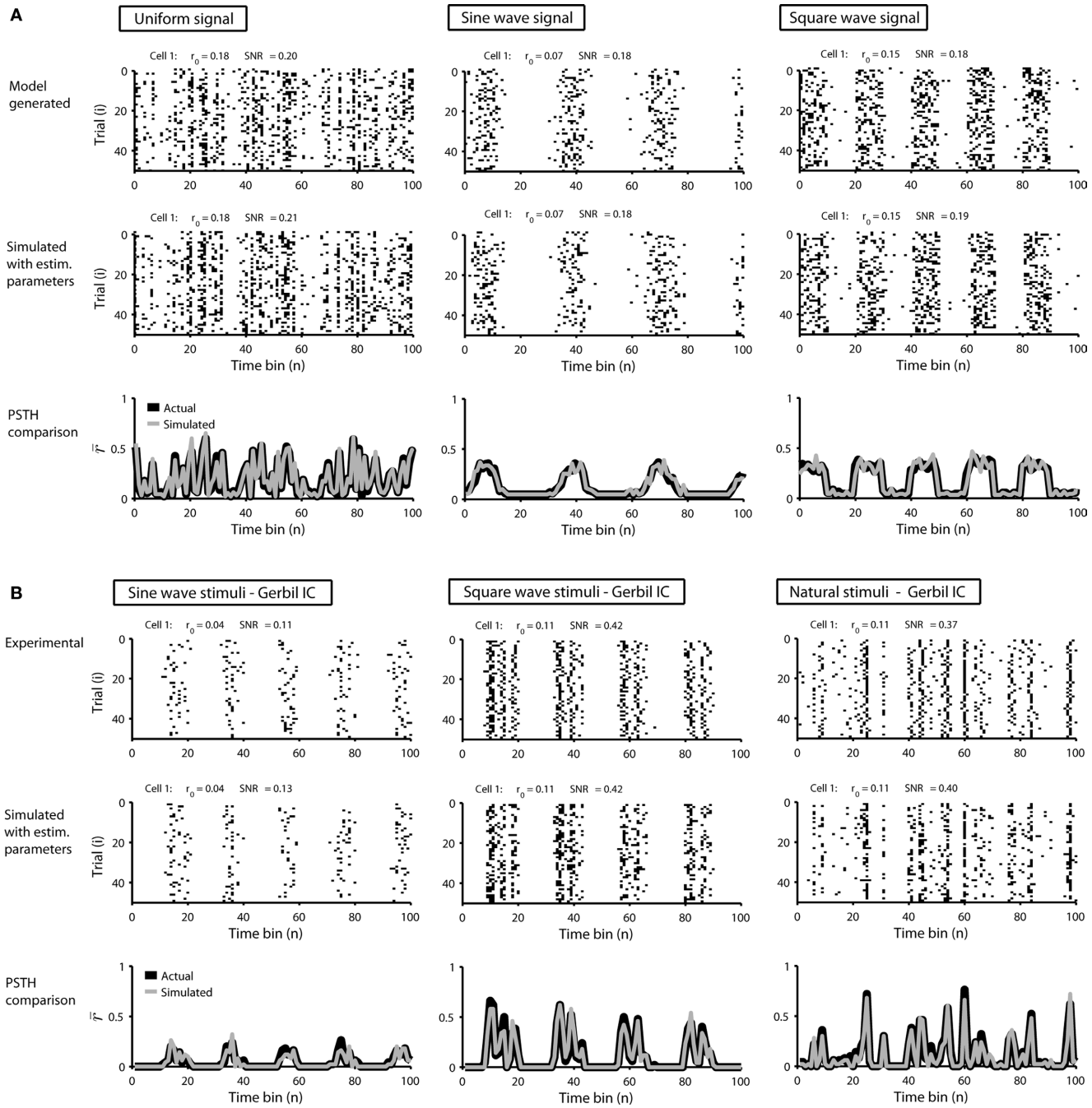


FIGURE 1 | Simulated responses match the single-cell properties of model generated and experimental data. (A) The top row shows model generated responses to repeated trials with a variety of waveforms for *s*: a uniform random noise signal in the range [–0.5 to 0.5]; a sinusoidal signal 0.75 sin(*n*/5) – 1; a square wave signal with a 50% duty cycle, a period of 20 samples, and the same mean value and peak-to-peak amplitude as the sinusoid. The middle row shows simulated responses with *s* estimated from the responses in the top row. The bottom row shows the PSTHs of the original model generated (black) and simulated (gray) responses. **(B)** The top row shows experimental responses of a cell recorded in the inferior colliculus of an anesthetized gerbil to repeated trials of a variety of sounds (experimental methods are described in Lesica and Grothe, 2008a,b): a sinusoidally amplitude modulated (SAM) tone with a carrier frequency of 6 KHz, a modulation depth of 100%, an intensity of 70 dB, and a modulation frequency of 50 Hz; a square wave modulated tone with the same carrier frequency, modulation depth, and intensity, a 50% duty cycle, and a period of 25 ms; a tone with the same carrier frequency and intensity modulated by a signal with power spectra matched to that measured from a series of animal vocalizations. The bin size Δ for these responses was 1 ms. The middle and bottom rows show the simulated responses simulated and PSTHs presented as in **(A)**.

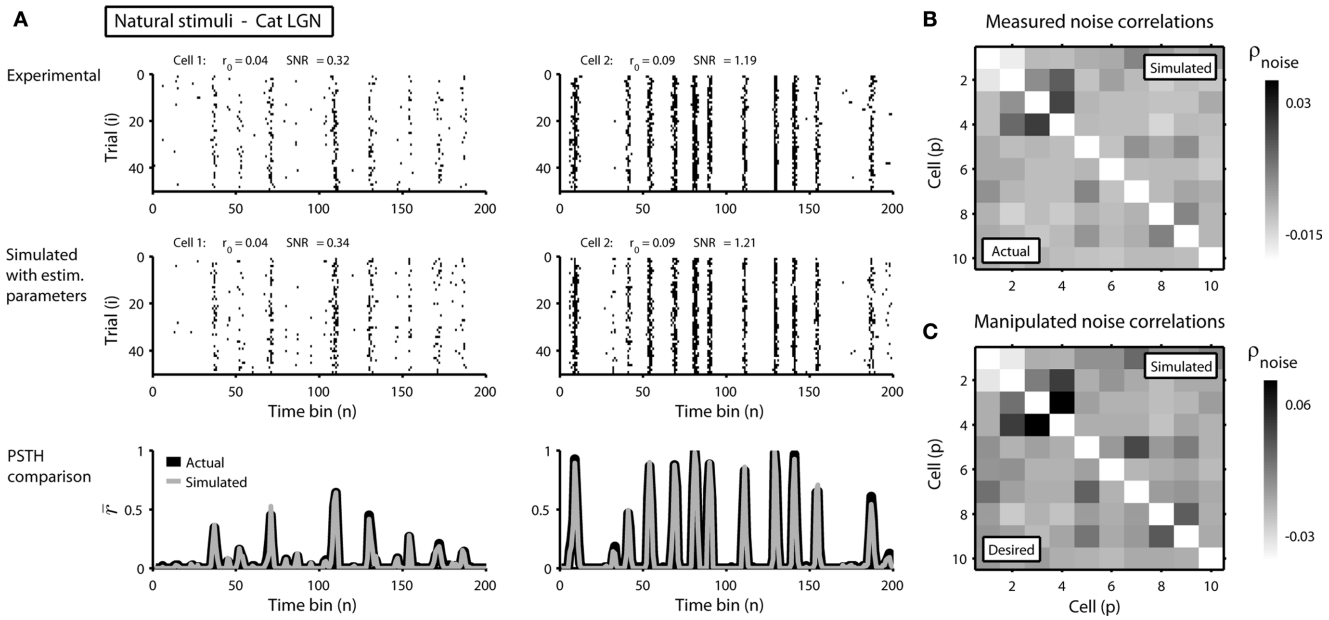


FIGURE 2 | Simulated responses match the single-cell properties and pairwise correlations in experimental data. (A) The top row shows the experimental responses of two cells recorded simultaneously in the lateral geniculate nucleus of an anesthetized cat to repeated trials of a natural scene movie (experimental methods are described in Lesica et al., 2007) with an RMS contrast of 0.4. The bin size Δ for these responses was 4 ms. The middle and bottom rows show the simulated responses and PSTHs presented as in **Figure 1**. **(B)** The image shows the pairwise noise correlations in the experimental (lower triangular portion) and simulated (upper triangular portion) responses of a population of 10 neurons recorded in the cat LGN. The responses of all 10 cells were recorded simultaneously. **(C)** The image shows the desired pairwise noise correlations (lower triangular portion; values are double those measured in the original data) and the pairwise noise correlations realized in the simulated responses (upper triangular portion).

framework. The goodness of fit can be measured by separating the available spike trains into "training" and "testing" sets, fitting the model parameters on the training spike trains, and calculating the (log) likelihood of the testing spike trains from the resulting model. The absolute likelihood may be difficult to interpret, but the ratio of the likelihoods from two different models can give be an informative measure.

To demonstrate the use of likelihood as a measure of goodness of fit, we simulated population spike trains from a known model (see figure legend for model parameters), split the spike trains into training and testing sets, and estimated the model parameters from the training spike trains. To determine whether including noise correlations in the estimated model improved the goodness of fit, we then compared the likelihood of the testing spike trains from the estimated model with and without noise correlations (i.e., with Σ_z estimated as described above or set to the identity matrix) for different numbers of training trials. The likelihood of a given testing spike train was computed as

$$L(\boldsymbol{r}_i) = \sum_{n=1}^{N} \log \boldsymbol{\Phi}_P\left((-2\boldsymbol{r}_i[n]+1)\cdot \boldsymbol{s}[n], \Sigma_z\right) \quad (5)$$

where $\boldsymbol{r}_i$ is the $N \times P$ binary matrix of the responses of a population of P cells on a given trial i, $\boldsymbol{r}_i[n]$ is the vector of the responses $r_i^p[n]$ for each cell, $\boldsymbol{s}[n]$ is the vector of the signals $s^p[n]$ for each cell, $\boldsymbol{\Phi}_P(\vec{x}, \Sigma)$ is the CDF for a P-dimensional Gaussian with zero mean and covariance Σ evaluated at $\vec{x}$, and $\cdot$ denotes a point-by-point vector product. To isolate the effects of the noise correlations on the goodness of fit, we set the PSTHs in the estimated model to be the same as those in the actual model and used only the estimated noise correlations. As shown in **Figure 3**, as the number of training trials increased, the measurement noise in noise correlations

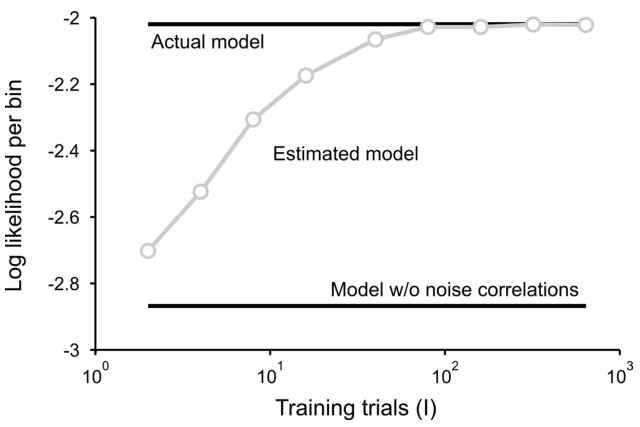


FIGURE 3 | Model goodness of fit increases with increasing trials. The gray circles show the log likelihood (per bin) for the estimated model as a function of the number of trials used for fitting Σ_z. The likelihood was computed for 100 trials not used for fitting the model. The likelihoods for the actual model with and without noise correlations are also shown. Spike trains were generated using the model described in Eq. 10 with the following parameters: $N = 500$, $P = 10$, σ_s^2 and θ were chosen so that $r_0 \sim \mathcal{M}(0.16, 0.04)$ and SNR $\sim \mathcal{M}(0.5, 0.35)$, and ρ_s^{pq} and ρ_z^{pq} were chosen so that $\rho_{\text{signal}}^{pq} \sim \mathcal{N}(0.12, 0.03)$ and $\rho_{\text{noise}}^{pq} \sim \mathcal{N}(0.07, 0.02)$.

decreased, and the likelihood from the estimated model with noise correlations approached that of the actual model, reaching the same value with $I = 80$ trials.

NOISE WITH TEMPORAL CORRELATIONS

The model as described above captures both the instantaneous and long-term signal correlations between cells by matching their individual PSTHs, but captures only instantaneous noise correlations because z is uncorrelated in time. While instantaneous noise correlations are likely to be sufficient to describe population spike trains in early sensory systems, the model can also be extended to capture long-term noise correlations if necessary, for example, to capture the high level of trial-to-trial variability in higher cortical areas. Long-term noise correlations can be captured by adding temporal correlations to z via Gaussian conditioning (MacKay, 2003; Macke et al., 2009) so that z in each time bin is drawn from a distribution with mean and covariance dependent on the values of z in the preceding time bins:

$$z[n+1 \mid n-K+1,...,n] \sim \mathcal{N}\left(CB^{-1}z'[n-K+1,...,n], \Sigma_0 - CB^{-1}C^T\right)$$

$$\text{where} \quad B = \begin{bmatrix} \Sigma_0 & \Sigma_1 & \cdots & \Sigma_{K-1} \\ \Sigma_1 & \Sigma_0 & \cdots & \Sigma_{K-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K-1} & \Sigma_{K-2} & \cdots & \Sigma_0 \end{bmatrix} \tag{6}$$

$$C = \begin{bmatrix} \Sigma_1 & \Sigma_2 & \cdots & \Sigma_K \end{bmatrix}$$

$$\Sigma_k = \begin{bmatrix} \rho_z^{pp}[k] & \rho_z^{pq}[k] \\ \rho_z^{qp}[k] & \rho_z^{qq}[k] \end{bmatrix}$$

$$z'[n-K+1,...,n] = \begin{bmatrix} z^p[n-K+1] & ,z^q[n-K+1],\cdots,z^p[n], & z^q[n] \end{bmatrix}$$

and K is the number of preceding time bins to condition on. An example of such conditioning is shown in **Figure 4**. We used a known model to generate population spike trains with the noise auto- and cross-correlation functions shown by the thick black lines (see figure legend for model parameters). We then estimated model parameters from those spike trains, including those required for conditioning z on the preceding $K = 4$ time bins. The noise auto- and cross-correlation functions for the spike trains simulated from the estimated model (shown by the gray circles) closely match those of the original spike trains.

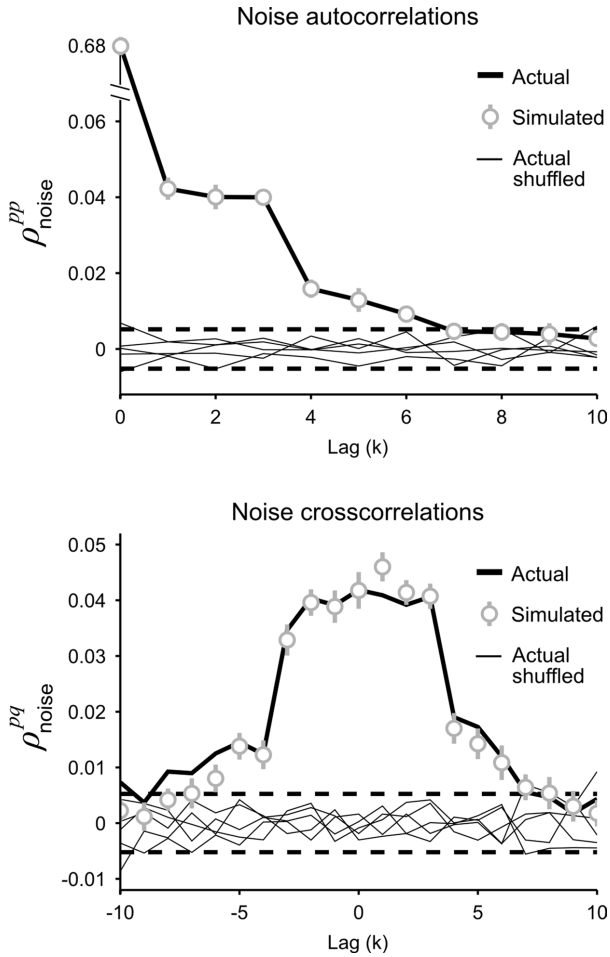


FIGURE 4 | Gaussian conditioning captures long-term noise correlations. The gray circles show the noise auto- and cross-correlation functions for spike trains simulated from the estimated model. Error bars represent the standard deviation in the correlation values computed from independent realizations of the model. The model parameters were estimated from spike trains simulated by the model in Eq. 10, including Gaussian conditioning of z, with the following parameters: $N = 500$, $P = 10$, σ_s^2 and θ were chosen so that $r_0 \sim \mathcal{N}(0.16, 0.04)$ and SNR $\sim \mathcal{N}(0.5, 0.35)$, $\rho_s^{pq} = 0.5$, and $\rho_z^{pp}[k]$ and $\rho_z^{pq}[k] = 0.12$ for $k \leq 4$ and zero thereafter. The noise auto- and cross-correlation functions of the original spike trains are shown by the thick black lines. Examples of the noise auto- and cross-correlation functions of the original spike trains after trial shuffling are shown by the thin black lines. The dashed lines indicate the mean ±2 standard deviations of the distribution of the correlation values after shuffling.

A GENERAL MODEL FOR POPULATION SPIKE TRAINS

SINGLE-CELL RESPONSES

When modeling experimental spike trains as described above, the noise correlations can be chosen arbitrarily, but the mean spike rate, trial-to-trial variability, and signal correlations are dependent on the PSTH. It may also be useful to have a general model for population spike trains in which all of the response properties can be specified independently. For a single cell, this is achieved by replacing the deterministic signal s in the model framework described above with a Gaussian random process:

$$r_i[n] = \begin{cases} 1, & (s[n] + z_i[n]) > \theta \\ 0, & (s[n] + z_i[n]) \leq \theta \end{cases} \tag{7}$$

where $z \sim \mathcal{N}(0, 1)$ is again a Gaussian random process that is different on every trial, s is a Gaussian random process $s \sim \mathcal{N}(0, \sigma_s^2)$ that is the same on every trial, and the threshold θ is allowed to take on any value (note that in this case, θ cannot simply be set to an arbitrary value; in order to achieve any combination of r_0 and SNR, 2 degrees of freedom are required). Such a model could be used, for example, to simulate spike trains with any mean spike rate and trial-to-trial variability. Furthermore, because the model is based on Gaussian processes, it may enable certain population response properties to be investigated analytically or numerically directly from the model parameters, without the need for simulations.

To specify the model parameters, the equations for r_0 and SNR can be written in terms of σ_s^2 and θ and solved numerically to obtain the appropriate values:

$$r_0 = \Phi(-\theta, \sigma_s^2 + 1) \tag{8}$$

$$\mathrm{SNR} = \frac{\mathrm{var}(\bar{r})}{\left\langle \mathrm{var}(\xi_i) \right\rangle_i} \tag{9}$$

While Eq. 8 is already written in terms of σ_s^2 and θ, Eq. 9 requires some manipulation. The numerator can be written as

$$\mathrm{var}(\bar{r}) = \mathrm{var}\left(\frac{1}{I}\sum_{i=1}^{I} r_i\right) = \frac{1}{I^2}\left(I\left\langle \mathrm{var}(r_i)\right\rangle_i + 2I(I-1)\left\langle \mathrm{cov}(r_i, r_j)\right\rangle_{i\neq j}\right)$$

where, because r_i is binary,

$$\left\langle \mathrm{var}(r_i)\right\rangle_i = r_0(1-r_0)$$

and, because s and z are Gaussian,

$$\left\langle \mathrm{cov}(r_i, r_j)\right\rangle_{i\neq j} = \Phi_2\left(-\theta, \begin{bmatrix} \sigma_s^2+1 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2+1 \end{bmatrix}\right) - r_0^2.$$

The denominator can be written as

$$\left\langle \mathrm{var}(\xi)\right\rangle_i = \left\langle \mathrm{var}(\bar{r} - r_i)\right\rangle_i = \mathrm{var}(\bar{r}) + \left\langle \mathrm{var}(r_i)\right\rangle_i - 2\left\langle \mathrm{cov}(\bar{r}, r_i)\right\rangle_i$$

where, because r_i is binary and s and z are Gaussian,

$$\begin{aligned}\left\langle \mathrm{cov}(\bar{r}, r_i)\right\rangle_i &= \left\langle \bar{r}\cdot r_i\right\rangle_{n,i} - r_0^2 = \left\langle \bar{r}^2\right\rangle_n - r_0^2 = \left\langle \left(\frac{1}{I}\sum_{i=1}^{I} r_i\right)^2\right\rangle_n - r_0^2 \\ &= \frac{1}{I^2}\left(I\left\langle r_i^2\right\rangle_{n,i} + I(I-1)\left\langle r_i\cdot r_j\right\rangle_{n,i\neq j}\right) - r_0^2 \\ &= \frac{1}{I^2}\left(Ir_0^2 + I(I-1)\Phi_2\left(-\theta, \begin{bmatrix} \sigma_s^2+1 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2+1 \end{bmatrix}\right)\right) - r_0^2\end{aligned}$$

Note that in these equations, $\bar{r}\cdot r_i$ and $r_i \cdot r_j$ denotes point-by-point vector products, and $\bar{r}^2$ denotes point-by-point squaring. Thus, for any realizable combination of r_0 and SNR, appropriate σ_s^2 and θ can be found (the minimum realizable SNR depends on the number of trials, see Appendix). To demonstrate this approach, we randomly chose a variety of values for r_0 and SNR, estimated the corresponding values of σ_s^2 and θ, and generated responses using the estimated values. As shown in **Figure 5A**, r_0 and SNR of the simulated responses closely match the desired values.

POPULATION RESPONSES

The model described above for a single cell is easily extended to a population, where the response of cell $p \in \{1, 2, \ldots, P\}$ is given by

$$r_i^p[n] = \begin{cases} 1, & \left(s^p[n] + z_i^p[n]\right) > \theta^p \\ 0, & \left(s^p[n] + z_i^p[n]\right) \leq \theta^p \end{cases} \tag{10}$$

where both s and z are multivariate Gaussian random process $s \sim \mathcal{N}(0, \Sigma_s)$ and $z \sim \mathcal{N}(0, \Sigma_z)$ with covariance matrices

$$\Sigma_s = \begin{bmatrix} \sigma_{s^1}^2 & \rho_s^{12}\sigma_{s^2}\sigma_{s^1} & \cdots & \rho_s^{1P}\sigma_{s^P}\sigma_{s^1} \\ \rho_s^{21}\sigma_{s^1}\sigma_{s^2} & \sigma_{s^2}^2 & & \\ \vdots & & \ddots & \\ \rho_s^{P1}\sigma_{s^1}\sigma_{s^P} & & & \sigma_{s^P}^2 \end{bmatrix} \text{ and}$$

$$\Sigma_z = \begin{bmatrix} 1 & \rho_z^{12} & \cdots & \rho_z^{1P} \\ \rho_z^{21} & 1 & & \\ \vdots & & \ddots & \\ \rho_z^{P1} & & & 1 \end{bmatrix}.$$

After determining σ_s^2 and θ for each cell based on the desired r_0 and SNR as described above, the pairwise correlation coefficients ρ_s^{pq} and ρ_z^{pq} required to obtain the desired spike train signal and noise correlations $\rho_{\mathrm{signal}}^{pq}$ and $\rho_{\mathrm{noise}}^{pq}$ can be found by solving the following equations numerically:

$$\rho_{\mathrm{signal}}^{pq} = \frac{\left\langle \mathrm{cov}\left(r_i^p, r_j^q\right)\right\rangle_{i\neq j}}{\sqrt{\left\langle \mathrm{var}\left(r_i^p\right)\right\rangle_i \left\langle \mathrm{var}\left(r_i^q\right)\right\rangle_i}} \quad \text{and}$$

$$\rho_{\mathrm{noise}}^{pq} = \rho_{\mathrm{total}}^{pq} - \rho_{\mathrm{signal}}^{pq} = \frac{\left\langle \mathrm{cov}\left(r_i^p, r_i^q\right)\right\rangle_i}{\sqrt{\left\langle \mathrm{var}\left(r_i^p\right)\right\rangle_i \left\langle \mathrm{var}\left(r_i^q\right)\right\rangle_i}} - \frac{\left\langle \mathrm{cov}\left(r_i^p, r_j^q\right)\right\rangle_{i\neq j}}{\sqrt{\left\langle \mathrm{var}\left(r_i^p\right)\right\rangle_i \left\langle \mathrm{var}\left(r_i^q\right)\right\rangle_i}}$$

where, because r_i is binary,

$$\left\langle \mathrm{var}\left(r_i^p\right)\right\rangle_i = r_0^p\left(1 - r_0^p\right)$$

and, because s and z are Gaussian,

$$\left\langle \mathrm{cov}\left(r_i^p, r_j^q\right)\right\rangle_{i\neq j} = \Phi_2\left(-\begin{bmatrix}\theta^p \\ \theta^q\end{bmatrix}, \begin{bmatrix} \sigma_{s^p}^2+1 & \rho_s^{pq}\sigma_{s^p}\sigma_{s^q} \\ \rho_s^{pq}\sigma_{s^p}\sigma_{s^q} & \sigma_{s^q}^2+1 \end{bmatrix}\right) - r_0^p r_0^q$$

and

$$\left\langle \mathrm{cov}\left(r_i^p, r_i^q\right)\right\rangle_i = \Phi_2\left(-\begin{bmatrix}\theta^p \\ \theta^q\end{bmatrix}, \begin{bmatrix} \sigma_{s^p}^2+1 & \rho_s^{pq}\sigma_{s^p}\sigma_{s^q} + \rho_z^{pq} \\ \rho_s^{pq}\sigma_{s^p}\sigma_{s^q} + \rho_z^{pq} & \sigma_{s^q}^2+1 \end{bmatrix}\right) - r_0^p r_0^q$$

Again, the functions are monotonic and each ρ_s^{pq} and ρ_z^{pq} can be solved for independently.

To demonstrate this approach, we generated a random set of pairwise signal and noise correlation coefficients $\rho_{\mathrm{signal}}^{pq}$ and $\rho_{\mathrm{noise}}^{pq}$, estimated the corresponding values of ρ_s^{pq} and ρ_z^{pq}, and simulated population spike trains with these values. As shown in **Figures 5B,C**, the correlations in the simulated spike trains closely matched the desired values.

DISCUSSION

We have described a model for simulating population spike trains typical of early sensory systems. The model has two forms: the first requires the specification of PSTHs and noise correlations and

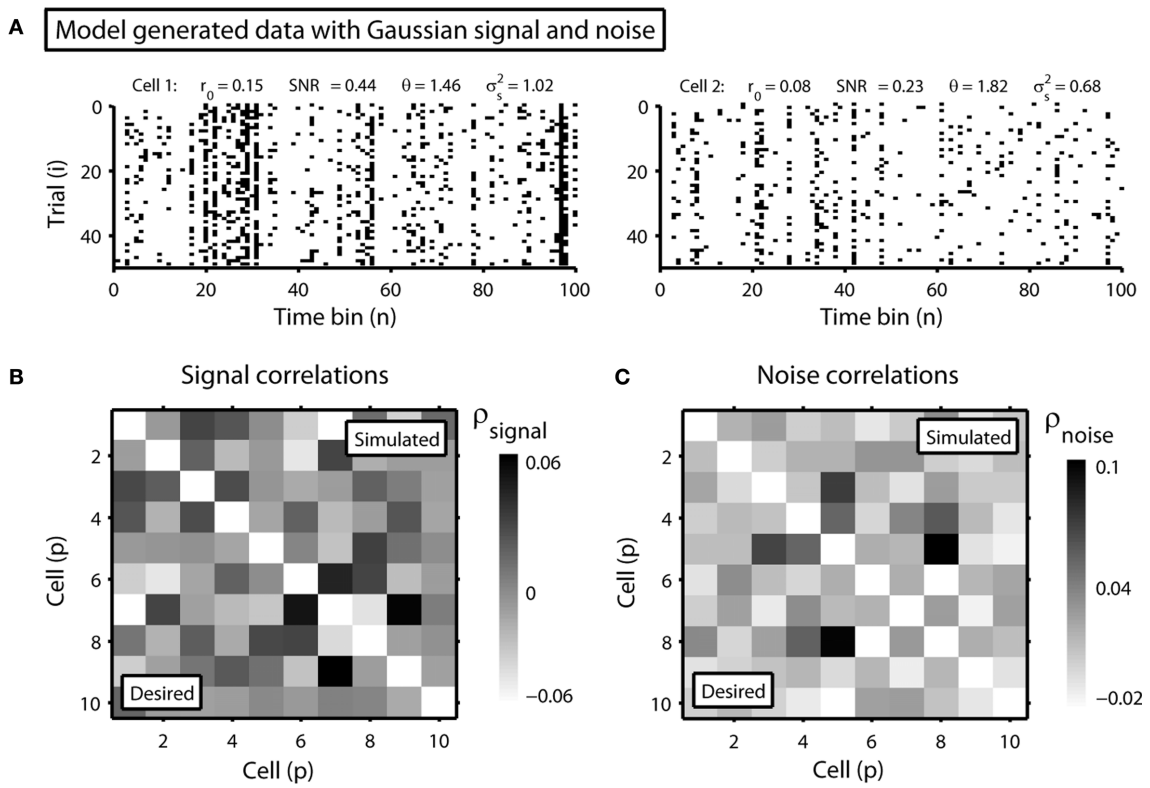


FIGURE 5 | Simulated population responses with specified single-cell properties and pairwise correlations. (A) The top row shows the simulated responses for two cells with specified r_0 and SNR. **(B)** The image shows the desired pairwise signal correlations (lower triangular portion) and the pairwise signal correlations realized in the simulated responses (upper triangular portion) for a population of 10 cells. **(C)** The desired and realized pairwise noise correlations for a population of 10 cells presented as in **(B)**.

can be used to match and manipulate experimental data, and the second is more general and allows for population spike trains with any mean spike rates, trial-to-trial variabilities, signal correlations, and noise correlations. Both forms of the model are easily implemented as parameter fitting requires simply finding the level crossings of monotonic functions and correlations can be determined independently for each pair of cells. The Matlab code required to fit the model parameters is available for download at http://www.ucl.ac.uk/ear/research/lesicalab.

Our model improves on the existing methods for generating population spike trains described in the Introduction in several important ways. First, the model framework is explicitly designed around the response properties that are important for early sensory neurons: time-varying spike rate (PSTH), trial-to-trial variability, and signal and noise correlations. Second, the model allows independent and straightforward manipulation of one response property without changes in the other properties. One can imagine a number of potential uses for a model with these properties. The fact that the model matches the single-cell properties and correlations observed experimentally is in itself of some utility, such as providing a simple framework for computing the likelihood of observed spike trains given only pairwise interactions. These likelihoods could be used to, for example, test how important noise correlations are in determining population spike patterns by comparing models with and without noise correlations. The model also provides the ability to manipulate noise correlations without affecting the signal correlations or single-cell properties. In the brain, these properties are coupled to each other – for example, one can decrease the spike rate of visual neurons by decreasing the contrast of the stimulus, but this will also likely change the trial-to-trial variability and the correlations. Thus, a question such as whether or not changes in correlations with changes in contrast are detrimental or beneficial to a population code is impossible to answer experimentally. With our model, one could compare simulated populations with high contrast single-cell properties and correlations to simulated populations with high contrast single-cell properties and low contrast correlations to directly test whether or not the change in correlations is important. A similar example can be used to illustrate the utility of the general form of the model: Because the general form of the model allows for spike trains with any mean spike rate, trial-to-trial variability, and pairwise signal and noise correlations (within statistical constraints), it could be used to perform a systematic investigation of the effects of noise correlations on populations with different levels of signal correlations that would be impossible to conduct experimentally.

There are several ways in which the formulation of our model described here could potentially be improved. For example, the assumption that no more than one spike can occur in any time

bin could potentially be relaxed, using a formulation analogous to that described in (Macke et al., 2009). Also, the current formulation of the model cannot be directly connected to biophysical quantities, so its parameters cannot be used, for example, to determine the current necessary to achieve a desired set of correlations in a stimulation experiment. With further effort, however, it may be possible to extend some of the desirable properties of our model into a more realistic framework such as an integrate-and-fire model (Paninski et al., 2004). Another limitation of the model that may be difficult to overcome is that in estimating a single value for each parameter across all trials, it implicitly assumes that the population is stationary. Certain types of non-stationarities, such as trial-to-trial fluctuations in the PSTH, may be captured by introducing long-term noise correlations via Gaussian conditioning, but others may require integrating the model into an adaptive framework (Eden et al., 2004).

There are also certain limitations of the model that are inherent in its underlying statistical framework. For example, it is difficult to generate spike trains with very low SNRs with a small number of trials – even if the variance of the internal signal in the model $\sigma_s^2 = 0$, because the spikes are generated stochastically, the PSTH computed from a small number of trials will not have zero variance, and, thus, the SNR will not be zero. There are also constraints on the values of signal and noise correlations that can be achieved. Because the model is based on multivariate Gaussian processes which must have positive semi-definite covariance matrices, it is difficult, for example, to generate spike trains with strong negative correlations. Finally, while the model can be specified and fit to a population of any size, it is only guaranteed to capture second-order correlations. As with any model that includes only pairwise interactions, our model's ability to capture the higher-order structure in population spike trains will depend on the nature of the population (Schneidman et al., 2006; Shlens et al., 2006; Roudi et al., 2009; Ohiorhenuan et al., 2010).

ACKNOWLEDGMENTS

We thank Benedikt Grothe for the use of the gerbil data and Chong Weng, Jianzhong Jin, Chun-I Yeh, Dan Butts, Garrett Stanley, and Jose-Manuel Alonso for the use of the cat data. Dmitry R. Lyamzin and Nicholas A. Lesica were supported by the Deutsche Forschungsgemeinschaft (LE2522/1-1) and the Wellcome Trust. Jakob H. Macke was supported by the German Ministry of Education, Science, Research and Technology through the Bernstein award to Matthias Bethge (BMBF; FKZ: 01GQ0601).

REFERENCES

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366.

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

Brenner, N., Strong, S. P., Koberle, R., Bialek, W., de Ruyter van Steveninck, R. R. (2000). Synergy in a neural code. *Neural Comput.* 12, 1531–1552.

Brette, R. (2009). Generation of correlated spike trains. *Neural Comput.* 21, 188–215.

Chornoboy, E., Schramm, L., and Karr, A. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybern.* 59, 265–275.

Cox, D. R., and Wermuth, N. (2002). On some models for multivariate binary variables parallel in complexity with the multivariate Gaussian distribution. *Biometrika* 89, 462–469.

De La Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* 448, 802–806.

Destexhe, A., and Pare, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *J. Neurophysiol.* 81, 1531.

Dorn, J. D., and Ringach, D. L. (2003). Estimating membrane voltage correlations from extracellular spike trains. *J. Neurophysiol.* 89, 2271–2278.

Eden, U. T., Frank, L. M., Barbieri, R., Solo, V., and Brown, E. N. (2004). Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Comput.* 16, 971–998.

Emrich, L. J., and Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *Am. Stat.* 45, 302–304.

Feng, J., and Brown, D. (2000). Impact of correlated inputs on the output of the integrate-and-fire model. *Neural Comput.* 12, 671–692.

Galan, R. F., Fourcaud-Trocme, N., Ermentrout, G. B., and Urban, N. N. (2006). Correlation-induced synchronization of oscillations in olfactory bulb neurons. *J. Neurosci.* 26, 3646.

Gutig, R., Aharonov, R., Rotter, S., and Sompolinsky, H. (2003). Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *J. Neurosci.* 23, 3697.

Gutnisky, D. A., and Josic, K. (2010). Generation of spatiotemporally correlated spike trains and local field potentials using a multivariate autoregressive process. *J. Neurophysiol.* 103, 2912–2930.

Krumin, M., and Shoham, S. (2009). Generation of spike trains with controlled auto-and cross-correlation functions. *Neural Comput.* 21, 1642–1664.

Kuhn, A., Aertsen, A., and Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Comput.* 15, 67–101.

Kulkarni, J. E., and Paninski, L. (2007). Common-input models for multiple neural spike-train data. *Network* 18, 375–407.

Lesica, N. A., and Grothe, B. (2008a). Efficient temporal processing of naturalistic sounds. *PLoS ONE* 3, e1655. doi: 10.1371/journal.pone.0001655.

Lesica, N. A., and Grothe, B. (2008b). Dynamic spectrotemporal feature selectivity in the auditory midbrain. *J. Neurosci.* 28, 5412–5421.

Lesica, N. A., Jin, J., Weng, C., Yeh, C. I., Butts, D. A., Stanley, G. B., and Alonso, J. M. (2007). Adaptation to stimulus contrast and correlations during natural visual stimulation. *Neuron* 55, 479–491.

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.

Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S., and Bethge, M. (2009). Generating spike trains with specified correlation coefficients. *Neural Comput.* 21, 397–423.

Niebur, E. (2007). Generation of synthetic spike trains with defined pairwise correlations. *Neural Comput.* 19, 1720–1738.

Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* 466, 617–621.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15, 243–262.

Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog. Brain Res.* 165, 493–507.

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, 2533–2561.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995.

Roudi, Y., Nirenberg, S., and Latham, P. E. (2009). Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS Comput. Biol.* 5, e1000380. doi: 10.1371/journal.pcbi.1000380.

Salinas, E., and Sejnowski, T. J. (2002). Integrate-and-fire neurons driven by correlated stochastic input. *Neural Comput.* 14, 2111–2155.

Schneidman, E., Berry, M. J. II, R. S., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007.

Shea-Brown, E., Josic, K., De La Rocha, J., and Doiron, B. (2008). Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Phys. Rev. Lett.* 100, 108102.

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M., and Chichilnisky, E. J. (2006). The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* 26, 8254–8266.

Song, S., and Abbott, L. (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron* 32, 339–350.

Stroeve, S., and Gilen, S. (2001). Correlation between uncoupled conductance-based integrate-and-fire neurons due to common and synchronous presynaptic firing. *Nat. Neurosci.* 13, 2005–2029.

Tchumatchenko, T., Malyshev, A., Geisel, T., Volgushev, M., and Wolf, F. (2008). Correlations and synchrony in threshold neuron models. *Quant. Biol. arXiv* 810, 2–6.

Toyoizumi, T., Rad, K. R., and Paninski, L. (2009). Mean-field approximations for coupled populations of generalized linear model spiking neurons with Markov refractoriness. *Neural Comput.* 21, 1203–1243.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 November 2009; paper pending published: 19 January 2010; accepted: 05 October 2010; published online: 15 November 2010.

Citation: Lyamzin DR, Macke JH and Lesica NA (2010) Modeling population spike trains with specified time-varying spike rates, trial-to-trial variability, and pairwise signal and noise correlations. Front. Comput. Neurosci. ***4****:144. doi: 10.3389/fncom.2010.00144*

APPENDIX

MINIMUM ACHIEVABLE SNR

For any binary response r with N time bins and I trials, the minimum realizable SNR depends only on the number of trials. Recalling the definition of SNR

$$\mathrm{SNR} = \frac{\mathrm{var}(\bar{r})}{\left\langle \mathrm{var}(\xi_i) \right\rangle_i}$$

the minimum value within the context of our model framework is clearly achieved when the variance of the signal $\sigma_s^2 = 0$ (in the case of $I = \infty$, this results in $\mathrm{var}(\bar{r}) = 0$). Recalling the expansion of $\mathrm{var}(\bar{r})$ in section "Single Cell Responses" and the fact that $r_0 = \Phi(-\theta, \sigma_s^2 + 1)$ within our framework, we find that when $\sigma_s^2 = 0$,

$$\begin{aligned}
\mathrm{var}(\bar{r}) &= \frac{1}{I^2}\left(I\left\langle \mathrm{var}\left(r_i\right)\right\rangle_i + 2I(I-1)\left\langle \mathrm{cov}\left(r_i, r_j\right)\right\rangle_{i\neq j}\right) \\
&= \frac{1}{I^2}\left(Ir_0(1-r_0) + 2I(I-1)\Phi_2\left(-\theta, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) - r_0^2\right) \\
&= \frac{1}{I^2}\left(Ir_0(1-r_0)\right)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\left\langle \mathrm{var}(\xi)\right\rangle_i &= \left\langle \mathrm{var}\left(\bar{r} - r_i\right)\right\rangle_i \\
&= \mathrm{var}(\bar{r}) + \left\langle \mathrm{var}\left(r_i\right)\right\rangle_i - 2\left\langle \mathrm{cov}\left(\bar{r}, r_i\right)\right\rangle_i \\
&= \frac{1}{I^2}\left(Ir_0\left(1-r_0\right)\right) + r_0(1-r_0) \\
&\quad - 2\left(\frac{1}{I^2}\left(Ir_0^2 + I(I-1)\Phi_2\left(-\theta, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\right) - r_0^2\right) \\
&= \frac{1}{I^2}\left(Ir_0\left(1-r_0\right)\right) + r_0\left(1-r_0\right)
\end{aligned}$$

Simplifying the resulting expression gives the minimum realizable SNR:

$$\mathrm{SNR} = \frac{\frac{1}{I^2}\left(Ir_0\left(1-r_0\right)\right)}{\frac{1}{I^2}\left(Ir_0\left(1-r_0\right)\right) + r_0\left(1-r_0\right)} = \frac{1}{(I+1)}$$

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 19 November 2010
doi: 10.3389/fncom.2010.00147

Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input

Michael Krumin, Inna Reutsky and Shy Shoham*

Faculty of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa, Israel

Edited by:
Jakob H. Macke, University College London, UK

Reviewed by:
Taro Toyoizumi, RIKEN Brain Science Institute, Japan
Kresimir Josic, University of Houston, USA

****Correspondence:***
Shy Shoham, Faculty of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel.
e-mail: sshoham@bm.technion.ac.il

The correlation structure of neural activity is believed to play a major role in the encoding and possibly the decoding of information in neural populations. Recently, several methods were developed for exactly controlling the correlation structure of multi-channel synthetic spike trains (Brette, 2009; Krumin and Shoham, 2009; Macke et al., 2009; Gutnisky and Josic, 2010; Tchumatchenko et al., 2010) and, in a related work, correlation-based *analysis* of spike trains was used for blind identification of single-neuron models (Krumin et al., 2010), for identifying compact auto-regressive models for multi-channel spike trains, and for facilitating their causal network analysis (Krumin and Shoham, 2010). However, the diversity of correlation structures that can be explained by the feed-forward, non-recurrent, generative models used in these studies is limited. Hence, methods based on such models occasionally fail when analyzing correlation structures that are observed in neural activity. Here, we extend this framework by deriving closed-form expressions for the correlation structure of a more powerful multivariate self- and mutually exciting Hawkes model class that is driven by exogenous non-negative inputs. We demonstrate that the resulting Linear–Non-linear-Hawkes (LNH) framework is capable of capturing the dynamics of spike trains with a generally richer and more biologically relevant multi-correlation structure, and can be used to accurately estimate the Hawkes kernels or the correlation structure of external inputs in both simulated and real spike trains (recorded from visually stimulated mouse retinal ganglion cells). We conclude by discussing the method's limitations and the broader significance of strengthening the links between neural spike train analysis and classical system identification.

Keywords: spike train analysis, linear system identification, point process, recurrent, multi-channel recordings, correlation functions, integral equations, retinal ganglion cells

INTRODUCTION

Linear system models enjoy a fundamental role in the analysis of a wide range of natural and engineered signals and processes (Kailath et al., 2000). Hawkes (Hawkes, 1971a,b; cf. Johnson, 1996) introduced the basic point processes equivalent of the linear auto-regressive and multi-channel auto-regressive process models, and derived expressions for their output correlations and spectral densities. The Hawkes model was later used as a model for neural activity in small networks of neurons (Brillinger, 1975, 1988; Brillinger et al., 1976; Chornoboy et al., 1988), where maximum likelihood (ML) parameter estimation procedures can be used to estimate the synaptic strengths between connected neurons, but where no external modulating processes were considered. Interestingly, the recent renaissance of interest in explicit modeling and model-based analysis of neural spike trains (e.g., Brown et al., 2004; Paninski et al., 2007; Stevenson et al., 2008), has largely disregarded the Hawkes-type models, focusing instead on their non-linear generalizations: the generalized linear models (GLMs), and related multiplicative models (Cardanobile and Rotter, 2010). GLMs are clearly powerful and flexible models of spiking processes, and are also related to the popular Linear–Non-linear encoding models (Chichilnisky, 2001; Paninski et al., 2004; Shoham et al., 2005). However, they do not enjoy the same level of mathematical simplicity as their Hawkes counterparts – only approximate analytical expressions for the correlation and the spectral properties of a GLM model were derived (Nykamp, 2007; Toyoizumi et al., 2009) under fairly restrictive conditions, while exact parameters for detailed, heterogeneous GLM models can only be evaluated numerically (Pillow et al., 2008).

The significance and applications of spike train models with closed-form expressions for the output correlation/spectral structure have begun to emerge in a number of recent studies. These include: (1) the ability to generate synthetic spike trains with a given auto- and cross-correlation structure (Brette, 2009; Krumin and Shoham, 2009; Macke et al., 2009; Gutnisky and Josic, 2010); (2) the ability to identify neural input-output encoding models "blindly" by analyzing the spectral and correlation distortions they induce (Krumin et al., 2010); (3) the ability to fit compact multivariate auto-regressive (MVAR) models to multi-channel neural spike trains (Krumin and Shoham, 2010); and (4) the ability to apply the associated powerful framework of Granger causality analysis (Granger, 1969; Krumin and Shoham, 2010). These early studies relied on the analysis of tractable non-linear spiking models such as threshold models (Macke et al., 2009; Gutnisky and Josic, 2010; Tchumatchenko et al., 2010) or the Linear–Non-linear-Poisson (LNP) models (Krumin and Shoham, 2009) driven by Gaussian input processes.

In this paper we revisit the Hawkes model within this new emerging framework for correlation-based, closed-form identification and analysis of spike trains models. The framework is

thereby extended from the exclusive treatment of feed-forward models to treating more general and neuro-realistic (yet analytically tractable) models that also include feedback terms. In Section "Methods" we begin by reviewing some basic results for the correlation structure of the classical, homogenous (constant input) single and multivariate Hawkes model, derive new integral equations for the correlation structure of a Hawkes model driven by a *time-varying* (*inhomogeneous*) stationary random non-negative process input (see **Figure 1**), and propose a numerical method for solving them. In Section "Results," we present the results of applying these methods to real neural recordings from isolated mouse retina, and the required methodological adaptations. We conclude with a discussion in Section "Discussion."

METHODS

In this section we begin by defining the Hawkes model, recalling its auto-correlation structure and then generalizing to multivariate (mutually exciting) non-homogeneous Hawkes model of point processes. Next, we propose a method for the solution of the resulting equations, and for the estimation of the different parameters of the model. In the final subsection the experimental methods of stimulation and data acquisition are presented.

THEORETICAL BACKGROUND

Let us consider the intensity of a self-exciting point process to be defined by the following expression:

$$\mu(t) = \lambda + \sum_k g(t - t_k) \tag{1}$$

Here, the instantaneous firing intensity $\mu(t)$ is the exogenous input λ summed together with multiple shifted replicas of the self-excitation kernel $g(t)$. The kernels are causal ($g(t) = 0, t < 0$), and t_k represents all the past spike-times. For technical reasons we will write the expression using the Stieltjes integral:

$$\mu(t) = \lambda + \int_{-\infty}^{t} g(t-u)dN(u) \tag{2}$$

where $N(t)$ is the counting process (number of spikes up to time t). The sum term in Eq. 1 is now replaced by a convolution of the spiking history with a linear kernel. The mean firing rate (denoted throughout the paper by $\langle dN \rangle$) of this point process is given by:

$$\langle dN \rangle \triangleq \mathbb{E}\left\{\frac{dN(t)}{dt}\right\} = \mathbb{E}\left\{\lambda + \int_{-\infty}^{t} g(t-u)dN(u)\right\} =$$

$$= \lambda + \int_{-\infty}^{t} g(t-u)\mathbb{E}\left\{\frac{dN(u)}{du}\right\}du = \lambda + \langle dN \rangle \cdot \int_0^{\infty} g(u)du \tag{3}$$

Resulting in:

$$\langle dN \rangle = \frac{\lambda}{1 - \int_0^{\infty} g(u)du} \tag{4}$$

The stability (and stationarity) condition for this model ($\int_0^{\infty} g(u)du < 1$) can easily be inferred from this equation. An expression for the auto-covariance function of such a point process was derived in Hawkes (1971a), and we will briefly review here the main results (adapted from his auto-covariance notation into auto-correlation function notation used here for simplicity). We will distinguish between two different auto-correlation functions, the first:

$$\tilde{R}_{dN}(\tau) \triangleq \frac{\mathbb{E}\{dN(t+\tau)dN(t)\}}{dt^2}, \tag{5}$$

which has a delta function singularity $\langle dN \rangle \cdot \delta(\tau)$ at $\tau = 0$ due to the nature of point processes, and the second:

$$R_{dN}(\tau) \triangleq \tilde{R}_{dN}(\tau) - \langle dN \rangle \cdot \delta(\tau), \tag{6}$$

from which this singularity was subtracted.

Using these definitions we get the following integral equation for the auto correlation of the output point process of the Hawkes model:

$$R_{dN}(\tau) = \lambda \cdot \langle dN \rangle + g(\tau) \cdot \langle dN \rangle + \int_{-\infty}^{\tau} g(\tau - u)R_{dN}(u)du \tag{7}$$

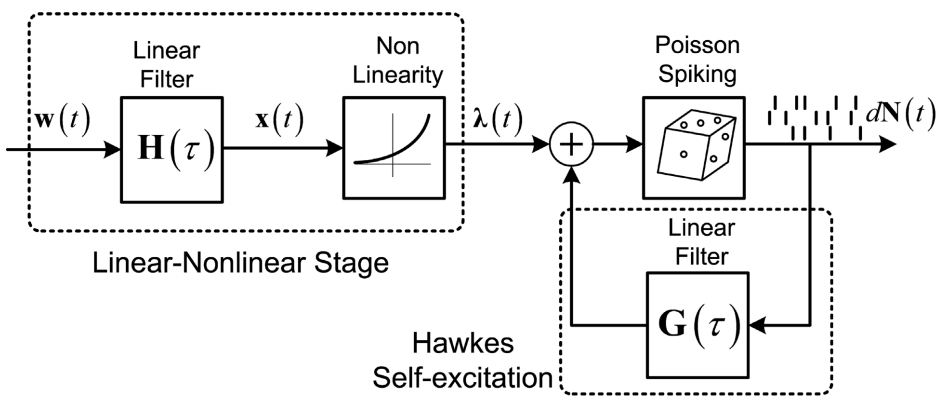


FIGURE 1 | Linear–Non-linear-Hawkes model diagram. White multivariate Gaussian noise $\mathbf{w}(t)$ passes through a Linear–Non-linear cascade, resulting in an exogenous input, $\boldsymbol{\lambda}(t)$, to the Hawkes model. By setting the Hawkes self- and mutual-excitation feedback filter to equal zero we remain with a multivariate Linear–Non-linear-Poisson (LNP) model. By setting $\boldsymbol{\lambda}(t) = \boldsymbol{\lambda}$ we get the Hawkes model.

This equation can be solved numerically (Mayers, 1962) or by using Wiener–Hopf related techniques (Noble, 1958; Hawkes, 1971b).

Similarly, Hawkes (1971a) generalized this solution (Eqs 4 and 7) to *multivariate* mutually exciting point processes by using matrix notation. The intensity of mutually exciting process becomes:

$$\boldsymbol{\mu}(t) = \boldsymbol{\lambda} + \int_{-\infty}^{t} \mathbf{G}(t-u)d\mathbf{N}(u) \tag{8}$$

with mean firing rates:

$$\langle d\mathbf{N} \rangle = \left(\mathbf{I} - \int_{0}^{\infty} \mathbf{G}(u)du \right)^{-1} \cdot \boldsymbol{\lambda} \tag{9}$$

and the cross-correlation matrix as a solution of:

$$\mathbf{R}_{d\mathbf{N}}(\tau) = \boldsymbol{\lambda} \cdot \langle d\mathbf{N} \rangle^{T} + \mathbf{G}(\tau) \cdot diag\left(\langle d\mathbf{N} \rangle\right) + \int_{-\infty}^{\tau} \mathbf{G}(\tau-u)\mathbf{R}_{d\mathbf{N}}(u)du \tag{10}$$

THE LINEAR–NON-LINEAR-HAWKES MODEL AND ITS CORRELATIONS

Let us now consider a more general case of a non-homogeneous Hawkes model, where the exogenous input $\lambda(t)$ can be a time-varying (stationary) process:

$$\mu(t) = \lambda(t) + \int_{-\infty}^{t} g(t-u)dN(u) \tag{11}$$

For example, this class of models includes the important special case (**Figure 1**) where $\lambda(t)$ is itself a non-negative stationary random process generated by a Linear–Non-linear cascade acting on a Gaussian process input (possibly a stimulus). Note the difference between the proposed linear–non-linear-Hawkes (LNH) model and the GLM-type models, in which the feedback term is summed with the $\boldsymbol{x}(t)$ and not with the $\boldsymbol{\lambda}(t)$ (according to the notation in **Figure 1**). This effectively changes the locus of the non-linearity present in the model and affects the model's properties and analytical tractability.

The mean firing rate of this point process can, in general, be found in a similar way as in Eqs 3 and 4:

$$\langle dN \rangle = \frac{\mathbb{E}\{\lambda(t)\}}{1 - \int_{0}^{\infty} g(u)du} \tag{12}$$

Next, the auto-correlation function $R_{dN}(\tau)$ of this process can be derived using a similar procedure to the derivation of Eq. 7 (the detailed derivation can be found in Section "Correlation Structure of the LNH Model" of Appendix). This time, the auto-correlation function is governed by two coupled integral equations:

$$R_{dN}(\tau) = R_{\lambda dN}(\tau) + g(\tau) \cdot \langle dN \rangle + \int_{-\infty}^{\tau} g(\tau-u)R_{dN}(u)du$$

$$R_{\lambda dN}(\tau) = R_{\lambda}(\tau) + \int_{\tau}^{\infty} g(u-\tau)R_{\lambda dN}(u)du \tag{13}$$

These two equations provide the solution for the output auto-correlation function $R_{dN}(\tau)$ and for the cross-correlation $R_{\lambda dN}(\tau) \triangleq \mathbb{E}\{\lambda(t+\tau)(dN(t)/dt)\}$ between the exogenous input $\lambda(t)$ and the point process whose intensity is defined by Eq. 11. Here, the input auto-correlation function $R_{\lambda}(\tau)$ and the self-exciting kernel $g(\tau)$ serve as given parameters (see also Identification of the LNH Model).

Equations 12 and 13 can be further generalized to a multivariate case (mutually exciting point processes), and be written using the matrix notation:

$$\langle d\mathbf{N} \rangle = \left(\mathbf{I} - \int_{0}^{\infty} \mathbf{G}(u)du \right)^{-1} \cdot \mathbb{E}\{\boldsymbol{\lambda}(t)\}$$

$$\mathbf{R}_{d\mathbf{N}}(\tau) = \mathbf{R}_{\boldsymbol{\lambda} d\mathbf{N}}(\tau) + \mathbf{G}(\tau) \cdot diag\left(\langle d\mathbf{N} \rangle\right) + \int_{-\infty}^{\tau} \mathbf{G}(\tau-u)\mathbf{R}_{d\mathbf{N}}(u)du$$

$$\mathbf{R}_{\boldsymbol{\lambda} d\mathbf{N}}(\tau) = \mathbf{R}_{\boldsymbol{\lambda}}(\tau) + \int_{\tau}^{\infty} \mathbf{R}_{\boldsymbol{\lambda} d\mathbf{N}}(u)\mathbf{G}^{T}(u-\tau)du \tag{14}$$

Note that for constant $\boldsymbol{\lambda}$ these equations are reduced to Eqs 9 and 10.

IDENTIFICATION OF THE LNH MODEL

The equations for the correlation structure of a single self-exciting point process and multivariate mutually exciting point processes (Eqs 13 and 14 respectively) can be solved numerically by switching from continuous time integral notation to discrete time matrix notation, and consequently performing matrix calculations. The integration operations in the Eqs 13 and 14 are thus converted to matrix multiplication operations. This allows a simple and straightforward way to solve the equations for the output correlation structure. Here, we only briefly present the main results. All the detailed explanations on the notation used, on how the appropriate matrices and vectors are built, and how the equations are solved in both single- and multi-channel cases can be found in Section "Solution of the Integral Equations" of Appendix. Using the new notation the output correlation is estimated by:

$$\underline{\mathbf{R}}_{d\mathbf{N}} = \left(\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_{2} \right)^{-1} \cdot \left(\underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}} + \underline{\mathbf{G}} \cdot diag\left(\langle d\mathbf{N} \rangle\right) \right),$$

$$\underline{\mathbf{R}}^{T}_{\boldsymbol{\lambda} d\mathbf{N}} = \left(\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_{1} \right)^{-1} \cdot \underline{\mathbf{R}}^{T}_{\boldsymbol{\lambda}} \tag{15}$$

where $\underline{\mathbf{R}}_{d\mathbf{N}}, \underline{\mathbf{R}}_{\boldsymbol{\lambda}}, \underline{\mathbf{R}}^{T}_{\boldsymbol{\lambda}}, \underline{\mathbf{R}}^{T}_{\boldsymbol{\lambda} d\mathbf{N}}$ and $\underline{\mathbf{G}}$ are block column vectors that represent the sampled versions of the correlations $\mathbf{R}_{d\mathbf{N}}(\tau), \mathbf{R}_{\boldsymbol{\lambda}}(\tau), \mathbf{R}_{\boldsymbol{\lambda} d\mathbf{N}}(\tau)$, and the feedback kernel $\mathbf{G}(\tau)$. Block matrices $\underline{\underline{\mathbf{G}}}_{1}$ and $\underline{\underline{\mathbf{G}}}_{2}$ are built from $\mathbf{G}(\tau)$, and $\underline{\underline{\mathbf{I}}}$ is the unity matrix of appropriate dimensions (see also Solution of the Integral Equations of Appendix). The generalized Hawkes model has three different sets of parameters – the input correlation structure $\mathbf{R}_{\boldsymbol{\lambda}}(\tau)$, the output correlation structure $\mathbf{R}_{d\mathbf{N}}(\tau)$, and the Hawkes feedback kernel $\mathbf{G}(\tau)$. Thus, in addition to the forward problem solution presented in Eq. 15, there are three other possible basic scenarios for the identification of the different parts of the proposed generalized Hawkes model from the correlation structure of the observed spike train(s).

(I) $\mathbf{R}_{d\mathbf{N}}(\tau), \mathbf{G}(\tau) \Rightarrow \hat{\mathbf{R}}_{\boldsymbol{\lambda}}(\tau)$

(II) $\mathbf{R}_{d\mathbf{N}}(\tau), \mathbf{R}_{\boldsymbol{\lambda}}(\tau) \Rightarrow \hat{\mathbf{G}}(\tau)$

(III) $\mathbf{R}_{d\mathbf{N}}(\tau) \Rightarrow \hat{\mathbf{G}}(\tau), \hat{\mathbf{R}}_{\boldsymbol{\lambda}}(\tau)$

In the first scenario we are interested in the estimation of the input correlation structure, given the output correlation structure $\mathbf{R}_{d\mathbf{N}}(\tau)$ and the Hawkes kernel $\mathbf{G}(\tau)$. By using the aforementioned matrix notation the solution can be achieved in a straightforward manner, akin to the forward problem:

$$\underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}} = \left(\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_2\right) \cdot \underline{\mathbf{R}}_{d\mathbf{N}} - \underline{\mathbf{G}} \cdot diag\left(\langle d\mathbf{N} \rangle\right)$$
$$\underline{\mathbf{R}}^T_{\boldsymbol{\lambda}} = \left(\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_1\right) \cdot \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} \tag{16}$$

After $\mathbf{R}_{\boldsymbol{\lambda}}(\tau)$is estimated one can proceed, if interested, with the estimation of an LN cascade model for this correlation structure by applying the correlation pre-distortion procedures developed and detailed in (Krumin and Shoham, 2009) and (Krumin and Shoham, 2010). Estimation of the Linear–Non-linear cascade model, in addition to the connectivity kernels $\mathbf{G}(\tau)$, can provide additional insights about the stimulus-driven neural activity.

The second possible scenario is to estimate the Hawkes kernels when the output and the input correlation structures are known (see, e.g., **Figure 3B**). Here, once again, we can use the advantage of the same matrix notation (block column vector $\underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}}$ and block matrix $\underline{\underline{\mathbf{R}}}_{d\mathbf{N}}$ represent the $\mathbf{R}_{\boldsymbol{\lambda} d\mathbf{N}}(\tau)$ and $\mathbf{R}_{d\mathbf{N}}(\tau)$ correlation functions, respectively) and solve the following equations in an iterative manner to estimate $\mathbf{G}(\tau)$:

$$\underline{\underline{\mathbf{G}}}^T = \left(\underline{\underline{\langle d\mathbf{N} \rangle}} + \underline{\underline{\mathbf{R}}}^T_{d\mathbf{N}}\right)^{-1} \left(\underline{\underline{\mathbf{R}}}^T_{d\mathbf{N}} - \underline{\underline{\mathbf{R}}}^T_{\boldsymbol{\lambda} d\mathbf{N}}\right)$$
$$\underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} = \left(\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_1\right)^{-1} \cdot \underline{\mathbf{R}}^T_{\boldsymbol{\lambda}} \tag{17}$$

where $\langle d\mathbf{N} \rangle$ stands for the block diagonal matrix with $diag\left(\langle d\mathbf{N} \rangle\right)$ as its $\overline{\text{block}}$ elements on the main diagonal.

The iterative solution of this set of equations is explained in detail in Section "Solution of the Integral Equations" of Appendix, Eq. A23.

The third possible scenario is to estimate both the kernels $\mathbf{G}(\tau)$ and the input correlation structure $\mathbf{R}_{\boldsymbol{\lambda}}(\tau)$, given only the output correlation structure $\mathbf{R}_{d\mathbf{N}}(\tau)$. In general, this problem is not well-posed and does not have a unique solution, and additional application-driven constraints on the structure of $\mathbf{G}(\tau)$ and/or $\mathbf{R}_{\boldsymbol{\lambda}}(\tau)$ should be considered. We will leave additional discussion on the uniqueness of the solution to the results (see Application to Neural Spike Trains – Single Cells) and in Sections "Discussion."

Refractoriness and strong inhibitory connections

In general, the connectivity between the different units ($\mathbf{G}(\tau)$ feedback terms in the Hawkes model) is not limited to non-negative values. Hence, the firing intensity $\mu(t)$ defined in Eqs 1 or 8 can occasionally become negative. However, the analytical derivations for the output mean rate and correlation structure are based on the assumption that $\mu(t)$ is non-negative for all t. The violation of this assumption results in a discrepancy between the actual and the analytical results. Simulation of the estimated LNH model [while using the effective firing intensity $\mu_{\text{eff}}(t) = \max\{\mu(t), 0\}$] yields output spike trains with a correlation structure $R^{\text{sim}}_{dN}(\tau)$ that is different from the desired output correlation structure $R_{dN}(\tau)$ (used for the estimation of the model parameters). To address this issue an additional procedure was developed for the estimation of the actual feedback kernel $g(\tau)$ from the input and the output correlations ($R_{\lambda}(\tau)$ and $R_{dN}(\lambda)$, respectively). The procedure is summarized in the following algorithm:

1. Estimate initial $g(\tau)$ from $R_{\lambda}(\tau)$ and $R_{dN}(\lambda)$ by solving Eq. 13 (in its matrix form of Eq. 17).
2. Simulate a Hawkes point process using the original input correlation $R_{\lambda}(\tau)$ and the estimated kernel $g(\tau)$. Use $\mu_{\text{eff}}(t) = \max\{\mu(t), 0\}$.
3. Estimate the output correlation $R^{\text{sim}}_{dN}(\tau)$ of the simulated spike train. The violation of the $\mu(t) \geq 0$ assumption will result in a difference between the desired ($R_{dN}(\tau)$) and the estimated ($R^{\text{sim}}_{dN}(\tau)$) output correlation structures.
4. Use the estimated $R^{\text{sim}}_{dN}(\tau)$ instead of the input correlation $R_{\lambda}(\tau)$ in the Eq. 13 to estimate the kernel $\Delta g(\tau)$. The output correlation that should be used is the desired $R_{dN}(\tau)$ throughout the iterative solution, only the input correlation $R_{\lambda}(\tau)$ changes from iteration to iteration.
5. Update $g(\tau) \leftarrow g(\tau) + \alpha \cdot \Delta g(\tau)$. The scalar $\alpha \leq 1$ is used for controlling the speed and/or smoothness of the convergence. In Section "Application to Neural Spike Trains – Single Cells" we have used a relatively small $\alpha = 0.1$ to ensure smooth convergence to the solution.
6. Loop through steps 2–5 until the actual $R^{\text{sim}}_{dN}(\tau)$ of the simulated spike train converges to the desired $R_{dN}(\tau)$.

The above procedure uses the difference between the model-based (simulated) and the desired (data-estimated) correlation structures of the output spike trains to systematically update the feedback kernel $g(\tau)$ until the difference between these two correlation structures becomes small enough. The resulting model allows to relax the assumption of $\mu(t) \geq 0$ and to use $\mu_{\text{eff}}(t) = \max\{\mu(t), 0\}$ instead.

EXPERIMENTAL METHODS

Retina preparation

Animal experiments and procedures were approved by the Institutional Animal Care Committee at the Technion – Israel Institute of Technology and were in accordance with the NIH Guide for the Care and Use of Laboratory Animals. Six-week-old wild type mice (C57/BL) were euthanized using CO_2 and then decapitated. Eyes were enucleated and immersed in Ringer's solution containing (in mM): NaCl, 124; KCl, 2.5; $CaCl_2$, 2; $MgCl_2$, 2; $NaHCO_3$, 26; NaH_2PO_4, 1.25; and Glucose, 22 (pH 7.35–7.4 with 95% O_2 and 5% CO_2 at RT). An incision was made at the ora serrata using a scalpel and the anterior chamber of the eye was separated from the posterior chamber cutting along the ora serrata with fine scissors. The lens was removed and the retina was gently cleaned of the remaining vitreous. Retinal tissue was isolated from the retinal pigmented epithelium. Three radial cuts were made and the isolated retina was flattened with the retinal ganglion cells facing the multi electrode array (MEA). During the experiment the retina was continuously perfused with oxygenated Ringer's solution.

Electrophysiology

The retina was stimulated by wide-field intensity-modulated light flashes using a DLP-based projector. The stimulus intensities were normally distributed and updated at the rate of 60 Hz. Resulting activity was recorded using 60-channel MEA with 10 μm diameter,

planar electrodes spaced at 100 μm. The data was acquired with custom written data acquisition software using Matlab 7.5.0 data acquisition toolbox.

RESULTS

SIMULATION STUDIES

We performed a number of simulation studies to validate the methods proposed for the solution of the integral Eqs 13 and 14.

In **Figures 2A–C** the forward model solution by the Eq. 15 is compared to the auto-correlation function estimated from single simulated point processes with different self-excitation kernels, $g(\tau)$, under two different conditions – constant input λ (pure Hawkes model), or time-varying input $\lambda(t)$ with an exponentially shaped auto-correlation function (LNH model). In **Figure 2D** an example of a bivariate case is presented with a more complex correlation structure of the input $\boldsymbol{\lambda}(t)$ and a set of self- and mutually exciting kernels $\mathbf{G}(\tau)$.

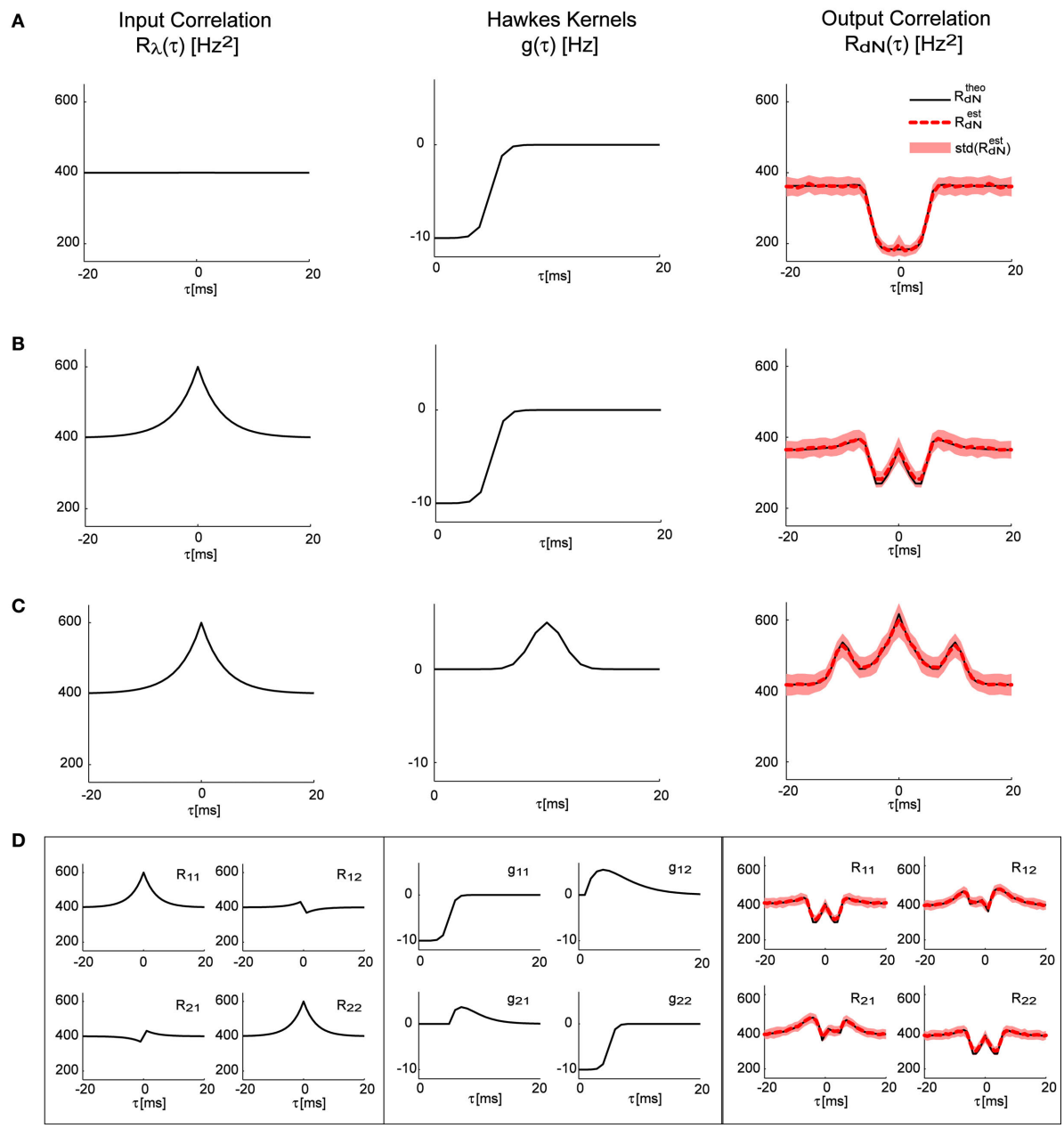


FIGURE 2 | Correlation structure of the homogeneous and inhomogeneous Hawkes models can be accurately predicted. Predicted theoretical correlation structure is compared to the correlation structure estimated from simulated point processes in several cases: **(A)** Constant λ and a refractory period-like self-exciting kernel $g(\tau)$. **(B)** Same as in **(A)**, but with time-varying $\lambda(t)$ that has an exponentially shaped auto-correlation function. **(C)** Similar to **(B)**, but with a different self-excitation kernel $g(\tau)$. **(D)** Bivariate mutually exciting point processes driven by time-varying exogenous inputs with complex correlation structure. Mean values and standard deviations of the estimators were calculated from 100 simulations (each 10 min long) of corresponding Hawkes models.

As can be seen in all of these examples, the analytically predicted correlation functions had a near-perfect match with the mean correlation functions of the simulated spike trains (correlation coefficient ≥0.99). Individual correlation functions calculated from 10-min traces were more noisy, thus the forward analytical prediction vs. simulation correlation coefficients for single traces were significantly lower: 0.83 ± 0.06.

Figure 3A shows the result of applying the "scenario I" solution (Eq. 16) to spike trains generated by the model presented in **Figure 2D**; the mean identified input correlations have an excellent match with the ones used for generating the data (correlation coefficients: 0.99 and 0.92 respectively for the auto- and cross-correlations).

Figure 3B shows the result of applying the "scenario II" solution (Eq. 17) to spike trains generated by the model presented in **Figure 2D**; the mean identified kernels greatly match the ones used in generating the data (correlation coefficients >0.99 for all kernels.

APPLICATION TO NEURAL SPIKE TRAINS – SINGLE CELLS

Next, we applied the method on the data recorded from the retina (see Methods for the experimental protocol). We started by analyzing the spike trains using reverse-correlation techniques (Ringach and Shapley, 2004) based on a feed-forward Linear–Non-linear–Poisson (LNP) model. The LNP-based estimates of the linear filter, and the static non-linearity (**Figure 4A**) were further used for the calculation of the expected output auto-correlation function of the estimated LNP model. This LNP-based output auto-correlation function was found to be noticeably different from the actual auto-correlation function of the measured spike trains (**Figure 4B**).

This LN cascade was then used for generating the input ($\lambda(t)$ in **Figure 1**) to the Hawkes feedback stage of the LNH model. The auto-correlation function of $\lambda(t)$ is exactly that of the LNP model's output estimated previously and found inconsistent with the real recordings. Now, the input auto-correlation function $R_{\lambda}(\tau)$ was used together with the measured output auto-correlation function $R_{dN}(\tau)$ to estimate the Hawkes feedback kernel $g(\tau)$ (**Figure 4C**) from Eq. 13 (including the procedure described in the Refractoriness and Strong Inhibitory Connections). Interestingly, the output auto-correlation function of the newly estimated LNH model (as measured from the simulated spike trains) was in excellent agreement with the auto-correlation function of the actual neural data (**Figure 4D**). The addition of the linear Hawkes feedback stage to the classical feed-forward LNP model proved beneficial to the model's capability of explaining more complex spike train correlation structures of real neural recordings (**Figure 4E**).

Finally, we validated that the improved fit of the LNH model to the data compared with the LNP model, does not result from a model overfitting due to the larger number of parameters in the LNH model. For each unit, we computed an LN-Hawkes for a different data set from the same unit (Gaussian distribution, different mean intensity). Next, we simulated an output spike train using a "hybrid" LNH model ("original" LN model + "new" feedback kernel $g(\tau)$), and estimated its correlation function. This output correlation function was compared to the correlation function of the original data by calculating the correlation coefficient between the two functions ρ_{LNH}. This procedure was applied to the nine units in our data set where the mean firing rates were >2 Hz. In eight out of these nine units the hybrid LNH model provided considerably better fits to the output

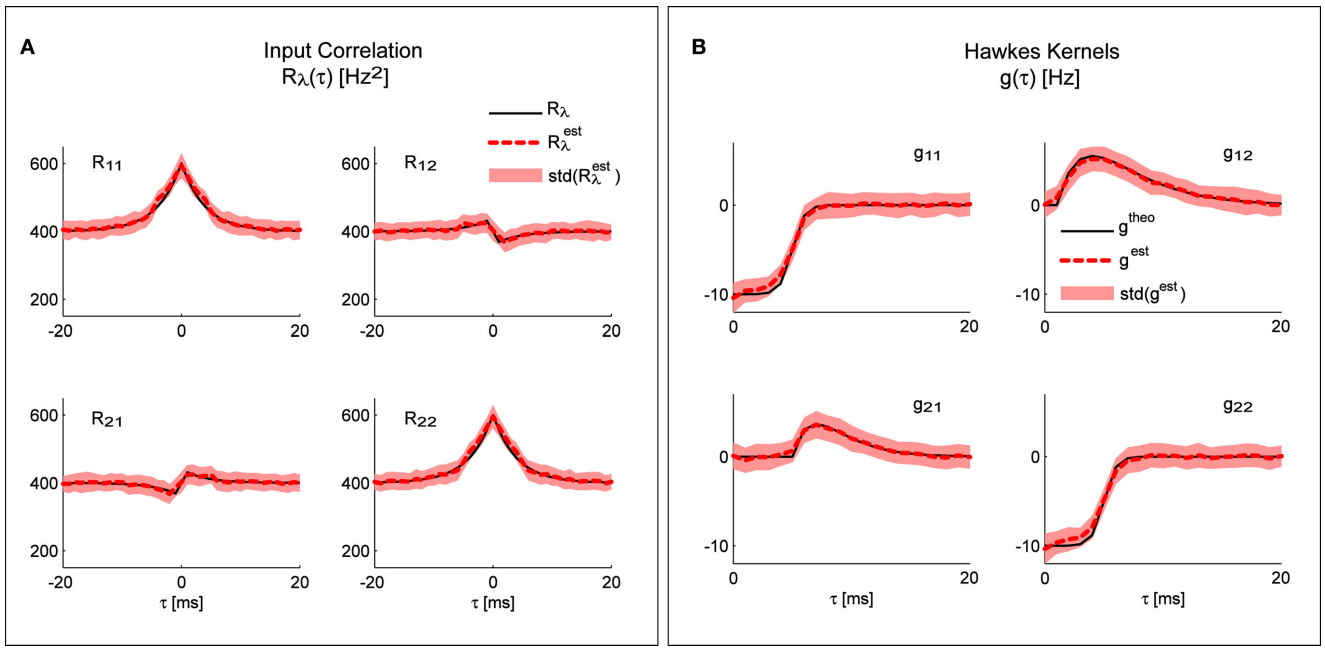


FIGURE 3 | System identification. Any of the three different parts of the system can be identified from the other two. **(A)** Comparison of the input correlation structure estimated from the simulated point processes and the real values used in the simulation. **(B)** Hawkes kernels estimated from the simulated point processes and input correlation structure are compared to their real value used for the simulation. Mean values and standard deviations of the estimators were calculated from 100 simulations (each 10 min long) of the bivariate inhomogeneous Hawkes models from **Figure 2D**.

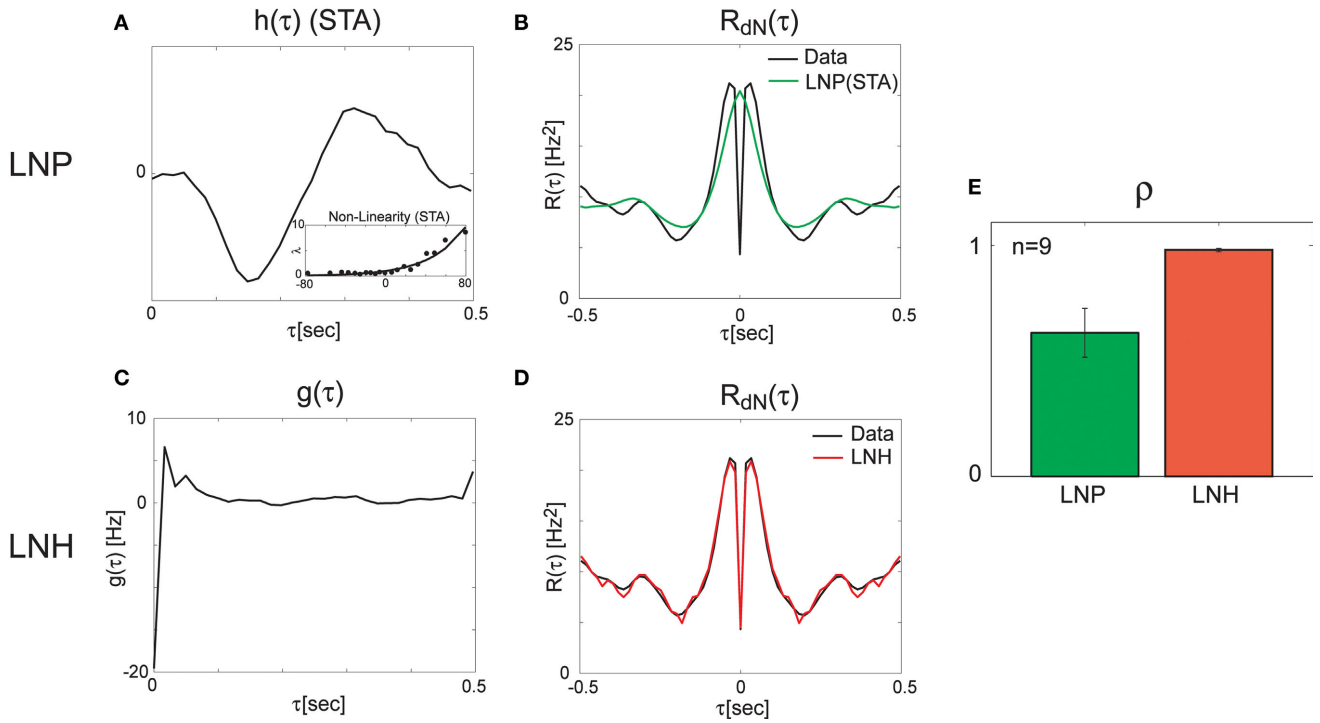


FIGURE 4 | Linear–non-linear-Hawkes and LNP model fits to single-unit retinal neural spike train auto-correlations. Single-unit recordings from mouse retinal ganglion cells were analyzed using the LNP and the LNH model-based approaches with the LNH model succeeding to explain the spike trains' correlations much better than the LNP model. **(A)** Linear filter $h(\tau)$ and the non-linearity estimated using reverse-correlation approach (spike triggered average). **(B)** The expected output auto-correlation function of the LNP model calculated from the parameters in **(A)** does not fit the actual auto-correlation function of the spike train well. **(C)** The self-excitation kernel $g(\tau)$ of the LNH model shows strong refractoriness that cannot be explained by the LNP model. **(D)** The LNH model output auto-correlation precisely fits the actual spike train auto-correlation measured from the data. **(E)** The correlation coefficients between the model and the actual output auto-correlation functions are significantly ($p = 0.005$) higher for the LNH model (with mean ± SE of $\rho_{LNP} = 0.62 \pm 0.11$ and $\rho_{LNH} = 0.98 \pm 0.01$, $n = 9$).

correlation function than the corresponding LNP model, providing in those cases an average improvement of $\langle \Delta\rho \rangle = \langle \rho_{LNH} - \rho_{LNP} \rangle = 0.19$ with $\langle \Delta\rho \rangle / \langle \rho_{LNP} \rangle = 30\%$. Note that this procedure is over-conservative, since there is no guarantee that kernels calculated for different input stimulus ensembles will be the same or conversely, that neural models will generalize across different stimulus ensembles.

DISCUSSION

In this paper, we extended previous work on the correlation-based simulation, identification and analysis of multi-channel spike train models with a feed-forward Linear–Non-linear (LN) stage driven by Gaussian process inputs (Krumin and Shoham, 2009; Krumin et al., 2010), by allowing the non-negative process to drive a feedback stage in the form of a multi-channel Hawkes process. The move from doubly stochastic Poisson (Cox) models in our previous work to doubly stochastic Hawkes models employed here vastly expands the range of realizable correlation structures, thus relaxing the main limitation of the previous results, and allowing for a superior, excellent fit ($\rho \simeq 0.98$) of the auto-correlation structures of spike trains recorded from real visually driven retinal ganglion cells. At the same time, it preserves the analytical tractability and closed-form correspondence between model parameters and the second-order statistical properties of the output spike trains, and thus, essentially, all of the advantages and potential applications of the general model-based correlation framework, which was limited, thus far, to feed-forward models. These currently include the synthetic generation of spike trains with a pre-defined correlation structure (Brette, 2009; Krumin and Shoham, 2009; Macke et al., 2009; Gutnisky and Josic, 2010; Tchumatchenko et al., 2010), "blind" correlation-based identification of single-neuron encoding models (Krumin et al., 2010), the compact representation of multi-channel spike trains in terms of multivariate auto-regressive processes and the framework of causality (Granger) analysis (Nykamp, 2007; Krumin and Shoham, 2010). As noted above, the LNH model is related to the commonly used GLM model, with the LNH feedback kernels paralleling the GLM history terms. Both ways of altering the underlying feed-forward LNP model lead to more flexible models capable of fitting more complex correlation structures, but the preferred fitting procedures for the two models differ: the GLM model is typically fit using a maximum likelihood approach, but this does not suit the LNH model (due to possible zero firing rates), where a method of moments (like the one introduced here) is more appropriate for the estimation of the linear kernels. A systematic study on the differences between the statistical properties of the two approaches falls beyond the scope of the current manuscript.

The model and analysis presented here also provide a new context and results to a significant body of related previous work on the second-order statistics of Hawkes models, which we will now review very briefly. The basic properties of the output correlation

structure and the spectrum of a univariate self-exciting and a multivariate mutually exciting linear point process model without an exogenous drive were derived in the original works of Hawkes (1971a,b) using the linear representation of this process (Eq. 2). Brillinger (1975) also analyzes linear point process models and uses spectral estimators for the kernels, which he applies to the analysis of synaptic connections (Brillinger et al., 1976). Bremaud and Massoulie (2002) and Daley and Vere-Jones (2003) (exercise 8.3.4) present expressions for the output spectrum of a univariate Hawkes model excited by an exogenous correlated point process derived using an alternative, cluster process representation of the Hawkes process:

$$d\tilde{N}(\omega) = \frac{\Gamma \cdot \mathbb{E}\{\lambda(t)\}/(1-\Gamma) + \tilde{\lambda}(\omega)}{\left|1-\tilde{g}(\omega)\right|^2},$$

where $\Gamma \triangleq \int_0^\infty g(u)du$ and $d\tilde{N}(\omega)$, $\tilde{\lambda}(\omega)$, $\tilde{g}(\omega)$ represent the respective spectra of $dN(t), \lambda(t), g(t)$. Our derivation in the Section "Methods" and "Correlation Structure of the LNH Model" of Appendix focused on expressions for the correlation structure of exogenously driven Hawkes process and was based on the linear representation, similar to Hawkes (1971a). Adding the exogenous input introduces a new term into the Hawkes integral Eq. 10, and a second integral equation for the cross-covariance term between the exogenous input and the output spike trains $\mathbf{R}_{\boldsymbol{\lambda} dN}(\tau)$. The parameters of these generalized models, i.e., the kernels $\mathbf{G}(\tau)$ and/or the input correlation structure $\mathbf{R}_{\boldsymbol{\lambda}}(\tau)$, can be directly estimated from the output process correlation structure using an iterative application of this set of equations, as illustrated in Section "Results," or they could, alternatively, be estimated from the spectral expressions.

We next turn to discuss certain limitations of the proposed framework. First, the analytical equations for the auto-correlation structure of the point processes (Eqs 7, 10, 13, and 14) are *exactly* true under the assumption $\mu(t) \geq 0$ (Eqs 2, 8, and 11) or when the stochastic intensity is always non-negative. These exact results could also provide an excellent agreement to many practical cases wherein the self-exciting Hawkes kernel $g(\tau)$ is only weakly negative (e.g., **Figure 2**), leading in such cases to slight systematic deviations at "negative" peaks. In cases of strong refractoriness or other inhibitory interactions, $g(\tau)$ becomes strongly negative, and the rectification of the stochastic intensity around zero leads to strong deviations from the assumptions underlying Eqs 7 and 13. For such cases we introduced an intuitive iterative procedure for computing $g(\tau)$ (see Refractoriness and Strong Inhibitory Connections), and it is likely that related alternatives are also possible. Although the convergence of this procedure is not proven, in practice, it was capable of estimating kernels for real neural spike trains that not only dramatically improved the auto-correlation fits relative to LNP cascades, but also generalized across different stimulus ensembles (a very conservative cross-validation test). Second, we have not addressed the important but complex issue of uniqueness of the different identification problems encountered here. Interestingly, in the examples we have examined, an excellent match was found, in practice, between the Hawkes kernels and their estimates (**Figure 3B**), although we are not aware of any guarantees of uniqueness here (these may perhaps be related to the nature of point processes). In the more general problem where both $\hat{\mathbf{G}}(\tau)$, $\hat{\mathbf{R}}_{\boldsymbol{\lambda}}(\tau)$ are simultaneously estimated, it seems obvious that unique solutions can only be obtained by imposing additional constraints on the solutions (i.e., degree of smoothness and/or sparseness). In section "Application to Neural Spike Trains – Single Cells" we presented an example of the "scenario III"-type problem, where only the output correlation structure is actually observable. In this example we used additional application-driven constraints on the input correlation structure $\mathbf{R}_{\boldsymbol{\lambda}}(\tau)$ to infer the feedback kernels $\mathbf{G}(\tau)$. Interestingly, the exact same "scenario III"-type framework can be used for generating synthetic spike trains with a controlled correlation structure. This application will benefit from using the LNH feedback model by harnessing the capability of generating spike trains with a much richer ensemble of possible correlation structures in comparison with the feed-forward-only models like LNP. Additionally, once $\hat{\mathbf{R}}_{\boldsymbol{\lambda}}(\tau)$ is determined there is an additional level of non-uniqueness in the determination of the underlying LN structure, which can also be overcome by imposing constraints (e.g., a minimum phase constraint (Krumin et al., 2010)).

When considering the broader relevance of this work, and the directions to which it may develop in the future, it is worth noting that some of the most fundamental and widely applied tools for the identification of systems rely on the use of second-order statistical properties (Ljung, 1999) (correlation or spectral). The increasing arsenal of tools for identifying spike train models from their correlations, rather than from their full observed realizations could form a welcome bridge between "classical" signal processing ideas and tools and the field of neural spike train analysis.

ACKNOWLEDGMENTS

This work was supported by Israeli Science Foundation grant #1248/06 and European Research Council starting grant #211055. We thank A. Tankus and the two anonymous reviewers for their comments on the manuscript.

REFERENCES

Bremaud, P., and Massoulie, L. (2002). Power spectra of general shot noises and Hawkes point processes with a random excitation. *Adv. Appl. Probab.* 34, 205–222.

Brette, R. (2009). Generation of correlated spike trains. *Neural. Comput.* 21, 188–215.

Brillinger, D. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cybern.* 59, 189–200.

Brillinger, D. R. (1975). The identification of point process systems. *Ann. Probab.* 3, 909–924.

Brillinger, D. R., Bryant, H. L., and Segundo, J. P. (1976). Identification of synaptic interactions. *Biol. Cybern.* 22, 213–228.

Brown, E. N., Kass, R. E., and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 7, 456–461.

Cardanobile, S., and Rotter, S. (2010). Multiplicatively interacting point processes and applications to neural modeling. *J. Comput. Neurosci.* 28, 267–284.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network* 12, 199–213.

Chornoboy, E., Schramm, L., and Karr, A. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybern.* 59, 265–275.

Daley, D. J., and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes,* Vol. 1. New York: Springer.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.

Gutnisky, D. A., and Josic, K. (2010). Generation of spatio-temporally correlated spike-trains and local-field potentials using a multivariate autoregressive process. *J. Neurophysiol.* 103, 2912–2930.

Hawkes, A. G. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 83–90.

Hawkes, A. G. (1971b). Point spectra of some mutually exciting point processes. *J. R. Stat. Soc. Series B Methodol.* 33, 438–443.

Johnson, D. H. (1996). Point process models of single-neuron discharges. *J. Comput. Neurosci.* 3, 275–299.

Kailath, T., Sayed, A. H., and Hassibi, B. (2000). *Linear Estimation*. Upper Saddle River, NJ: Prentice Hall.

Krumin, M., Shimron, A., and Shoham, S. (2010). Correlation-distortion based identification of Linear-Nonlinear-Poisson models. *J. Comput. Neurosci.* 29, 301–308.

Krumin, M., and Shoham, S. (2009). Generation of spike trains with controlled auto- and cross-correlation functions. *Neural. Comput.* 21, 1642–1664.

Krumin, M., and Shoham, S. (2010). Multivariate auto-regressive modeling and granger causality analysis of multiple spike trains. *Computat. Intell. Neurosci.* 2010, Article ID 752428.

Ljung, L. (1999). *System Identification – Theory for the User,* 2nd Edn. Upper Saddle River, NJ: Prentice Hall PTR.

Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S., and Bethge, M. (2009). Generating spike trains with specified correlation coefficients. *Neural. Comput.* 21, 397–423.

Mayers, D. F. (1962). "Part II. Integral equations, Chapters 11–14," in *Numerical Solution of Ordinary and Partial Differential Equations*, ed. L. Fox (London: Pergamon), 145–183.

Noble, B. (1958). *Methods Based on the Wiener-Hopf Technique*. London: Pergamon.

Nykamp, D. Q. (2007). A mathematical framework for inferring connectivity in probabilistic neuronal networks. *Math. Biosci.* 205, 204–251.

Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog. Brain Res.* 165, 493–507.

Paninski, L., Shoham, S., Fellows, M. R., Hatsopoulos, N. G., and Donoghue, J. P. (2004). Superlinear population encoding of dynamic hand trajectory in primary motor cortex. *J. Neurosci.* 24, 8551–8561.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999.

Ringach, D., and Shapley, R. (2004). Reverse correlation in neurophysiology. *Cogn. Sci.* 28, 147–166.

Shoham, S., Paninski, L. M., Fellows, M. R., Hatsopoulos, N. G., Donoghue, J. P., and Normann, R. A. (2005). Statistical encoding model for a primary motor cortical brain-machine interface. *IEEE Trans. Biomed. Eng.* 52, 1312–1322.

Stevenson, I. H., Rebesco, J. M., Miller, L. E., and Koerding, K. P. (2008). Inferring functional connections between neurons. *Curr. Opin. Neurobiol.* 18, 582–588.

Tchumatchenko, T., Malyshev, A., Geisel, T., Volgushev, M., and Wolf, F. (2010). Correlations and synchrony in threshold neuron models. *Phys. Rev. Lett.* 104, 058102.

Toyoizumi, T., Rad, K. R., and Paninski, L. (2009). Mean-field approximations for coupled populations of generalized linear model spiking neurons with Markov refractoriness. *Neural. Comput.* 21, 1203–1243.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 December 2009; accepted: 25 October 2010; published online: 19 November 2010.

Citation: Krumin M, Reutsky I and Shoham S (2010) Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. Front. Comput. Neurosci. ***4****:147. doi: 10.3389/fncom.2010.00147*

APPENDIX

CORRELATION STRUCTURE OF THE LNH MODEL

Part I – Derivation of the output correlation of the inhomogeneous Hawkes point process

We consider the Hawkes point process driven by a time-varying exogenous input, with the intensity defined in Eq. 11:

$$\mu(t) = \lambda(t) + \int_{-\infty}^{t} g(t-u)dN(u)$$

For the mean firing rate we receive:

$$\langle dN \rangle \triangleq \mathbb{E}\left\{\frac{dN(t)}{dt}\right\} = \mathbb{E}\left\{\lambda(t) + \int_{-\infty}^{t} g(t-u)dN(u)\right\}$$
$$= \mathbb{E}\{\lambda(t)\} + \int_{-\infty}^{t} g(t-u)\mathbb{E}\left\{\frac{dN(t)}{dt}\right\}du = \mathbb{E}\{\lambda(t)\} + \langle dN \rangle \cdot \int_{0}^{\infty} g(u)du \tag{A1}$$

resulting in:

$$\langle dN \rangle = \frac{\mathbb{E}\{\lambda(t)\}}{1 - \int_{0}^{\infty} g(u)du} \tag{A2}$$

Next, we expand the expressions for the correlation structure of the output spike trains, following a similar formalism to the derivation found in Hawkes (1971a) for the correlations of homogeneous Hawkes processes:

$$\begin{aligned} R_{dN}(\tau) &\triangleq \tilde{R}_{dN}(\tau) - \langle dN \rangle \cdot \delta(\tau) \\ &= \mathbb{E}\{dN(t+\tau)dN(t)\}/dt^2 - \langle dN \rangle \cdot \delta(\tau) \\ &= \mathbb{E}\left\{\frac{dN(t)}{dt} \cdot \left[\lambda(t+\tau) + \int_{-\infty}^{t+\tau} g(t+\tau-u)dN(u)\right]\right\} \\ &= \mathbb{E}\left\{\frac{dN(t)}{dt} \cdot \lambda(t+\tau)\right\} + \mathbb{E}\left\{\frac{dN(t)}{dt}\int_{-\infty}^{t+\tau} g(t+\tau-u)dN(u)\right\} \\ &= R_{\lambda dN}(\tau) + \int_{-\infty}^{t+\tau} g(t+\tau-u)\tilde{R}_{dN}(t-u)du \end{aligned} \tag{A3}$$

Now, substituting $\tilde{R}_{dN}(\tau) = R_{dN}(\tau) + \langle dN \rangle \cdot \delta(\tau)$ we get:

$$R_{dN}(\tau) = R_{\lambda dN}(\tau) + g(\tau) \cdot \langle dN \rangle + \int_{-\infty}^{\tau} g(\tau-u)R_{dN}(u)du \tag{A4}$$

We have arrived to a solution similar to Eq. 7 with one additional term $R_{\lambda dN}(\tau)$ that will be derived in Part II.

Part II – Derivation of the cross-correlation between the exogenous input λ (t) and the output point process

The derivation of $R_{\lambda dN}(\tau)$ has much in common with the derivations in Part I above.

$$\begin{aligned} R_{\lambda dN}(\tau) &\triangleq \mathbb{E}\left\{\lambda(t+\tau) \cdot \frac{dN(t)}{dt}\right\} \\ &= \mathbb{E}\left\{\lambda(t+\tau) \cdot \left[\lambda(t) + \int_{-\infty}^{t} g(t-u)dN(u)\right]\right\} \\ &= \mathbb{E}\{\lambda(t+\tau) \cdot \lambda(t)\} + \mathbb{E}\left\{\lambda(t+\tau) \cdot \int_{-\infty}^{t} g(t-u)dN(u)\right\} \\ &= R_{\lambda}(\tau) + \int_{-\infty}^{t} g(t-u)\mathbb{E}\left\{\lambda(t+\tau)\frac{dN(u)}{du}\right\}du \\ &= R_{\lambda}(\tau) + \int_{-\infty}^{t} g(t-u)R_{\lambda dN}(t+\tau-u)du \end{aligned} \tag{A5}$$

To summarize, the derivations in Part I and Part II of the current Appendix result in two coupled integral equations:

$$R_{dN}(\tau) = R_{\lambda dN}(\tau) + g(\tau) \cdot \langle dN \rangle + \int_{-\infty}^{\tau} g(\tau-u)R_{dN}(u)du$$
$$R_{\lambda dN}(\tau) = R_{\lambda}(\tau) + \int_{\tau}^{\infty} g(u-\tau)R_{\lambda dN}(u)du \tag{A6}$$

Part III – Derivation of the output correlation structure for the multidimensional LNH model

Let us now consider a multivariate inhomogeneous Hawkes process:

$$\boldsymbol{\mu}(t) = \boldsymbol{\lambda}(t) + \int_{-\infty}^{t} \mathbf{G}(t-u)d\mathbf{N}(u), \tag{A7}$$

where $\boldsymbol{\mu}(t)$, $\boldsymbol{\lambda}(t)$, and $d\mathbf{N}(t)$are now column vectors, and $\mathbf{G}(\tau)$ is a square matrix. The values in the row #*r* and column #*s* of the matrix $\mathbf{G}(\tau)$ correspond to the mutual-excitation kernel that explains the effect of the firing history of the process #*s* on the stochastic intensity of the process #*r*.

The expression for the mean firing rate $\langle d\mathbf{N} \rangle$ of the process is derived in the following way:

$$\langle d\mathbf{N} \rangle \triangleq \mathbb{E}\left\{\frac{d\,\mathbf{N}(t)}{dt}\right\} = \mathbb{E}\left\{\boldsymbol{\lambda}(t) + \int_{-\infty}^{t} \mathbf{G}(t-u)d\,\mathbf{N}(u)\right\}$$
$$= \mathbb{E}\{\boldsymbol{\lambda}(t)\} + \int_{-\infty}^{t} \mathbf{G}(t-u)\mathbb{E}\left\{\frac{d\mathbf{N}(t)}{dt}\right\}du = \mathbb{E}\{\boldsymbol{\lambda}(t)\} + \langle d\mathbf{N} \rangle \cdot \int_{0}^{\infty} \mathbf{G}(u)du, \tag{A8}$$

resulting in

$$\langle d\mathbf{N} \rangle = \left(\mathbf{I} - \int_{0}^{\infty} \mathbf{G}(u)du\right)^{-1} \cdot \mathbb{E}\{\boldsymbol{\lambda}(t)\} \tag{A9}$$

The output correlation structure is now defined by:

$$\begin{aligned}\mathbf{R}_{d\mathbf{N}}(\tau) &\triangleq \tilde{\mathbf{R}}_{d\mathbf{N}}(\tau) - diag\left(\langle d\mathbf{N}\rangle\right)\cdot\delta(\tau)\\ &= \mathbb{E}\left\{d\mathbf{N}(t+\tau)d\mathbf{N}^T(t)\right\}/dt^2 - diag\left(\langle d\mathbf{N}\rangle\right)\cdot\delta(\tau)\\ &= \mathbb{E}\left\{\left[\boldsymbol{\lambda}(t+\tau)+\int_{-\infty}^{t+\tau}\mathbf{G}(t+\tau-u)d\mathbf{N}(u)\right]\cdot\frac{d\mathbf{N}^T(t)}{dt}\right\}\\ &= \mathbb{E}\left\{\boldsymbol{\lambda}(t+\tau)\cdot\frac{d\mathbf{N}^T(t)}{dt}\right\}+\mathbb{E}\left\{\int_{-\infty}^{t+\tau}\mathbf{G}(t+\tau-u)d\mathbf{N}(u)\cdot\frac{d\mathbf{N}^T(t)}{dt}\right\}\\ &= \mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(\tau)+\int_{-\infty}^{t+\tau}\mathbf{G}(t+\tau-u)\tilde{\mathbf{R}}_{d\mathbf{N}}(t-u)du\\ &= \mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(\tau)+\mathbf{G}(\tau)\cdot diag\left(\langle d\mathbf{N}\rangle\right)+\int_{-\infty}^{t+\tau}\mathbf{G}(t+\tau-u)\mathbf{R}_{d\mathbf{N}}(t-u)du\\ &= \mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(\tau)+\mathbf{G}(\tau)\cdot diag\left(\langle d\mathbf{N}\rangle\right)+\int_{-\infty}^{\tau}\mathbf{G}(\tau-u)\mathbf{R}_{d\mathbf{N}}(u)du\end{aligned} \tag{A10}$$

Similarly to the Eq. A5 we can also derive:

$$\begin{aligned}\mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(\tau) &\triangleq \mathbb{E}\left\{\boldsymbol{\lambda}(t+\tau)\cdot\frac{d\mathbf{N}^T(t)}{dt}\right\}\\ &= \mathbb{E}\left\{\boldsymbol{\lambda}(t+\tau)\cdot\left[\boldsymbol{\lambda}(t)+\int_{-\infty}^{t}\mathbf{G}(t-u)d\mathbf{N}(u)\right]^T\right\}\\ &= \mathbb{E}\left\{\boldsymbol{\lambda}(t+\tau)\cdot\boldsymbol{\lambda}^T(t)\right\}+\mathbb{E}\left\{\boldsymbol{\lambda}(t+\tau)\cdot\int_{-\infty}^{t}d\mathbf{N}^T(u)\mathbf{G}^T(t-u)\right\}\\ &= \mathbf{R}_{\boldsymbol{\lambda}}(t)+\int_{-\infty}^{t}\mathbb{E}\left\{\boldsymbol{\lambda}(t+\tau)\frac{d\mathbf{N}^T(u)}{du}\right\}\mathbf{G}^T(t-u)du\\ &= \mathbf{R}_{\boldsymbol{\lambda}}(t)+\int_{-\infty}^{t}\mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(t+\tau-u)\mathbf{G}^T(t-u)du\\ &= \mathbf{R}_{\boldsymbol{\lambda}}(\tau)+\int_{\tau}^{\infty}\mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(u)\mathbf{G}^T(u-\tau)du\end{aligned} \tag{A11}$$

SOLUTION OF THE INTEGRAL EQUATIONS

Part I – Developing the discrete time matrix notation formalism for the integral equations

The following coupled equations govern the relationship between the input correlation structure $\mathbf{R}_{\boldsymbol{\lambda}}(\tau)$, the output correlation structure $\mathbf{R}_{d\mathbf{N}}(\tau)$, and the feedback linear kernel $\mathbf{G}(\tau)$ of the generalized Hawkes model:

$$\begin{aligned}\mathbf{R}_{d\mathbf{N}}(\tau) &= \mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(\tau)+\mathbf{G}(\tau)\cdot diag\left(\langle d\mathbf{N}\rangle\right)+\int_{-\infty}^{\tau}\mathbf{G}(\tau-u)\mathbf{R}_{d\mathbf{N}}(u)du\\ \mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(\tau) &= \mathbf{R}_{\boldsymbol{\lambda}}(\tau)+\int_{\tau}^{\infty}\mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(u)\mathbf{G}^T(u-\tau)du\end{aligned} \tag{A12}$$

We can rewrite these equations in the following manner:

$$\begin{aligned}\mathbf{R}_{d\mathbf{N}}(\tau) &= \mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(\tau)+\mathbf{G}(\tau)\cdot diag\left(\langle d\mathbf{N}\rangle\right)+\int_{-\infty}^{\tau}\mathbf{G}(\tau-u)\mathbf{R}_{d\mathbf{N}}(u)du\\ &= \mathbf{R}_{\boldsymbol{\lambda}d\mathbf{N}}(\tau)+\mathbf{G}(\tau)\cdot diag\left(\langle d\mathbf{N}\rangle\right)\\ &\quad+\int_{0}^{\tau}\mathbf{G}(\tau-u)\mathbf{R}_{d\mathbf{N}}(u)du+\int_{0}^{\infty}\mathbf{G}(\tau+u)\mathbf{R}_{d\mathbf{N}}(u)du\\ \mathbf{R}^T_{\boldsymbol{\lambda}d\mathbf{N}}(\tau) &= \mathbf{R}^T_{\boldsymbol{\lambda}}(\tau)+\int_{\tau}^{\infty}\mathbf{G}(u-\tau)\mathbf{R}^T_{\boldsymbol{\lambda}d\mathbf{N}}(u)du\end{aligned} \tag{A13}$$

To solve these equations numerically we use the following discretized representation:

$$\begin{aligned}\underline{\mathbf{R}}_{d\mathbf{N}} &= \underline{\mathbf{R}}_{\boldsymbol{\lambda}d\mathbf{N}}+\underline{\mathbf{G}}\cdot diag\left(\langle d\mathbf{N}\rangle\right)+\underline{\underline{\mathbf{G}}}_T\cdot\underline{\mathbf{R}}_{d\mathbf{N}}+\underline{\underline{\mathbf{G}}}_H\cdot\underline{\mathbf{R}}_{d\mathbf{N}}\\ &= \underline{\mathbf{R}}_{\boldsymbol{\lambda}d\mathbf{N}}+\underline{\mathbf{G}}\cdot diag\left(\langle d\mathbf{N}\rangle\right)+\underline{\underline{\mathbf{G}}}_2\cdot\underline{\mathbf{R}}_{d\mathbf{N}}\\ \underline{\mathbf{R}}^T_{\boldsymbol{\lambda}d\mathbf{N}} &= \underline{\mathbf{R}}^T_{\boldsymbol{\lambda}}+\underline{\underline{\mathbf{G}}}_1\cdot\underline{\mathbf{R}}^T_{\boldsymbol{\lambda}d\mathbf{N}},\end{aligned} \tag{A14}$$

where $\langle d\mathbf{N}\rangle$ – is a block column vector representing the mean firing rates of the output spike trains

$\underline{\mathbf{G}},\underline{\mathbf{R}}_{d\mathbf{N}},\underline{\mathbf{R}}_{\boldsymbol{\lambda}},\underline{\mathbf{R}}_{\boldsymbol{\lambda}d\mathbf{N}}$ – block column vectors of N block elements with the first block element representing $\tau=0$, and the last block element representing $\tau=\tau_{max}$. The choice of the discretization time-step $d\tau$ depends on the desired time resolution of the solution.

$\underline{\mathbf{R}}^T_{\boldsymbol{\lambda}},\underline{\mathbf{R}}^T_{\boldsymbol{\lambda}d\mathbf{N}}$ – also block column vectors, but with their block elements transposed (in the univariate case $\underline{\mathbf{R}}^T_{\boldsymbol{\lambda},\boldsymbol{\lambda}d\mathbf{N}}=\underline{\mathbf{R}}_{\boldsymbol{\lambda},\boldsymbol{\lambda}d\mathbf{N}}$)

$\underline{\underline{\mathbf{G}}}_1,\underline{\underline{\mathbf{G}}}_T,\underline{\underline{\mathbf{G}}}_H,\underline{\underline{\mathbf{G}}}_2$ – square block matrices of size $N\times N$ blocks that match the dimensions of the block column vectors.

To convert the integration operations into matrix multiplication operations we define the matrices $\underline{\underline{\mathbf{G}}}_1$ and $\underline{\underline{\mathbf{G}}}_2=\underline{\underline{\mathbf{G}}}_T+\underline{\underline{\mathbf{G}}}_H$ ($d\tau$ – time resolution) in the following way:

$$\underline{\underline{\mathbf{G}}}_1 = d\tau\cdot\begin{bmatrix}\mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \cdots & \mathbf{G}_{N-1}\\ 0 & \mathbf{G}_0 & \mathbf{G}_1 & & \vdots\\ \vdots & \ddots & \ddots & \ddots & \\ & & & & \mathbf{G}_1\\ 0 & \cdots & & 0 & \mathbf{G}_0\end{bmatrix} \tag{A15}$$

is a block Toeplitz matrix with the elements of the block vector $\underline{\mathbf{G}}$ in the first row, and zeros in the first block column (excluding the main diagonal). The block elements of the matrix are;

$$\mathbf{G}_k \triangleq \mathbf{G}\left(\tau=k\cdot d\tau\right) \tag{A16}$$

$\underline{\underline{\mathbf{G}}}_2$ is a sum of two other matrices: $\underline{\underline{\mathbf{G}}}_2=\underline{\underline{\mathbf{G}}}_T+\underline{\underline{\mathbf{G}}}_H$, where

$$\underline{\underline{\mathbf{G}}}_T = d\tau\cdot\begin{bmatrix}\mathbf{G}_0 & 0 & \cdots & & 0\\ \mathbf{G}_1 & \mathbf{G}_0 & & & \vdots\\ \mathbf{G}_2 & \ddots & \ddots & & \\ \vdots & \ddots & & & 0\\ \mathbf{G}_{N-1} & \cdots & & \mathbf{G}_1 & \mathbf{G}_0\end{bmatrix} \tag{A17}$$

is a block Toeplitz matrix with the elements of the block vector $\underline{\mathbf{G}}$ in the first block column, and zeros in the first block row (excluding the main diagonal).

$$\underline{\underline{\mathbf{G}}}_H = d\tau \cdot \begin{bmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \cdots & \mathbf{G}_{N-1} \\ \mathbf{G}_1 & \mathbf{G}_2 & & \iddots & 0 \\ \mathbf{G}_2 & & \iddots & & \vdots \\ \vdots & \iddots & & & \\ \mathbf{G}_{N-1} & 0 & \cdots & & 0 \end{bmatrix} \tag{A18}$$

is a block Henkel matrix with the elements of the block vector $\underline{\mathbf{G}}$ in the first block column, and zeros in the last block row (excluding the secondary diagonal).

Part II – Solution of the equations for different scenarios

The solution of the Eq. A14 for the output correlation structure $\underline{\mathbf{R}}_{d\mathbf{N}}$ (the forward model) is straightforward:

$$\left.\begin{aligned} \underline{\mathbf{R}}_{d\mathbf{N}} &= \underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}} + \underline{\mathbf{G}} \cdot diag\left(\langle d\mathbf{N}\rangle\right) + \underline{\underline{\mathbf{G}}}_2 \cdot \underline{\mathbf{R}}_{d\mathbf{N}} \\ \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} &= \underline{\mathbf{R}}^T_{\boldsymbol{\lambda}} + \underline{\underline{\mathbf{G}}}_1 \cdot \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} \end{aligned}\right\} \Rightarrow$$

$$\left.\begin{aligned} \underline{\mathbf{R}}_{d\mathbf{N}} &= (\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_2)^{-1} \cdot \left(\underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}} + \underline{\mathbf{G}} \cdot diag\left(\langle d\mathbf{N}\rangle\right)\right) \\ \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} &= (\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_1)^{-1} \cdot \underline{\mathbf{R}}^T_{\boldsymbol{\lambda}} \end{aligned}\right\}, \tag{A19}$$

where the second equation is solved in the beginning and then substituted into the first (after the appropriate rearrangement of $\underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}}$).

For scenario (**I**) $\mathbf{R}_{d\mathbf{N}}(\tau), \mathbf{G}(\tau) \Rightarrow \hat{\mathbf{R}}_{\boldsymbol{\lambda}}(\tau)$ the solution is also straightforward:

$$\left.\begin{aligned} \underline{\mathbf{R}}_{d\mathbf{N}} &= \underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}} + \underline{\mathbf{G}} \cdot diag\left(\langle d\mathbf{N}\rangle\right) + \underline{\underline{\mathbf{G}}}_2 \cdot \underline{\mathbf{R}}_{d\mathbf{N}} \\ \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} &= \underline{\mathbf{R}}^T_{\boldsymbol{\lambda}} + \underline{\underline{\mathbf{G}}}_1 \cdot \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} \end{aligned}\right\} \Rightarrow$$

$$\left.\begin{aligned} \underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}} &= \left(\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_2\right) \cdot \underline{\mathbf{R}}_{d\mathbf{N}} - \underline{\mathbf{G}} \cdot diag\left(\langle d\mathbf{N}\rangle\right) \\ \underline{\mathbf{R}}^T_{\boldsymbol{\lambda}} &= \left(\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_1\right) \cdot \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} \end{aligned}\right\} \tag{A20}$$

For scenario (**II**) $\mathbf{R}_{d\mathbf{N}}(\tau), \mathbf{R}_{\boldsymbol{\lambda}}(\tau) \Rightarrow \hat{\mathbf{G}}(\tau)$ we will reorganize the equations and the matrix notation. Let us rewrite the first equation of Eq. A12 in the following way:

$$\begin{aligned} \mathbf{R}_{d\mathbf{N}}(\tau) &= \mathbf{R}_{\boldsymbol{\lambda} d\mathbf{N}}(\tau) + \mathbf{G}(\tau) \cdot diag\left(\langle d\mathbf{N}\rangle\right) + \int_{-\infty}^{\tau} \mathbf{G}(\tau - u)\mathbf{R}_{d\mathbf{N}}(u)du \\ &= \mathbf{R}_{\boldsymbol{\lambda} d\mathbf{N}}(\tau) + \mathbf{G}(\tau) \cdot diag\left(\langle d\mathbf{N}\rangle\right) + \int_{0}^{\infty} \mathbf{G}(u)\mathbf{R}_{d\mathbf{N}}(u - \tau)du \\ \mathbf{R}^T_{d\mathbf{N}}(\tau) &= \mathbf{R}^T_{\boldsymbol{\lambda} d\mathbf{N}}(\tau) + diag\left(\langle d\mathbf{N}\rangle\right) \cdot \mathbf{G}^T(\tau) + \int_{0}^{\infty} \mathbf{R}^T_{d\mathbf{N}}(u - \tau)\mathbf{G}^T(u)du \end{aligned} \tag{A21}$$

This equation, written in the matrix form is:

$$\underline{\mathbf{R}}^T_{d\mathbf{N}} = \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} + \underline{\underline{\langle d\mathbf{N}\rangle}} \cdot \underline{\mathbf{G}}^T + \underline{\underline{\mathbf{R}}}^T_{d\mathbf{N}} \cdot \underline{\mathbf{G}}^T,$$

where the matrix $\underline{\underline{\langle d\mathbf{N}\rangle}}$ is a block diagonal matrix with blocks of $diag\,(\langle d\mathbf{N}\rangle)$ replicated N times (that corresponds to τ_{max}) on its diagonal to match the dimensions of the matrix $\underline{\underline{\mathbf{R}}}^T_{d\mathbf{N}}$. $\underline{\underline{\mathbf{R}}}^T_{d\mathbf{N}}$ is a block Toeplitz matrix with the block vector $\underline{\mathbf{R}}^T_{d\mathbf{N}}$ as its first block row and block column (note, that transpose is applied within-the-blocks, so that for the univariate case there is effectively no transpose):

$$\underline{\underline{\mathbf{R}}}^T_{d\mathbf{N}} \triangleq d\tau \cdot \begin{bmatrix} \mathbf{R}^T_{d\mathbf{N}}(0) & \mathbf{R}^T_{d\mathbf{N}}(1) & \mathbf{R}^T_{d\mathbf{N}}(2) & \cdots & \\ \mathbf{R}^T_{d\mathbf{N}}(1) & \mathbf{R}^T_{d\mathbf{N}}(0) & \ddots & \ddots & \\ \mathbf{R}^T_{d\mathbf{N}}(2) & \ddots & \ddots & & \\ \vdots & \ddots & & & \mathbf{R}^T_{d\mathbf{N}}(1) \\ & & & \mathbf{R}^T_{d\mathbf{N}}(1) & \mathbf{R}^T_{d\mathbf{N}}(0) \end{bmatrix} \tag{A22}$$

This, together with the matrix form of the second equation of Eq. A12 brings us to a couple of equations:

$$\begin{aligned} \underline{\mathbf{G}}^T &= \left(\underline{\underline{\langle d\mathbf{N}\rangle}} + \underline{\underline{\mathbf{R}}}^T_{d\mathbf{N}}\right)^{-1} \left(\underline{\mathbf{R}}^T_{d\mathbf{N}} - \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}}\right) \qquad (*) \\ \underline{\mathbf{R}}^T_{\boldsymbol{\lambda} d\mathbf{N}} &= \left(\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{G}}}_1\right)^{-1} \cdot \underline{\mathbf{R}}^T_{\boldsymbol{\lambda}} \qquad (**) \end{aligned} \tag{A23}$$

These can be solved iteratively:

(i) Start with a random $\underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}}$
(ii) Find $\underline{\mathbf{G}}$ from (∗)
(iii) Build matrix $\underline{\underline{\mathbf{G}}}_1$
(iv) Find $\underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}}$ from (∗∗)
(v) Goto ii)

We can alternatively set the initial condition to $\underline{\mathbf{R}}_{\boldsymbol{\lambda} d\mathbf{N}} = \underline{\mathbf{R}}_{\boldsymbol{\lambda}}$, which corresponds to $\underline{\mathbf{G}} = \underline{0}$.

This iterative solution converges very rapidly and, in practice, a single iteration brings us very close to the final solution.

frontiers in
COMPUTATIONAL NEUROSCIENCE

METHODS ARTICLE
published: 22 September 2010
doi: 10.3389/fncom.2010.00127

Surrogate spike train generation through dithering in operational time

Sebastien Louis[1], George L. Gerstein[2], Sonja Grün[1] and Markus Diesmann[1,3]*

[1] *RIKEN Brain Science Institute, Wako-shi, Japan*
[2] *Department of Neuroscience, University of Pennsylvania, Philadelphia, PA, USA*
[3] *RIKEN Computational Science Research Program, Wako-shi, Japan*

Edited by:
Philipp Berens, Max-Planck-Institute for Biological Cybernetics, Germany

Reviewed by:
Zoltan Nadasdy, Seton Brain and Spine Institute, USA
Matthew Harrison, Brown University, USA

****Correspondence:***
Markus Diesmann, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan.
e-mail: diesmann@brain.riken.jp

Detecting the excess of spike synchrony and testing its significance can not be done analytically for many types of spike trains and relies on adequate surrogate methods. The main challenge for these methods is to conserve certain features of the spike trains, the two most important being the firing rate and the inter-spike interval statistics. In this study we make use of operational time to introduce generalizations to spike dithering and propose two novel surrogate methods which conserve both features with high accuracy. Compared to earlier approaches, the methods show an improved robustness in detecting excess synchrony between spike trains.

Keywords: surrogate data, dithering, operational time, spike synchrony

INTRODUCTION

Surrogate generation has become a widespread tool for the statistical analysis of parallel spike trains (see Grün, 2009 for a review). As trial shuffling (Gerstein and Perkel, 1972) is limited to data consisting of a set of trials originating from an identical stochastic process, within trial approaches have been developed. In particular, dithering (Date et al., 1998) is often used in cross-correlation analysis and repeating pattern analysis, with the aim of identifying the time scales at which the neural code may be operating. The methods consist in randomly shifting individual spikes (Date et al., 1998; Nadasdy et al., 1999; Hatsopoulus et al., 2003; Shmiel et al., 2006; Stark and Abeles, 2009), patterns of spikes (Harrison and Geman, 2009), or the whole spike train (Perkel et al., 1967; Pipa et al., 2008) by an amount sufficient to destroy fine temporal spiking. A commonly tested hypothesis states that the firing rates of neurons are sufficient to explain the statistics of fine temporal spiking patterns. Rejecting such a hypothesis could suggest a form of coding beyond that of rate coding. One example, which we focus on in this study, is excess synchrony.

Unfortunately, spike dithering alters the original data in two undesirable ways; it smoothes the rate profile and distorts the inter-spike interval (ISI) distribution toward that of a Poisson process (Pazienti et al., 2008). We demonstrate in the present study that these effects need to be taken into consideration before applying the method to experimental data. Indeed, the outcome of a synchrony or pattern analysis is entirely determined by the adequacy of the surrogate method (Grün, 2009; Louis et al., 2010). Modifying the rate profile or the interval statistics is likely to affect the coincidence count statistics, and in turn could give grounds for false positive (FP) results. This is the main criticism against excess synchrony detection and Unitary Events (Grün et al., 2002a,b). This observation becomes all the more important as the number of parallel spike trains being analyzed increases. For example in the analysis of spatio-temporal patterns across multiple neurons (Nadasdy et al., 1999; Abeles and Gat, 2001). Previous work (Gerstein, 2004; Pipa et al., 2008; Harrison and Geman, 2009) only addressed the conservation of the ISI distribution, or the firing rate profile (Smith and Kohn, 2008), up to a precision determined by the dither width chosen. In the present study we propose dithering methods which simultaneously conserve both features with a higher level of precision.

The estimation of the intensity function within or across trials certainly only constitutes an approximate source of information, however we show that it can and should be used in the implementation of dithering methods. It may be argued that provided the rate profile, one can simply sample from it. That is, use it along with the ISI distribution, as a parameter of a chosen model of spike generation. The problems with this are that the spike count is not necessarily preserved, and strong assumptions on the process itself have to be made; for example that it is a Poisson process. The issues can be overcome if the estimated features of the process are integrated into the dithering method. The immediate benefit of having an estimate of the spike rate is that the process can be approximately transformed to a unit rate stationary process through rescaling of the time axis; a mapping to operational time. Once a process is stationary, the constraints on the dithering method are considerably relaxed.

We begin our study by indicating how a time invariant dithering applied in operational time instead of real time leads to a perfect conservation of the rate profile. However, the effective transformation undergone by the spikes in real time is not entirely obvious. We address this issue and demonstrate how a uniform dither in operational time maps to a variable range, non-uniform dither following the rate profile itself, in real time. This is verified through simulations.

A fixed time scale hypothesis can be tested for by fixing the dither range in real time and replicating directly the effect of an operational time dither by modulating the dithering profile according

to the firing rate. In so doing, simulations and calculations point to the use of taking a power of the rate profile as the shape of the dithering distribution. We then extend the two known methods of joint-ISI based dithering (Gerstein, 2004) and spike train shifting (Pipa et al., 2008) and apply them in operational time, leading to superior conservation properties.

After demonstrating in how far these methods are capable to preserve the firing rate profile and the ISI distribution, we proceed to compare the different surrogate methods in their ability to provide a good implementation of the null-hypothesis for testing the presence of excess precise spike coincidences. The quality of the surrogates is evaluated by testing for FP and false negative (FN) outcomes. Finally, we apply the different surrogate methods to responses of neurons recorded in the primary visual cortex of the anesthetized macaque monkey (Aronov et al., 2003). Preliminary results of the present study have been presented in abstract form (Diesmann et al., 2009).

VARIANTS OF DITHERING

DITHERING IN REAL TIME

We view the spike train of a neuron as a point process with continuous conditional intensity function $\lambda(t \mid H_t)$, where H_t is the history of the process up to time t. We refer from now on to $\lambda(t \mid H_t)$ as the rate profile of the neuron. Dithering as outlined above consists in shifting individual spikes randomly around their initial position in time, following a *dither distribution*. In the most general case, a dithering method $\mathcal{D}$ will map a spike train $\mathbf{t} = \{t_i \mid i = 1,\ldots,N\}$ of N spikes to

$$\mathcal{D}(\mathbf{t}) = \{t_i + \xi_i(\mathbf{t}) \mid i = 1,\ldots,N\}, \tag{1}$$

where the $\xi_i(\mathbf{t})$ are random variables distributed as $\xi_i(\mathbf{t}) \sim D_i$ and the D_i are dither distributions associated to each spike t_i. They can potentially depend on the spike train as a whole and be different for each spike. In the case of a uniform dither method with range $\pm w$ (dither width), the above simplifies to

$$\mathcal{D}(\mathbf{t}) = \{t_i + \xi_i \mid i = 1,\ldots,N\}, \tag{2}$$

with the ξ_i being independent and identically distributed random variables $\xi_i \sim D = U(-w,w)$. A further simplification is obtained by dithering all spikes together by the same amount $\xi \sim D$ such that $\mathcal{D}(\mathbf{t}) = \{t_i + \xi \mid i = 1,\ldots,N\}$, representing the spike train shifting surrogate (Pipa et al., 2008). Assuming an inhomogeneous Poisson process with intensity function $\lambda(t)$, the effect of dithering individual spikes with a fixed distribution D throughout time, yields the profile $\lambda_D(t)$, with

$$\begin{aligned}\lambda_D(t) &= \lim_{\delta u \to 0} \sum_i \lambda(t-u_i) D(u_i) \delta u, \\ &- \sum_{i \neq j} \lambda(t-u_i)\lambda(t-u_j) D(u_i) D(u_j) (\delta u)^2 + \ldots, \\ &= \int \lambda(t-u) D(u) du,\end{aligned} \tag{3}$$

where the sums and the integral are taken on the support of $D(u)$. The above, which is a simple convolution, is obtained by first applying the inclusion–exclusion principle (Grimaldi, 2003) and then letting δu tend to 0, removing all terms of order larger than 1 in δu. This result is in fact equivalent to a translated Poisson process, which itself is a Poisson process (Snyder and Miller, 1991). We note that for a finite time resolution the higher-order corrections are present and may be important. The result holds up to edge effects and so is valid at time t if the dithering operation is applied on all spikes in the region $[t - w, t + w]$. An immediate consequence of Eq. 3 is that a constant rate $\lambda(t) = c$ is always preserved through a constant dither operation

$$\lambda_D(t) = c \int D(u) du = \lambda(t). \tag{4}$$

Furthermore, any profile which is a linear function of time $\lambda(t) = at + b$ will be conserved under a dithering operation with mean displacement 0 ($\mathbb{E}[\xi] = 0$)

$$\begin{aligned}\lambda_D(t) &= at + b - a \int u D(u) du, \\ &= \lambda(t) - a\mathbb{E}[\xi] = \lambda(t).\end{aligned} \tag{5}$$

Thus given a mapping of non-stationary point processes to stationary ones, it is possible to implement a rate profile preserving dithering operation. First the process is mapped to a stationary one, the dithering is applied, and then the process is mapped back.

DITHERING IN OPERATIONAL TIME

The desired mapping to a stationary process is achieved by transforming t to a new variable known as operational time $\tilde{t}$ (Cox and Isham, 1980)

$$\tau(t) = \int_0^t \lambda(u) du = \mathbb{E}\big[N[0,t]\big] = \tilde{t}, \tag{6}$$

where $N[0,t]$ is the number of events on the interval $[0,t]$. The last equality means that in operational time, the point process becomes a process with unit rate. In other words, this transformation can be seen as a rescaling of the time axis, such that the rate now becomes constant at 1 Hz (Brown et al., 2001). So for a dithering operation $\mathcal{D}$ with fixed dither distribution, the above equations tell us that for an inhomogeneous Poisson point process, $\mathbf{t}$ and $(\tau^{-1} \circ \mathcal{D} \circ \tau)(\mathbf{t})$, where $\tau(\mathbf{t}) = \{\tau(t_i) \mid i = 1,\ldots,N\}$ are sampled from the same rate profile. Therefore a simple resampling procedure could consist in a fixed width uniform dithering (UD) in operational time.

To understand the effect of such a transformation in real time, we introduce the Perron–Frobenius (PF) operator $\mathcal{L}$ (Beck and Schlögl, 1995), used in non-linear dynamics to describe the time evolution of densities in phase space. After an iteration of the map f on the density $\rho(y)$, the output function becomes

$$(\mathcal{L}\rho)(y) = \sum_{x \in f^{-1}(y)} \frac{\rho(x)}{|f'(x)|}, \tag{7}$$

where $f'(x) = df(x)/dx$. In the present case, we wish to map a UD distribution from operational to real time. Thus f becomes the inverse mapping τ^{-1} and $\rho(y)$ becomes our dither distribution $D(\tilde{t}) = 1/(\tilde{t}_+ - \tilde{t}_-)$ for $t \in [\tilde{t}_-, \tilde{t}_+]$. Assuming $\tau(t)$ is a strictly increasing function, applying the PF operator yields

$$
\begin{aligned}
(\mathcal{L}D)(t) &= \frac{D(\tilde{t})}{\tau^{-1'}(\tilde{t})} \\
&= \frac{(\tau' \circ \tau^{-1} \circ \tau)(t)}{\tilde{t}_+ - \tilde{t}_-} \\
&= \frac{\frac{d}{dt}\int_0^t \lambda(u)du}{\tilde{t}_+ - \tilde{t}_-} \\
&= \frac{\lambda(t)}{\tilde{t}_+ - \tilde{t}_-}.
\end{aligned} \tag{8}
$$

The above states that a UD distribution in operational time, is equivalent to a dithering distribution following the rate profile, normalized over the mapped range $\left[\tau^{-1}(\tilde{t}_-), \tau^{-1}(\tilde{t}_+)\right] = [t_-, t_+]$. An obvious consequence of Eq. 8 is that the dither boundaries in real time are now modulated directly by the rate profile. Or conversely, a fixed dither width in real time will transform to a variable dither width in operational time (**Figure 1**). For a fixed range in operational time, the larger the firing rate, the smaller the effective dither width in real time (illustrated in **Figure 2**).

A fixed range in operational time may constitute an interesting surrogate generation method, as it preserves the estimated rate profile exactly, and has an intuitive interpretation: in order to stand out from the noise, spike synchrony needs to be more precise in regions of high rate requiring only a smaller dither for effective destruction. However, if we fix the dithering boundaries in real time, to $\pm w$, say, this produces a dithering distribution following the rate profile as shown in Eq. 8, with mapped boundaries $[t - w, t + w]$, meaning

$$
D_t(u) = \frac{\lambda(t+u)}{\tilde{t}_+ - \tilde{t}_-} = \frac{\lambda(t+u)}{\int_{t-w}^{t+w} \lambda(x)dx}. \tag{9}
$$

Non-uniform dithering distribution

Following Gerstein's (2004) use of the square root function to scale in the dithering, we allow here for a general dithering distribution shaped according to a composition $g \circ \lambda(t)$ of the rate profile. Using Eqs. 9 and 3 with a time dependent dithering distribution, a general dithering method following a function of the local rate profile would map this rate according to

$$
\begin{aligned}
\lambda_D(t) &= \int \lambda(t-u) D_{t-u}(u)du \\
&= (g \circ \lambda)(t) \int_{t-w}^{t+w} \frac{\lambda(y)}{\int_{y-w}^{y+w} (g \circ \lambda)(x)dx} dy.
\end{aligned} \tag{10}
$$

The question now becomes: how well is $\lambda(t)$ preserved, depending on the choice of g. Below we show how the two obvious choices, $g(x) = 1/2w$ (uniform dither in real time) and $g(x) = x$ (uniform dither in operational time with fixed range in real time), both affect $\lambda(t)$ in negative but opposite ways. For this we allude to Jensen's inequality (Gradshteyn and Ryzhik, 2000) which in the continuous case states that if φ is a convex function, then

$$
\varphi\left(\int_a^b f(x)dx\right) \le \int_a^b \frac{\varphi\big((b-a)f(x)\big)}{b-a} dx. \tag{11}
$$

It is straightforward to show that for λ convex on the interval $[y - w, y + w]$ the above yields

$$
2w\lambda(y) \le \int_{y-w}^{y+w} \lambda(x)dx, \tag{12}
$$

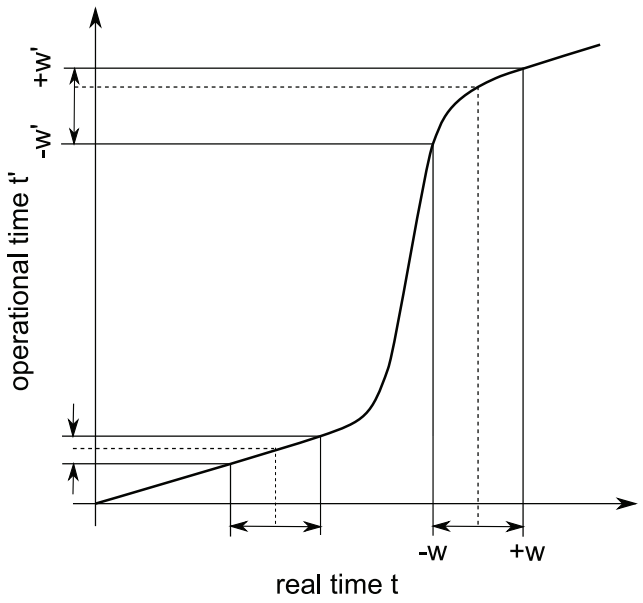


FIGURE 1 | Illustration of the conversion from real (horizontal) to operational (vertical) time. The thick curve shows a cumulative rate profile, which serves for the transformation from real time to operational time. The dashed lines indicate the positions of two example spikes prior to dithering. A constant dither window $\pm w$ in real time is converted to non-constant dither windows $\pm w'$ in operational time.

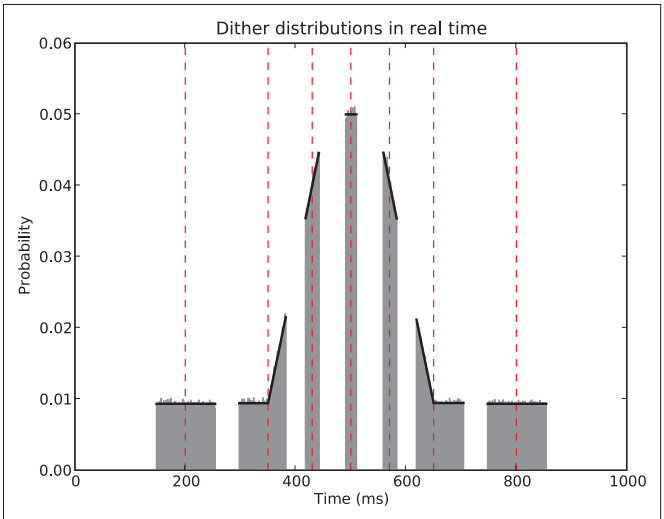


FIGURE 2 | Uniform dithering in operational time. Real time distributions of dithers at selected points (red vertical dashed lines) in the time course of the rate. For each time point an identical uniform, fixed width dither of $w' = \pm 30$ ms in operational time was used. The dither distributions (gray, bin width 5 ms) at each original spike position are obtained empirically by 100000 dither repetitions of each spike. The rate profile used for the mapping is shown in black.

while the converse holds for a concave λ. Starting with UD [$g(x) = 1/2w$] and assuming a locally convex profile, combining Eqs. 10 and 12 gives

$$\lambda_D(t) = \int_{t-w}^{t+w} \frac{\lambda(y)}{2w} dy \geq \lambda(t), \tag{13}$$

where $\lambda_D(t)$ can be interpreted as the "dithered" rate profile. Thus surrogates generated by UD have an increase in rate in convex regions of the profile, and conversely a decrease in concave regions, relative to the original rate profile. As expected, this is equivalent to a smoothing of the original profile (see **Figure 4**). For the case in which the dithering distribution follows the rate profile itself ($g(x) = x$), we obtain

$$\begin{aligned} \lambda_D(t) &= \lambda(t) \int_{t-w}^{t+w} \frac{\lambda(y)}{\int_{y-w}^{y+w} \lambda(x) dx} dy \\ &\leq \lambda(t) \int_{t-w}^{t+w} \frac{\lambda(y)}{2w\lambda(y)} dy \\ &\leq \lambda(t) \cdot 1. \end{aligned} \tag{14}$$

The effect here is the opposite to that of UD; $\lambda_D(t)$ now exaggerates the non-stationarities, decreasing in convex regions and increasing in concave ones, relative to $\lambda(t)$.

JOINT-ISI DITHERING

A similar exaggeration of the profile was previously noted in Gerstein (2004), where the feature to be preserved is the ISI distribution. In this surrogate method, the joint-ISI distribution is constructed from pairs of successive intervals (see **Figure 3**). Each spike is situated on the joint-ISI surface according to its pre- and post-intervals and moving a spike is equivalent to displacing the point along the perpendicular to the main diagonal in the joint-ISI plot. The surrogate method is then based on dithering the spikes following the local joint-ISI distribution, in the same way as dithering following the rate profile.

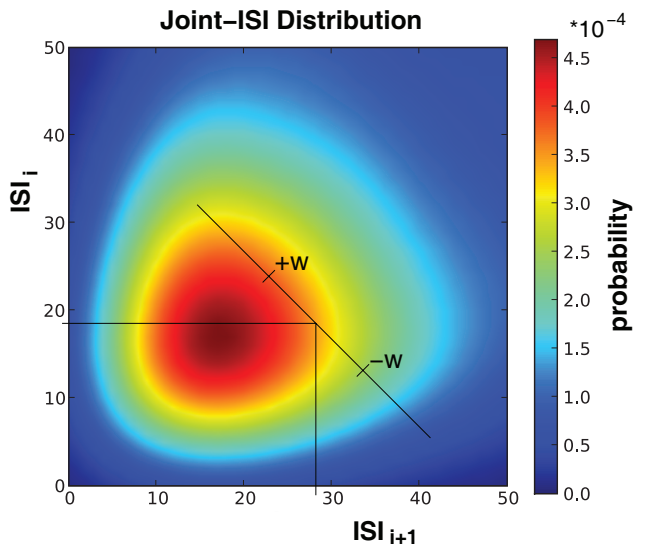


FIGURE 3 | Illustration of Joint-ISI dithering. Joint-ISI distribution obtained from 1000 spike train realizations (same parameters as in **Figure 4**). The two perpendicular lines mark one particular spike with a particular ISI relative to its preceding (ISI_i) and its following spike (ISI_{i+1}). The spike is dithered within the interval [$-w$, $+w$] according to the non-uniform probability given by the joint-ISI distribution (color coded).

Gerstein (2004) observed that the peakedness (kurtosis) of the ISI distribution of the resulting surrogates is increased relative to the original shape; meaning the surrogate spike trains are more regular in their activity than the original. We now understand that this occurs as result of Eq. 14; each of the perpendicular cuts of the joint-ISI distribution sees its shape emphasized increasing the peakedness of the two-dimensional distribution. Consequently the shape of the marginal ISI distribution is also emphasized.

To overcome this effect Gerstein (2004) proposes to take the square root of the joint-ISI distribution before applying the dithering procedure. The surrogates then exhibit an ISI distribution very close to the original for dither widths on the order of 10 ms. Setting $g(x) = \sqrt{x}$ in Eq. 10 does not lead to $\lambda_D(t) = \lambda(t)$, however its smoothing property counterbalances the emphasizing property of the profile itself, providing a significant improvement as can be seen in **Figure 4**.

ISI CONSERVING DITHERS IN OPERATIONAL TIME

Combining both, the ideas of operational time for rate conservation and joint-ISI based dithering for interval conservation, we propose a novel surrogate method. It consists in first mapping the spikes to operational time, then applying a joint-ISI based dithering in operational time with real time fixed boundaries, before finally mapping the spikes back to real time.

The dithering of the whole spike train (Pipa et al., 2008), that is adding a single uniformly distributed shift to all spikes, can also be applied in operational time. If the process is a renewal process in operational time, then such a surrogate constitutes an ideal surrogate, as it conserves both the ISI and rate features of the real time process. However the dithering ranges varies depending on the position of the individual spikes relative to the rate profile.

We show that both methods conserve the spike rate as well as the ISI distribution.

SIMULATION METHODS

Spike dithering is now widely used in the detection of excess synchrony in parallel spike trains and the investigation of patterns and temporal coding. However its exact effect on the statistics of the spike trains has not been studied in detail.

An analysis is only useful for the experimentalist if it is based on biologically realistic rate profiles and ISI distributions. Due to the restricted power of our present theoretical tools an appropriate level of realism is only achievable by computer simulation. Fortunately, the progress in computer hardware and methods for trivial parallelization in high-level programming languages has now considerably expanded our capabilities compared to the time when dithering was first considered. The algorithms described below are implemented in Python (Langtangen, 2006) and executed in parallel using the techniques described in Denker et al. (2010). Example code for implementing dithering in operational time is available at www.spiketrain-analysis.org.

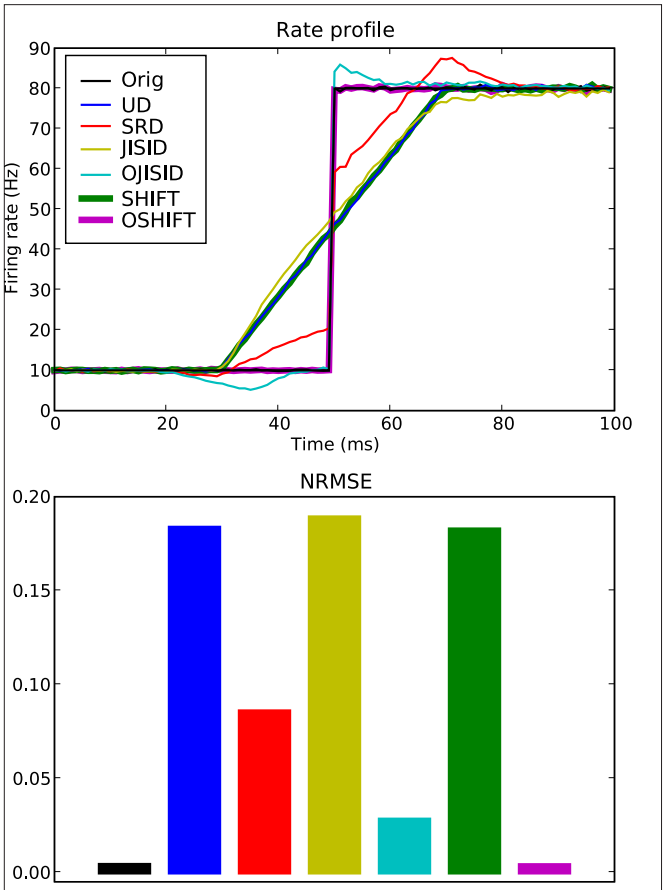


FIGURE 4 | Conservation of rate profile at ± 20 ms dither width. A total of 5×10^5 trials are used, with an underlying rate following the black curve in the top panel. The estimated rate profiles (top panel) yielded by the surrogate methods are shown in color (see legend and **Table 1**). The lower panel shows a bar plot of the normalized root mean square error (NRMSE) of each surrogate method compared to the original PSTH (resolution 1 ms). The leftmost bar is the NRMSE of the original PSTH compared to the true rate profile.

It is intuitively clear that spike dithering works for a Poisson process with a constant intensity in the dithering interval (see Diesmann et al., 2009 for a thorough introduction). Thus as soon as the spike train exhibits temporal structure such as refractory periods, dithering may be questionable. In addition, the question of the adequate choice of dithering width needs to be addressed. If the width is kept too small compared to the tolerated jitter in synchrony, the sensitivity of the detection may be affected (Pazienti et al., 2008). For excess synchrony detection, the dither width clearly depends on both the hypothesis being tested for (the allowed jitter) and the requirement to conserve the firing rate profile.

In the following sections we compare the different surrogate methods listed in **Table 1** in two steps. In the first step we examine the methods' ability to preserve the rate profile of the spike trains as well as their ISI distribution. Both are primary features of spike trains which ought to be conserved adequately. In the second step we quantify how the potential non-preservation of these features impacts spike correlation analysis, using coincidences as an example, and may lead to erroneous conclusions on the spike correlation structure of the data.

BENCHMARK DATA

In order to compare the different surrogate methods, we simulate continuous time spike trains exhibiting both rate non-stationarities and non-Poisson ISI statistics. The standard rate profile used is shown by the black curve in **Figure 4** consisting of a single step, with a base rate of 10 Hz. In our study we parameterize the profile by the size of the step $\Delta\lambda$. In the FP and FN analysis below we use 50 trials and restrict $\Delta\lambda$ to the range between 0 and 100 Hz, leading to an upper firing rate of at most 110 Hz.

The duration of 100 ms corresponds to the typical length of the analysis window in time-resolved correlation analysis (Grün, 2009). The individual trials are produced by mapping a unit rate gamma process (renewal process with gamma distributed ISIs and rate 1 Hz) through the inverse function of the integrated rate profile (τ^{-1} above) to real time. In other words a time rescaled stationary process. The spiking regularity is thereby defined through the shape parameter of the gamma process in operational time γ_{op} or alternatively its coefficient of variation CV_{op}, which due to the deterministic mapping leads to a constant $CV = CV_{op}$ in real time. The resulting spike trains exhibit non-stationary firing rates and a non-trivial total ISI statistics.

IMPLEMENTATION OF DITHERING METHODS

In the order of **Table 1**, we start with UD in real time. As explained in the previous sections, we implement UD by adding a random number drawn from the uniform distribution $U(-w,w)$ independently to each spike time in the spike train. Next for the rate profile dependent method SRD, we first estimate the firing rate profile through the peristimulus time histogram (PSTH) constructed over the trials on a 1-ms resolution. The amount of smoothing applied depends on the number of trials; for 50 trials we choose a 10-ms Gaussian smoothing. From this smoothed PSTH we construct a linearly interpolated function to provide us with a continuous rate profile $\lambda(t)$. Then each spike t_i is dithered according to the normalized segment of the exponentiated rate profile: $t_i \to s_i$ where $\mathbf{P}(s_i = t) = \lambda^\beta(t) / \int_{t_i-w}^{t_i+w} \lambda^\beta(u)du$ for $t \in [t_i - w, t_i + w]$ and 0 otherwise. SRD uses $\beta = 0.5$.

The SHIFT surrogate is constructed simply by adding the same random number drawn from the uniform distribution $U(-w,w)$ to each spike time in the spike train. Thus UD and SHIFT constitute the limits of the pattern-jittering method proposed by Harrison and Geman (2009). In broad terms, this method fixes a threshold for the ISIs of interest. ISIs larger than this threshold allow for a segmentation of the data into patterns, which are dithered independently (the same random number is added to each spike of a pattern). In UD, the patterns are individual spikes and thus the ISI threshold is at 0, leading to a maximal perturbation of the ISI distribution. In SHIFT, the whole spike train is a single pattern and the ISI threshold is larger than the largest ISI, leading to a minimal variability. Observing the performance of these two limiting methods will give us an idea on where to situate pattern-jittering, with respect to other methods. We also extend SHIFT to an operational time version OSHIFT, which dithers the whole spike train in operational time. The mapping is done through the integrated PSTH. The real time dither range is thus no longer fixed.

Table 1 | The investigated dither methods and their features.

Dither time	Dither distribution	ISI conservation	Abbreviation
Real	Uniform	No	UD
Real	$\sqrt{\text{Rate}}$	No	SRD
Real	$\sqrt{\text{J}-\text{ISI}}$	Yes	JISID
Operational	$\sqrt{\text{OJ}-\text{ISI}}$	Yes	OJISID
Real	Uniform	Yes	SHIFT
Operational	Uniform	Yes	OSHIFT

The first column indicates in which time coordinate the spikes are dithered. The second column lists the shape of the dithering distribution. UD (uniform dithering), SRD (dithering according to the normalized square rooted rate profile), JISID (dithering according to the joint-ISI distribution), and SHIFT (spike train shifting) are thus dithered in real time, with the exception of OJISID (dithering according to the joint-ISI distribution in operational time) and OSHIFT (spike train shifting in operational time). UD, SHIFT, and OSHIFT use a uniform distribution whereas SRD uses the square root of the estimated firing rate profile itself. JISID and OJISID use the joint-ISI distribution constructed from the data, in real time and operational time respectively. The third column indicates whether the method attempts to conserve the ISI statistics and the fourth shows the abbreviations used in the text and figures for each of the dithering methods.

For the ISI dependent methods, JISID and OJISID, we first construct the JISI matrix on a 1-ms resolution (in real and rescaled operational time respectively). The size $I_{\max}$ of this matrix was set to 100 ms × 100 ms. In general, the choice of $I_{\max}$ will depend on the mean and standard deviation of the ISI distribution. Once filled, the matrix is square rooted (Gerstein, 2004). In the case of a small number of trials, the matrix is additionally smoothed with a 2D Gaussian of width 3 ms. From this square rooted matrix we proceed to construct a 2D interpolated function $J(x,y)$ through bilinear interpolation (Press et al., 2007), where $0 < x,y \leq I_{\max}$ are the pre- and post-inter-spike intervals respectively. The dithered position of a spike t_i with pre- and post-intervals x_i and y_i is then given by $t_i \rightarrow t_i + z_i$ with $P(z_i = t) = J(x_i + t, y_i - t) / \int_{-w}^{w} J(x_i + u, y_i - u)du$ for $0 < x_i + t, y_i - t \leq I_{\max}$ and 0 otherwise. The spikes are dithered in parallel; that is the dither distribution is initially fixed for each spike at the beginning of procedure, based on the position of neighboring spikes. It may seem like a dynamic version, in which one first dithers (assuming the spikes are numbered) even spikes, then updates the joint-ISI coordinates before dithering the odd spikes, would be more accurate. However the conservation properties and the excess synchrony detection performance remain unaffected. The same procedure is used in the OJISID method once the spike trains were mapped to operational time. The joint-JISI matrix is then constructed in operational time. To make the matrix sizes compatible between the two methods we scaled the operational time back down to a duration of 100 ms, relative to a bin size of 1 ms. Spikes which fall out of the matrix are dithered uniformly within the initial dither width.

QUANTIFICATION OF FEATURE CONSERVATION

In order to have a reliable comparison of feature conservation across the surrogate methods, we simulate a total of 5×10^5 trials following the procedure outlined above. Except for UD and SHIFT which are independent of the rate profile, all methods make use of all the trials in estimating the rate profile and the JISI distribution. We then calculate the normalized root mean square error (NRMSE) of the surrogate PSTH obtained from all surrogate trials H_s with respect to the true PSTH obtained from the original trials H_T at a resolution of 1 ms as $\text{NRMSE} = 1/(H_T^{\max} - H_T^{\min}) \cdot \sqrt{\sum (H_T - H_S)^2 / N}$, where $H_T^{\max}$ and $H_T^{\min}$ are the maximum and the minimum spike count in the histogram of the original spike trains respectively and N is the number of bins of the histograms.

The NRMSE is also calculated for the respective ISI distributions (1 ms resolution) in a similar fashion, quantifying the destruction of the original ISI statistics.

FALSE POSITIVE AND FALSE NEGATIVE EVALUATION IN CORRELATION ANALYSIS

Feature conservation is only part of the assessment of a surrogate method. It may indicate its flaws and advantages, however it does not guarantee that it will be useful in the context of a specific analysis of the data. Here we concentrate on the example of spike correlation analysis, and show how surrogates can be used for testing for the presence or absence of correlation between spike trains and for deriving their significance. In the correlation analysis we are interested to detect the presence or not of excess precise spike synchrony, beyond that explained by the firing rate and the ISI statistics. We define a synchronous event by two spikes (one from each neuron) occurring within ±1 ms of each other. Surrogates serve as an implementation of the null-hypothesis of independent firing. By dithering the precise relationship of the spikes of the two spike trains is destroyed.

To evaluate the performance of the surrogates in the context of correlation analysis we look at FP rates in data containing no excess synchrony and FN rates in data containing excess synchrony. In the statistical analysis we are here following the terminology of Ventura et al. (2005). For each parameter configuration and surrogate method, the FP and FN rates are obtained as follows. We begin by generating 1000 data sets with the same parameter configuration. A data set consists of 50 trials (100 ms duration) of two parallel spike trains generated according to a defined rate profile and interval statistics as in the study of feature conservation estimation of single spike trains. In case of studying the FN rate, the parallel spike trains contain correlated spiking due to insertion of jittered (±1 ms) coincident spike events at rate λ_c. For the FP analysis, the insertion is omitted and the spike trains are independent on a fine temporal scale, but are correlated on a slower time scale due to correlated (identical) rate profiles.

For each data set, we produce 1000 surrogate versions. Each set of surrogates is analyzed as the original data set for the occurrences of coincident spike events. In each data set, the number of coincidences of an allowed temporal precision (here ± 1 ms) is counted. A coincidence is detected by testing if there is one spike (or more) of the second spike train within ±1 ms relative to a spike of the first (reference) spike train. If more spikes occur within an individual coincidence window, this is counted as one coincidence ("clipping"). From the coincidence counts derived from the surrogates we construct the surrogate coincidence count distribution, which serves to estimate the significance of the coincidence count of the original data by calculating the p-value.

Thus for one parameter configuration of the data and surrogate method, we obtain 1000 p-values (p_i for $i = 1,\ldots,1000$). Given a significance level α, which we fix to 0.01 in the following, we

convert these results into counts of positive (significant) results, i.e., $N_+ = \Sigma_i \varphi(p_i \leq \alpha)$, and counts of negative (non-significant) results, i.e., $N_- = \Sigma_i \varphi(p_i > \alpha)$, where $\varphi(x) = 1|0$ if x is true|false. If the chosen parameter configuration involves injected synchrony, then the FN rate in percentage is given by $100 \cdot N_-/N$ (where $N = N_+ + N_- = 1000$), i.e., the percentage of falsely undetected correlation. Conversely, if the parameter configuration does not involve injections, then the FP rate reads $100 \cdot N_+/N$ indicating the percentage of falsely detected correlation (Louis et al., 2010).

In general the FP rate (empirical type I error) does not coincide with the prespecified significance level α because the surrogate distribution only imperfectly resembles the distribution of coincidence counts of independent data. If the independent distribution is known a matched significance level α_m can be determined which restricts the FP rate to a prespecified value. However, with knowledge of the independent distribution, typically no surrogate method is required in the first place. The FN rate (empirical type II error) can be used to compare the sensitivity or test power of different surrogate methods if they are adjusted to produce the same FP rate. Differences in sensitivity may then, for example, originate from a different effectiveness of the surrogate methods in destroying injected coincidences.

CALIBRATION BASED ON SIMULATED DATA

EFFECT OF DITHER SURROGATES ON RATE PROFILE

We compare a total of six surrogate methods (UD, SRD, SHIFT, OSHIFT, JISID, and OJISID, listed in **Table 1**) in their ability to preserve the underlying rate profile. The dither width was fixed at 20 ms; intentionally large such as to accentuate the differences between methods. The parameters of the profile are $\Delta\lambda = 70$ Hz and $\gamma_{op} = 3$ (**Figure 4**).

We observe that the UD, SHIFT, and JISID methods perform worse (highest NRMSE), systematically deviating away from the rate in the vicinity of the step. The effect is intuitive for UD and SHIFT which lead to a smoothed profile, corresponding to a convolution as shown in Eq. 13. In the case of JISID, the effect can be understood by considering the likely positions of the interval borders in the JISI matrix given by the previous and the next spike in dependence of the rate profile. When the rate is increasing, as is the case at the step, the post-interval is likely to be smaller than the pre-interval, thus the JISI based dither will tend to recenter the spike and shift it back in time. Repeating this effect over trials leads to a systematic shift of the rate profile. The rate would be shifted in the other direction in downward transients.

Taking the square root of the profile as a dithering distribution more than halves the NRMSE, as can be seen from the performance of SRD. OJISID performs even better, apart from the slight emphasizing effect at the step, which was anticipated in Eq. 14. Finally OSHIFT is at the level of the variability in the original PSTH as it does not modify any of the statistics of the processes.

EFFECT OF DITHER SURROGATES ON ISI DISTRIBUTIONS

The same set of trials as in the previous section is then used to assess the conservation of the ISI distribution (**Figure 5**). The UD and SRD methods deviate most from the original distribution, showing the largest NRMSEs (computed for the surrogate distributions compared to the distribution of the original trials). Next we find the JISI method, which conserves the distribution far better. It is still quite far from the original distribution, which we attribute to the size of the dither window, which is larger than the width of the distribution. The method conserves the ISI distribution with much higher precision for dither widths on the order of 10 ms (not shown, see Gerstein, 2004). As anticipated, the OJISI method not only preserves the rate profile, it also yields surrogates with an ISI distribution even closer to the original, as can be seen from the reduced NRMSE.

However the most accurate methods are the SHIFT and OSHIFT methods, as expected. They perfectly conserve the ISI statistics of the processes. For SHIFT, this is obvious. For OSHIFT, the ISI statistics of the process in operational time are perfectly conserved, so applying exactly the same mapping to and from operational time leads to a perfect conservation in real time. However the ISI sequence of a single trial is modified, unlike in the SHIFT method.

Combining the last two results, we can safely conclude that the OSHIFT and OJISID methods are by far the most feature preserving surrogate methods amongst the six being compared.

EFFECT OF DITHER SURROGATES ON SENSITIVITY OF CORRELATION ANALYSIS

To see how the methods compare in the context of correlation analysis, we devised two separate analyses focusing on different parameters. The first analysis evaluates the dependence of FPs and FNs on the strength $\Delta\lambda$ of the non-stationarity in rate (**Figure 6**). In the second analysis $\Delta\lambda$ is set to 70 Hz and we investigate the dependence of FP and FN results on the coefficient of variation (CV) controlled by the shape parameter γ_{op} of the process (**Figure 8**). The task of the surrogates is to detect the presence of excess spike synchrony (±1 ms), beyond that explained by the firing rate and the ISI statistics. We set the dither range in the various surrogate methods to ±20 ms throughout this part of the study for the intended destruction of the precise temporal relationship between the spikes of the two neurons. This dither range is a lower bound for OSHIFT; spikes may be dithered by larger amounts in the low rate region. In addition, **Figure 7** shows a scenario where for the first analysis we progressively reduce the dither width of UD.

We begin with the rate dependence analysis (**Figure 6**) and clearly identify UD as the method with the highest FP rates (up to 35% for $\Delta\lambda = 100$ Hz) and correspondingly the lowest FN rates. Then comes the SRD method which attempts to better preserve the rate step but ignores the ISI statistics. With similar FP performances, we find SHIFT and JISID, producing FPs up to 10%. However the SHIFT method is clearly superior when looking at the FN rates which are consistently lower. Thus for the same accuracy, SHIFT is more sensitive than JISID.

The operational time methods OJISID and OSHIFT are considerably more conservative, with FP rates of at most 5%. However their level of sensitivity appears to be far lower, with high FN rates going up to 70%. In the Appendix we explain why in fact OSHIFT reaches the maximum sensitivity: The other surrogate methods are smoothing the rate profile. This leads to a distribution of coincidence counts which is shifted to a lower mean compared to the mean of the actual independent distribution. Consequently a smaller fraction of the dependent distribution is located to the left of the threshold coincidence count determined by the significance level. The FN rate appears to be reduced compared to OSHIFT but

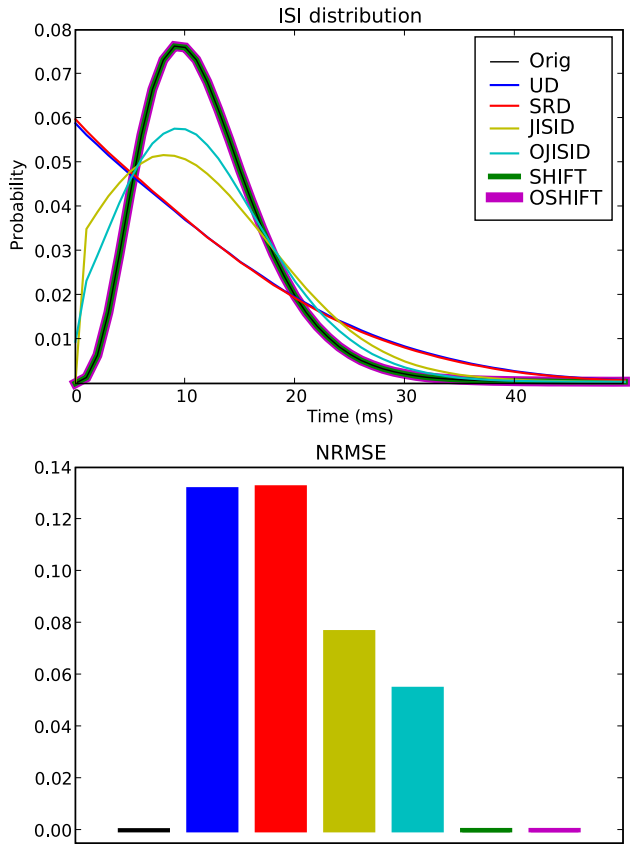


FIGURE 5 | Conservation of the ISI distribution at ±20 ms dither width. The top panel shows the original and estimated ISI distributions based on the various surrogates (colors, see legend). The original ISI distribution curve is covered both by SHIFT and OSHIFT traces. The lower panel shows a bar plot of the NRMSE performances of each surrogate method (colors as in top panel) compared to the ISI distribution of the original spike trains. The heights of the bars for SHIFT, OSHIFT and the NRMSE of the ISI distribution of the original spike trains compared to the true distribution (black) are amplified just to indicate that the NRMSE was computed for all cases. The data are the same as shown in **Figure 4.**

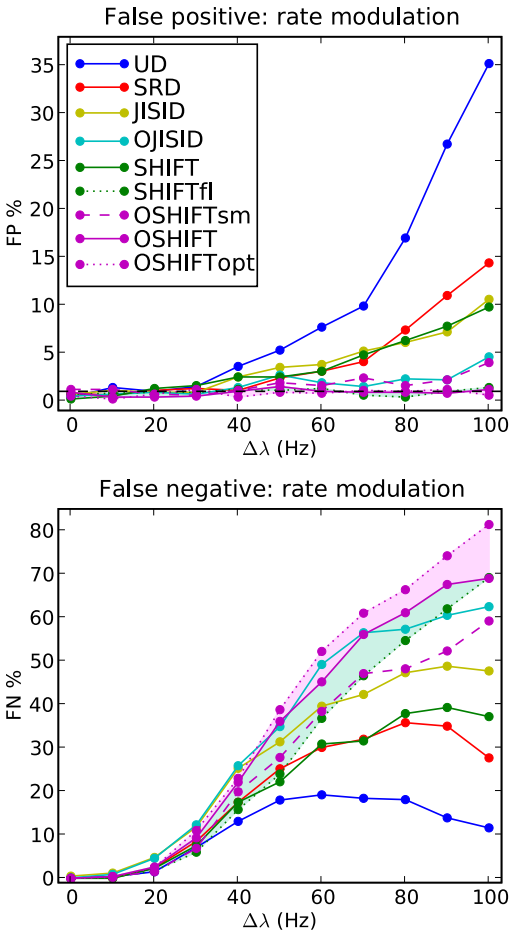


FIGURE 6 | FP and FN percentages as a function of the step amplitude Δλ, with fixed $\gamma_{op} = 3$. SHIFTfl is the SHIFT surrogate applied to a stationary process with the average spike rate of the rate step scenario. OSHIFTsm is a version of OSHIFT where the mapping to operational time constructed from the data is smoothed (10 ms Gaussian). OSHIFTopt is based on the true rate profile and the injection rate for the lower plot is $\lambda_c = 2$ Hz. A worst case estimate of the standard deviation of the FP/FN percentage is given by the error in the mean $\sqrt{p(1-p)/n} = 0.0158 = 1.6\%$, i.e., for a Bernoulli process with $p = 0.5$ and $n = 1000$ realizations as used here.

this is trivially so because the FP rate is larger than the significance level suggests. If a larger fraction of the independent distribution is to the right of the significance level also more of the dependent distribution is. The FN rates are only decisive if the FP rates are comparable.

Figure 6 shows three variants of OSHIFT: one in which the true rate profile is used (OSHIFTopt, short dashed purple), one in which the PSTH is smoothed before being used for the mapping (OSHIFTsm, long dashed purple) and one without the smoothing (OSHIFT, solid purple). An immediate observation is that smoothing the PSTH, which is initially quite variable due to the limited number of trials (50) induces a strong increase in FPs, from 0 to 5% at $\Delta\lambda = 100$ Hz. Thus integrating the PSTH for deriving the rate mapping provides an inherent reduction of noise and leads to a reliable mapping to and from operational time. SHIFT and JISID perturb the rate profile and exhibit the same dependence of the FP rate on Δλ. Using operational time, the FP rate of OJISI does not drop to the one of OSHIFT but reaches the level of OSHIFTsm. The reason is that the smoothing of the rate profile described above is also used in OJISI . The region shaded in light purple indicates the effect of using a PSTH instead of the true rate profile, and in some sense illustrates the distance to the optimal surrogate. Another way to explore the effect of non-stationarity in spike rate on a particular surrogate method is to compare the results for non-stationary data with the result of stationary data but otherwise similar characteristics. The curves labeled SHIFTfl (short dashed green) show the result of the SHIFT surrogate method applied to a data set generated by a stationary process parameterized by the average spike rate of the non-stationary process and identical regularity. The distance between SHIFT (solid green) and SHIFTfl (short dashed green) illustrates the cost of ignoring non-stationarity. For SHIFT the FP rate substantially increases with Δλ while it stays at the expected level for SHIFTfl. Thus, the smaller FN rate of SHIFT as compared to SHIFTfl is due to the higher FP rate of SHIFT. We

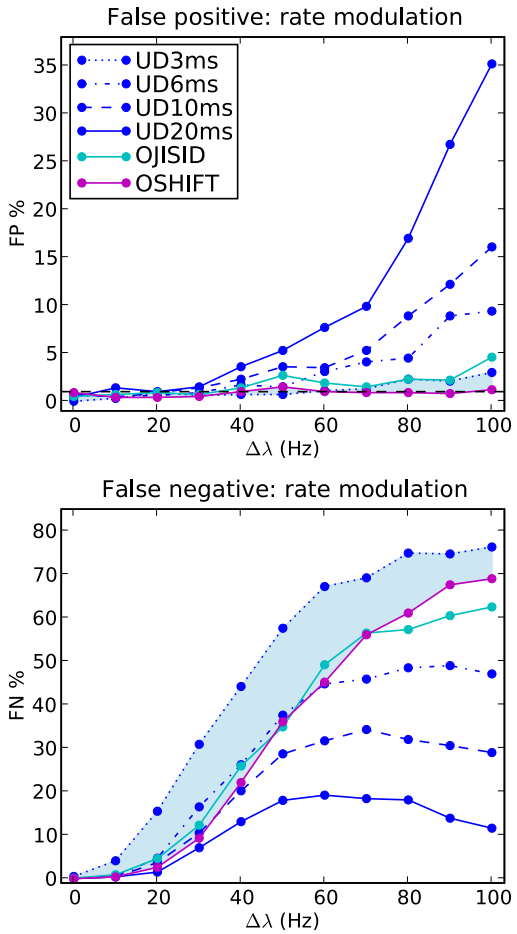


FIGURE 7 | Comparison of sensitivity at matched false positive rates. FP and FN rates as a function of the rate step amplitude, for varying dither widths of UD. Other parameters as in **Figure 6.**

find that OSHIFT (solid purple) lies between the optimal surrogate performances for non-stationary (OSHIFTopt, short dashed purple) and stationary (SHIFTfl, dashed green). The performance of OSHIFTopt and SHIFTfl is optimal in the sense that they destroy all coincidences, and have an FP rate at the expected level because the rate profiles are exactly respected. At the same FP rate no other method can have a lower FN rate. The rate of injected coincidences λ_c is stationary and the same in both cases. Therefore, in the data set with a rate step (OSHIFTopt) the distance between the number of coincidences in the data and the expected number of coincidences in the surrogates is smaller (see also Grün et al., 2003). This leads to the larger FN rate of OSHIFTopt compared to SHIFTfl illustrated by the conjunction of the light blue and light purple areas. The region shaded in light blue indicates that the increase in the FN rate (light blue) of OSHIFT compared to the stationary setting (SHIFTfl) is about half of the optimal value (conjunction of light blue and light purple) due to the remaining noise in the estimation of the integrated rate profile.

To illustrate the superiority of the operational time methods, we performed the same analysis using UD with reducing dither widths ±20, ±10, ±6, and ±3 ms. The results are shown in **Figure 7**. We observe that by the time the FP rates are brought down to the level of OJISID (still larger than OSHIFT: upper panel, light blue area) by using a dither width of ±3 ms, the UD method is far less sensitive than OSHIFT or OJISID (lower panel, light blue shaded area). This signifies that one cannot reach the level of performance of the more advanced methods proposed in this study by simply reducing the dither width of simpler methods. The Appendix shows that the loss of sensitivity of UD with decreasing dither width is due to an insufficient destruction of the injected coincidences.

The second parameter which we investigate is the spiking regularity, quantified by the shape parameter of the ISI (gamma) distribution in operational time γ_{op} (see **Figure 8**). More precisely we plot the FP and FN rates as a function of the coefficient of variation $CV = 1/\sqrt{\gamma_{op}}$. As in **Figure 5** we set $\Delta\lambda = 70$ Hz.

For $CV > 1$, we find that most methods, except SHIFT, are operating at reasonable FP rates. However, once the processes become more regular than Poisson, UD shows a strong increase in FP rates. The performance of SHIFT does not seem to depend too strongly on CV and the offset of a few percent above the significance level is due to the smoothing of the rate profile, as can be seen from the SHIFTfl curve. SRD and JISID show a fairly similar behavior, reaching 5% FP rates for highly regular processes, on par with SHIFT. Again, JISID is above the significance level as it shows poor rate conservation properties. Below we find the operational time methods, of which OSHIFT lies at the significance level, unaffected by the increasing regularity.

Turning to the FN rates (**Figure 8**, lower panel), we note that UD and SRD follow a similar trend, opposite to their FP rate trend. SHIFT proves to be fairly sensitive through the parameter range and shows again an upward slope with regularity. In contrast, JISID and OJISID become more sensitive as regularity increases. The reason for this increased FN rate in irregular regimes is that most spikes have small pre- and post-intervals (burst) and thus cannot be dithered by large enough amounts, relative to the coincidence width (±1 ms). The OSHIFT method looses in sensitivity as the process gains in regularity, maintaining its distance with the optimal surrogate for the non-stationary (light purple shading) and stationary (light green shading) cases. This suggests that as in **Figure 6** its performance is limited by the accuracy of the mapping to operational time.

Combining the observations made above, we conclude that OSHIFT is the most conservative method, and in terms of sensitivity, for a fixed accuracy, is the closest to optimum. The OJISID also outperforms simpler methods, however the constraints on the dither range by the previous spike and the next spike limits variability of the coincidence counts in the surrogates and its implementation is far more involved.

APPLICATION TO EXPERIMENTAL DATA

To assess the behavior of the various surrogates in an experimental setting, we consider a pair of neurons recorded non-simultaneously in the primary visual cortex of the anesthetized macaque monkey (Aronov et al., 2003). The reason for choosing non-simultaneous recordings is that we need to be in a situation in which we can be certain that there is no excess synchrony; only then can we be sure that we are observing FP results. The stimuli are transient presentations

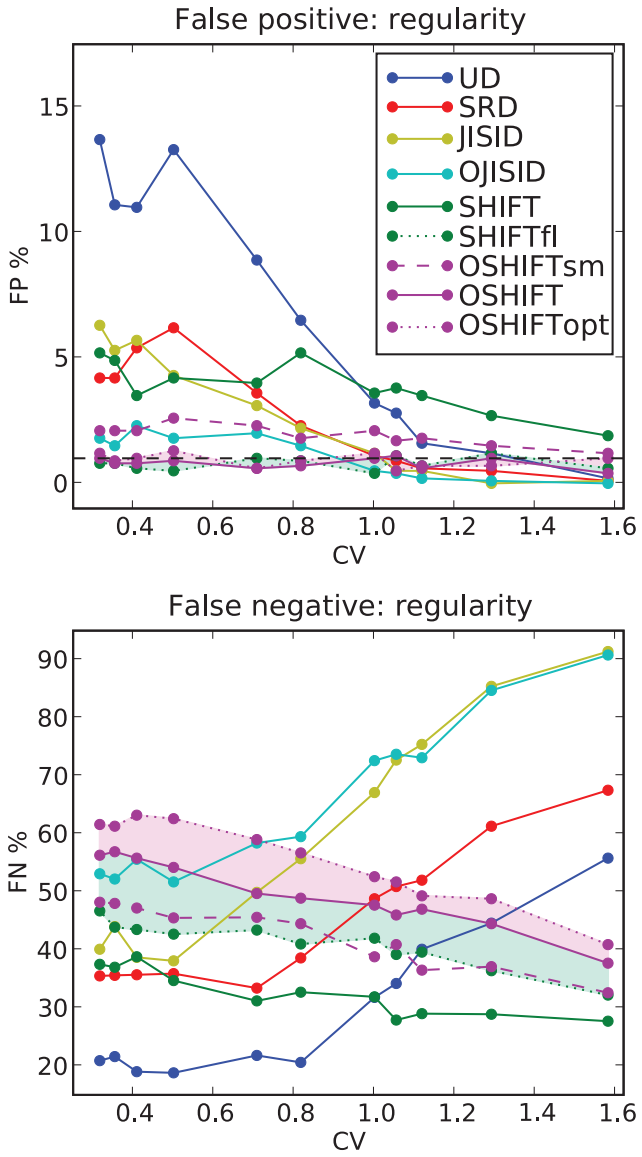


FIGURE 8 | FP and FN percentages as a function of the regularity of the processes. The firing rates of the parallel spike trains follow the step profile with $\Delta\lambda$ = 70 Hz. The regularity of the spike trains is parameterized by the coefficient of variation CV = CV_{op}. FP values are shown for CV values corresponding to shape parameters γ_{op} = 10,8,6,4,2,1.5,1,0.9,0.8,0.6, and 0.4 (from left to right). Each data point results from 1000 repetitions.

of stationary gratings of varying spatial phase. Each neuron was recorded from in different sessions and with a different stimulus presented. A total of 64 trials were obtained for each neuron; the responses are shown in **Figure 9**. As one can see from the dot display and the PSTH, both neurons exhibit strong rate transients within the analysis window of duration 100 ms covering the time interval [250, 350 ms]. The spike rates of the two neurons peak at 200 and 250 Hz, respectively, with highly regular spike sequences.

We treat these two neurons as if they were simultaneously recorded and test for excess coincidences with an allowed jitter of ±1 ms. Due to the large amplitude of the transient we use a dither width of 10 ms for all methods except OSHIFT, which applies a non-uniform dither with a range in real time of at least 10 ms. For each surrogate method we generate 10000 surrogate versions of the recordings from the neuron with the mildest rate transient. The resulting coincidence distributions and *p*-values of the observed coincidence count (black line) are shown in **Figure 10**.

The resulting surrogate distributions are compatible with the observations made in the previous section on synthetic data. The non-ISI conserving methods, UD and SRD, would detect a significant (above 1% level) number of coincidences (*p*-value below 10^{-3} in both cases). SHIFT does not detect significant excess synchrony. In turn, the more conservative methods, JISID, OJISID, and OSHIFT clearly do not observe any excess synchrony, as one would expect from independent recordings.

Thus in such strong rate transient regimes, we recommend the use of more advanced methods which take into account the observed rate and ISI properties of the recordings. The example data exhibit considerable cross-trial non-stationarity. Nevertheless OSHIFT is sufficiently robust and remains the method of choice.

Alternatively, one can reduce the dither width of more basic methods, however this will induce a substantial reduction in sensitivity, for a similar accuracy, as shown in **Figure 7**.

DISCUSSION

The result of our study of the family of surrogate methods based on dithering is that the methods considering the ISI distribution behave best with respect to rate modulations and regularity of the spike trains. The novel techniques of joint-ISI dithering (OJISID) and train dithering (OSHIFT) in operational time are the most robust methods, since they exhibit the lowest FP rates amongst the surrogate methods considered in the paper. The apparently lower FN rate of other methods is a direct consequence of the increased FP rate. At the same FP rate, simpler methods cannot match the sensitivity of OSHIFT and OJISID. Thus this surrogate approach should be restricted to the application to Poisson spike trains with small rate fluctuations. This is also illustrated by the analysis of the macaque V1 recordings considered above. Even though the neurons were recorded in different sessions and are expected not to exhibit excess synchrony, UD does consider the empirical coincidence count as highly significant.

In the Appendix we explore the theoretical relationship between FP and FN by assuming that the true independent and dependent distributions of coincidence counts are known. In this scenario the significance level used for a particular surrogate method can be adjusted to generate a desired FP rate. As a consequence all surrogate methods which effectively destroy coincidences also produce identical FN rates. Methods which leave a fraction of the coincidences intact have a lower sensitivity.

The spike exchange (Smith and Kohn, 2008; Grün, 2009) surrogate methods may seem to have an advantage over the methods covered in this study, as they conserve the PSTH exactly, account for non-stationarities across trials and keep the spike count per trial constant. Thus we expect that they perform better than UD or SRD. However they do not attempt to conserve ISI distribution and as we demonstrated in the FP rate analysis, high firing rates combined with spiking regularity place strict requirements on the surrogate method.

Having considered the limiting cases (UD and SHIFT) of the pattern-jitter method (Harrison and Geman, 2009), we believe that their performances situate the performance of the latter. We

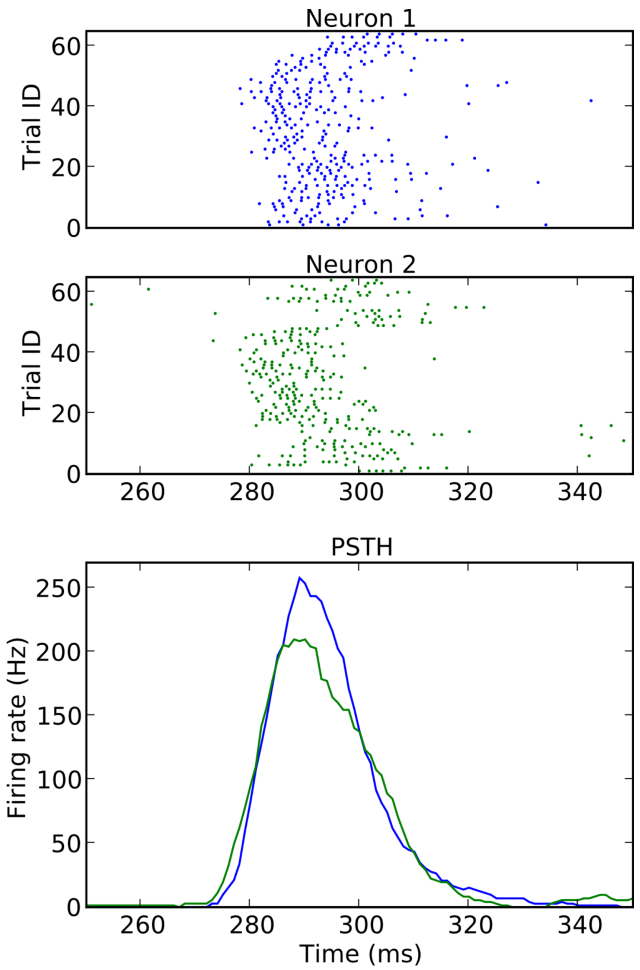


FIGURE 9 | Dot displays (upper panels) and smoothed PSTHs (bottom panel) of two simple cells recorded from monkey primary visual cortex recorded in the same session, however under different stimulus conditions. The stimuli for the two neurons are stationary transient gratings of different orientation. The data are described in (Aronov et al., 2003) and publicly available at neurodatabase.org. Neuron 1 corresponds to file 410106s in condition 5, and neuron 2 to file 410106t in condition 1.

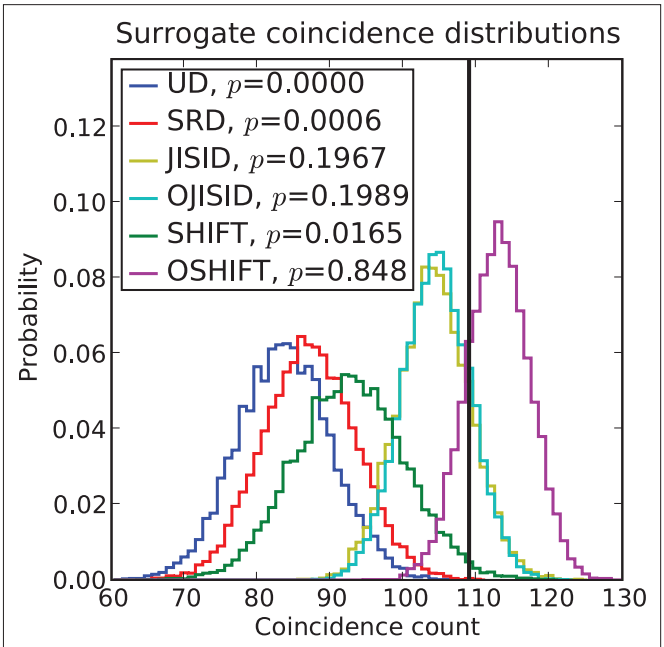


FIGURE 10 | Surrogate coincidence distributions based on experimental data shown in Figure 9, with a dither width set to 10 ms in range-bounded methods. The vertical black line indicates the coincidence count in the original data.

anticipate that pattern-jitter will not represent an improvement over OSHIFT for the processes considered in this study. However further studies including cross-trial non-stationarities are required to clearly differentiate the methods and define the conditions under which they are most applicable.

Experimentally recorded spike trains have the added complication that they not only exhibit rate non-stationarities but also non-stationarities in the CV (regularity) of the ISIs within trials (Shinomoto et al., 2003, 2009; Davies et al., 2006; Nawrot et al., 2008; Kilavik et al., 2009). In consequence single trials may have a rate profile and a potentially independent regularity profile. It remains to be investigated whether a concept similar to operational time for converting a non-stationary rate process to a stationary one, can be found to account for regularity non-stationarities. Again surrogate generation methods need to be thoroughly tested for processes that are non-stationary in both parameters.

In a given application it may be of interest to not only include the rate profiles and the interval statistics but also additional features like a baseline spike correlation into the null-hypothesis. Whether it is possible to limit the destructive power of dithering to achieve this needs to be investigated.

The comparison between the joint-ISI dithering methods in real and operational time highlights that given a non-stationary rate the ISI distribution is in a formal sense not defined. The long intervals in the ISI distribution in real time are dominated by contributions from the low rate regimes. Nevertheless, the full distribution is used to dither spikes also in high rate regions where short intervals dominate. The role of the transformation to operational time is to get access to a well defined ISI distribution valid at any point on the temporal axis. Nawrot et al. (2008) exploited the same idea to reliably assess the CV of ISIs in neuronal data.

Our theoretical framework explains why the square root profile introduced by Gerstein (2004) is superior to a flat dither profile and to the profile following the original distribution. We have no evidence, however, in what sense 0.5 is an optimal choice of the exponent. In fact, it seems that for an arbitrary rate profile an exponent with an improved performance can be found (not shown). This raises the intriguing question whether there is a locally optimized time dependent choice of the exponent $\beta(t)$.

The method of dithering in operational time emphasizes the dual role of the size of the sliding analysis window in time-resolved correlation analysis. The original paper on the Unitary Events analysis (Grün et al., 2002b) states the two characteristics controlled by the parameter: (1) The window needs to be narrow enough to assume stationarity of the rate. (2) The window size needs to be large enough to collect sufficient statistics but adapted

to a potentially temporally constrained occurrence ("hot region") of correlation. With reliable surrogate methods at hand to construct the distribution of coincidence counts for non-stationary rate profiles, the first condition can be relaxed and the experimenter has more freedom to optimize for the dynamics of correlation.

The use of the firing rate profile for a coordinate transformation has also been emphasized in the proposal to use order statistics for spike train resampling (Richmond, 2009). The method allows for the generation of surrogate spike trains with constrained spike numbers and a specified rate profile. Rooted in the theory of order statistics it constitutes a model in itself. However the effects of spiking regularity are so far neglected. It is of interest to find the relation between this method and different versions of dithering, as they have the same aim.

It has been argued that brain processing may be reflected in the higher-order correlation of multiple parallel spike trains (Gerstein et al., 1978; Abeles, 1991). The analysis in this paper, however, is restricted to two parallel processes. Future work needs to investigate whether the idea to work in a transformed temporal coordinate can be extended to higher dimensions. This is relevant because it is conceivable that in higher-order analysis perturbations of assumed surrogate invariants like rate profile and interval statistics may become more problematic.

We have studied a particular injection model where the rate of coincidence is constant while the overall spike rate is changing. This naturally leads to an increase of FN with increasing spike rate. An alternative model would be one where the rate of injected synchrony increases with the spike rate. In this sense our model is a worst case scenario for the detection of synchrony. Which model is best adapted to brain processes remains to be investigated.

While the struggle to construct sensitive and robust surrogates for neuronal spike data continues, we have presented some practical and conceptual advances. On the basis of our study we recommend not to hesitate to exploit the computer technology available today and to use surrogate methods based on operational time to simultaneously conserve the rate profile and the ISI distribution.

ACKNOWLEDGMENTS

The initial part of this work was conducted when S. Grün and M. Diesmann enjoyed a scientific stay with G. Gerstein in Philadelphia in May 2008. We acknowledge fruitful and encouraging discussions with the participants of the EPSRC funded "Spatio-temporal Patterns and Synfire Chains" workshop in August 2008 organized by S. Baker in Newcastle. We also thank J. Ito for useful suggestions. Partially funded by DIP F1.2, Helmholtz Alliance on Systems Biology (Germany), Next-Generation Supercomputer Project of MEXT (Japan), BMBF Grant 01GQ0420 to BCCN Freiburg, and EU Grant 15879 (FACETS). Experimental data used in this study were delivered via neurodatabase.org – a neuroinformatics resource funded by the Human Brain Project.

REFERENCES

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*, 1st Edn. Cambridge: Cambridge University Press.

Abeles, M., and Gat, I. (2001). Detecting precise firing sequences in experimental data. *J. Neurosci. Methods* 107, 141–154.

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

Beck, C., and Schlögl, F. (eds) (1995). "Transfer operator methods (chapter 17)," in *Thermodynamics of Chaotic Systems*. Springer Series in Computational Neuroscience (Cambridge, England: Cambridge University Press), 190–203.

Brown, E., Barbieri, R., Ventura, V., Kass, R., and Frank, L. (2001). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput.* 14, 325–346.

Cox, D., and Isham, V. (1980). *Point Processes*. *Monographs on Applied Probability and Statistics*. Boca Raton: Chapman and Hall.

Date, A., Bienenstock, E., and Geman, S. (1998). *On the Temporal Resolution of Neural Activity*. Technical Report, Division of Applied Mathematics, Brown University.

Davies, R., Gerstein, G., and Baker, S. (2006). Measurement of time-dependent changes in the irregularity of neural spiking. *J. Neurosci.* 96, 906–918.

Denker, M., Wiebelt, B., Fliegner, D., Diesmann, M., and Morrison, A. (2010). "Practically trivial parallel data processing in a neuroscience laboratory," in *Analysis of Parallel Spike Trains*, Springer Series in Computational Neuroscience, S. Grün, and S. Rotter, (New York: Springer), 416–436.

Diesmann, M., Louis, S., Grün, S., and Gerstein, G. (2009). "Spike train surrogates based on dithering in operational time," in *Proceedings of 57th Session of the International Statistical Institute*, Durban, South Africa.

Gerstein, G. L. (2004). Searching for significance in spatio-temporal firing patterns. *Acta Neurobiol. Exp. (Wars)* 64, 203–207 (review).

Gerstein, G., and Perkel, D. (1972). Mutual temporal relationships among neuronal spike trains. Statistical techniques for display and analysis. *Biophys. J.* 12, 453–473.

Gerstein, G., Perkel, D., and Subramanian, K. (1978). Identification of functionally related neural assemblies. *Brain Res.* 140, 43–62.

Gradshteyn, I., and Ryzhik, I. (2000). *Tables of Integrals, Series, and Products*, 6th Edn. San Diego, CA: Academic Press.

Grimaldi, R. P. (2003). *Discrete and Combinatorial Mathematics: An Applied Introduction*, 5th Edn. Reading, Massachusetts: Addison Wesley.

Grün, S. (2009). Data-driven significance estimation of precise spike correlation. *J. Neurophysiol.* 101, 1126–1140 (invited review).

Grün, S., Diesmann, M., and Aertsen, A. (2002a). 'Unitary Events' in multiple single-neuron spiking activity. I. Detection and significance. *Neural Comput.* 14, 43–80.

Grün, S., Diesmann, M., and Aertsen, A. (2002b). 'Unitary Events' in multiple single-neuron spiking activity. II. Non-Stationary data. *Neural Comput.* 14, 81–119.

Grün, S., Riehle, A., and Diesmann, M. (2003). Effect of cross-trial nonstationarity on joint-spike events. *Biol. Cybern.* 88, 335–351.

Harrison, M., and Geman, S. (2009). A rate and history-preserving resampling algorithm for neural spike trains. *Neural Comput.* 21, 1244–1258.

Hatsopoulus, N., Geman, S., Amarasingham, A., and Bienenstock, E. (2003). At what time scale does the nervous system operate? *Neurocomputing* 52–54, 25–29.

Kilavik, B., Roux, S., Ponce-Alvarez, A., Confais, J., Grün, S., and Riehle, A. (2009). Long-term modifications in motor cortical dynamics induced by intensive practice. *J. Neurosci.* 29, 12653–12663.

Langtangen, H. P. (2006). *Python Scripting for Computational Neuroscience*, 2nd Edn. Berlin/Heidelberg/New York: Springer.

Louis, S., Borgelt, C., and Grün, S. (2010). "Selecting appropriate surrogate methods for correlation analysis," in *Analysis of Parallel Spike Trains*. Springer Series in Computational Neuroscience, S. Grün, and S. Rotter, (New York: Springer), 359–382.

Nadasdy, Z., Hirase, H., Czurko, A., Csicsvari, J., and Buzsaki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *J. Neurosci.* 19, 9497–9507.

Nawrot, M. P., Boucsein, C., Rodriguez Molina, V., Riehle, A., Aertsen, A., and Rotter, S. (2008). Measurement of variability dynamics in cortical spike trains. *J. Neurosci. Methods* 169, 374–390.

Pazienti, A., Maldonado, P., Diesmann, M., and Grün, S. (2008). Effectiveness of systematic spike dithering depends on the precision of cortical synchronization. *Brain Res.* 1225, 39–46.

Perkel, D., Gerstein, G., and Moore, G. (1967). Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J.* 7, 419–440.

Pipa, G., Wheeler, D., Singer, W., and Nikolic, D. (2008). Neuroxidence: reliable and efficient analysis of an excess

or deficiency of joint-spike events. *J. Comput. Neurosci.* 25, 64–88.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd Edn. Cambridge: Cambridge University Press.

Richmond, B. (2009). Stochasticity, spikes and decoding: sufficiency and utility of order statistics. *Biol. Cybern.* 100, 447–457.

Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., Fujita, I., Tamura, H., Doi, T., Kawano, K., Inaba, N., Fukushima, K., Kurkin, S., Kurata, K., Taira, M., Tsutsui, K., Komatsu, H., Ogawa, T., Koida, K., Tanji, J., and Toyama, K. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Comput. Biol.* 7, 12591–12603. doi:10.1371/journal.pcbi.1000433.

Shinomoto, S., Shima, K., and Tanji, J. (2003). Differences in spiking patterns among cortical neurons. *Neural Comput.* 15, 2823–2842.

Shmiel, T., Drori, R., Shmiel, O., Ben-Shaul, Y., Nadasdy, Z., Shemesh, M., Teicher, M., and Abeles, M. (2006). Temporally precise cortical firing patterns are associated with distinct action segments. *J. Neurophysiol.* 96, 2645–2652.

Smith, M. A., and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* 28, 12591–12603.

Snyder, D., and Miller, M. (1991). *Random Point Processes in Time and Space.* New York: Springer Verlag.

Stark, E., and Abeles, M. (2009). Unbiased estimation of precise temporal correlations between spike trains. *J. Neurosci. Methods* 179, 90–100.

Ventura, V., Cai, C., and Kass, R. E. (2005). Statistical assessment of time-varying dependency between two neurons. *J. Neurophysiol.* 94, 2940–2947.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 November 2009; paper pending published: 10 December 2009; accepted: 04 August 2010; published online: 22 September 2010.

Citation: Louis S, Gerstein GL, Grün S and Diesmann M (2010) Surrogate spike train generation through dithering in operational time. Front. Comput. Neurosci. **4**:*127. doi: 10.3389/fncom.2010.00127*

APPENDIX

THE RELATIONSHIP OF FP AND FN

In the following we compare three characteristic dithering methods (UD, SHIFT, and OSHIFT) in a situation where the FP generated by the different methods are matched by adjusting the significance level α. This cannot be done for experimental data because, as shown below, the resulting α_m depends on the detailed shape of the surrogate distribution and the calibration requires access to the true distribution of coincidence counts for independent data.

Nevertheless, the analysis of a model situation enables us to investigate the relationship between FP and FN and the limit of sensitivity. Consider a situation similar to the one discussed in **Figure 6**. We generate two types of data sets consisting of 100 trials of 100 ms duration of $\gamma = 3$ process realizations with a rate step at 50 ms from a base of 10 Hz to a new rate level elevated by $\Delta\lambda$. One type is called the correlated or dependent data set. Here we inject coincidences with a jitter of ±1 ms using a Poisson process at rate $\lambda_c = 2$ Hz and reduce the baseline rate accordingly. The second type is left uncorrelated which we call the independent data set. As in **Figure 6** we vary $\Delta\lambda$ from 0 to 100 Hz and create 10^5 data sets for both types. Subsequently we apply the dithering methods UD, SHIFT, and OSHIFT to the data sets to generate one surrogate data set per method and original data set. Finally, for each of the methods we collocate the data into four distributions of coincidence counts: independent data, correlated data, and the corresponding two surrogate distributions. For comparison we also compile the four distributions for UD at a reduced dither width of ±3 ms.

Figure 11 verifies that at a dither width of ±20 ms the surrogate distributions for correlated and independent data are identical for all dithering methods because the dither width is large enough to destroy practically all injected coincidences. Note that in this Appendix we simplify the procedure compared to the main text in that we construct the distributions of coincidence counts by combining data from all original spike train realizations. This is less accurate because for a particular realization the surrogate distributions do not conserve the spike counts of the original data. As argued above and elsewhere (Grün, 2009) we do not recommend to do this in the analysis of experimental data but it is convenient to study the fundamental relationship between FP and FN.

Figure 11 illustrates the shapes of the distributions and their relationships at a particular $\Delta\lambda$. For the given significance level of $\alpha = 0.01$ the fractions of FP and FN differ considerably between the surrogate methods. This is mainly due to the different means of the surrogate distributions. The differences between the surrogate and the independent distribution for UD and SHIFT demonstrate the fact that these surrogates lead to lower mean coincidence counts than in the independent data due to the destruction of the rate profile (cf. **Figure 4**). For OSHIFT the surrogate distribution well resembles the independent distribution. UD exhibits a decreased level of FN simply because compared to OSHIFT the surrogate distribution is shifted to the left. The price is an increase of FP far exceeding α because the surrogate distribution is also shifted to the left with respect to the independent distribution. For UD 3 ms the surrogate distributions are closer to the independent distribution because the rate profile is less distorted as at the larger dither width. However, at the smaller dither width the surrogate distributions for correlated and independent data are no longer identical because the method does not destroy coincidences effectively enough.

In the right column of **Figure 11** we match the fraction of FP of the dithering methods at a target level of 1% by selecting the minimal coincidence count required for significance $n_{\mathrm{FP=1\%}}$ as the largest coincidence count for which the total probability of the independent distribution for counts larger or equal to this value does not exceed 1%. The total probability of the surrogate distribution corresponding to $n_{\mathrm{FP=1\%}}$ constitutes the matching significance level α_m of the dithering method used to generate the surrogate distribution. By definition $n_{\mathrm{FP=1\%}}$ has exactly the same value for all dithering methods because it only depends on the independent distribution, not the surrogate distribution. Therefore, not only the fraction of FP but also the fraction of FN are now identical for all dithering methods. What is different is the matched significance level α_m. Clearly, this calibration procedure is only possible in our model situation because we have access to the true coincidence count distribution of independent data.

An exception to the invariance of the relationship between FP and FN with respect to the dithering methods at matched α levels is UD 3 ms. Here the surrogate method does not destroy all injected coincidences. The surrogate distribution for correlated data is shifted to the right with respect to the surrogate distribution for independent data. Therefore, the α_m determined using the independent data leads to an increased fraction of FN.

The top panel of **Figure 12** shows the dependence of the fraction of FP on the magnitude of the rate step $\Delta\lambda$. With increasing $\Delta\lambda$ the discreteness of the distribution of coincidence counts reduces and therefore the optimal choices of $n_{\mathrm{FP=1\%}}$ better approximate the target FP fraction of 1%. A detailed discussion of the discreteness of the distribution of coincidence counts can be found in Grün (2009). All methods behave the same.

The fraction of FN increases with increasing $\Delta\lambda$ (**Figure 12**, middle panel) because with increasing mean spike rate the fraction of surplus coincidences compared to the number of chance coincides reduces. The dependence of FN on $\Delta\lambda$ only depends on the independent distribution and therefore is identical for all dithering methods and represents the optimal sensitivity (FN rate) for any surrogate method. An exception is UD 3 ms which does not manage to destroy the injected coincidences effectively enough. The result is a substantially reduced sensitivity. The FN converge again at large $\Delta\lambda$ when the injected coincidences contribute little to the large number of chance coincidences.

The bottom panel of **Figure 12** shows the dependence of the significance level α_m on $\Delta\lambda$. For OSHIFT the α_m stays close to the desired fraction of FP = 1% because the surrogate distribution well approximates the distribution of coincidence counts for independent data. Thus, using OSHIFT the experimenter can select the α of choice and obtain the expected FP level. Also UD 3 ms is well behaved in this respect. For UD at ±20 ms, however, α_m drops with increasing $\Delta\lambda$ by at least two orders of magnitude. This indicates that $n_{\mathrm{FP=1\%}}$ is located far out in the tail of the surrogate distribution. Consequently the precise value of α_m depends on the details of the time course of the original data. There is no universal mapping of a desired significance level α to α_m for UD.

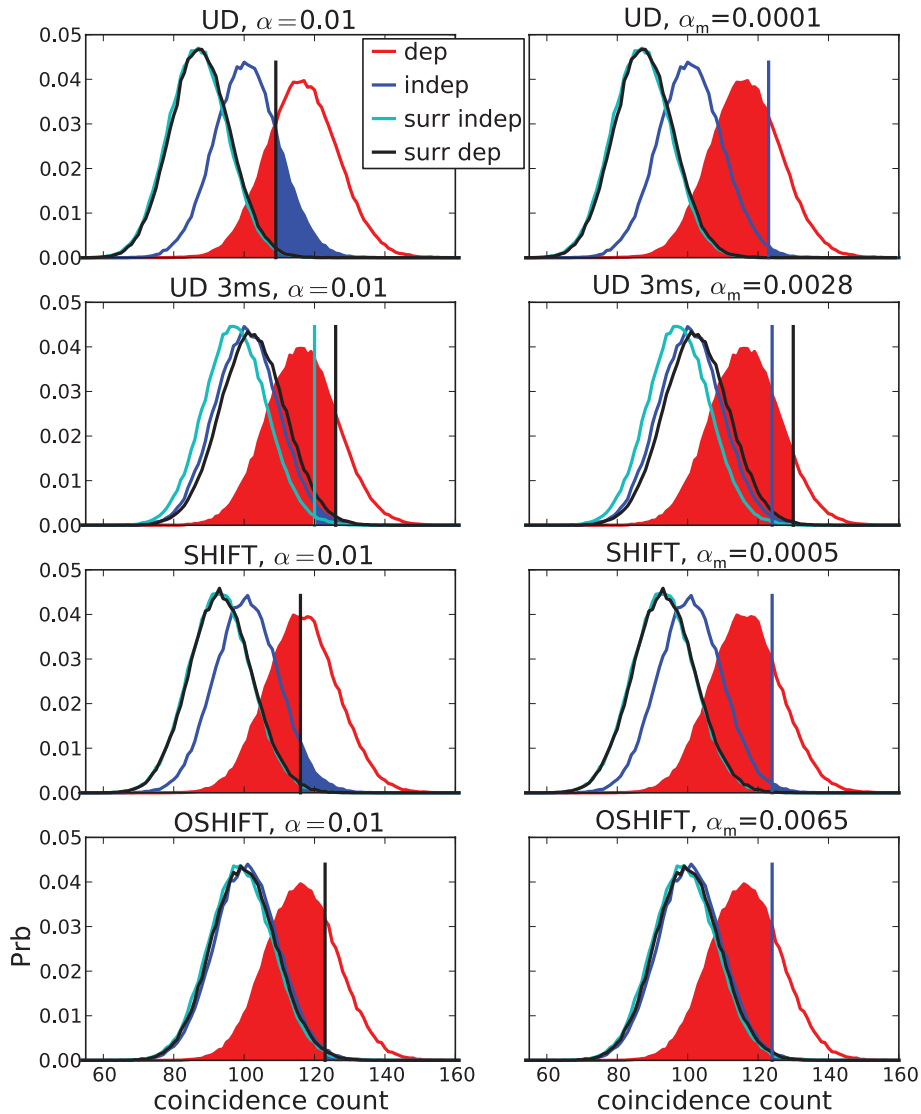


FIGURE 11 | False positives (FP, blue areas) and false negatives (FN, red areas) for unmatched (left) and matched (right) α. The curves show coincidence count distributions of correlated (red), independent (blue) and the respective surrogate (black, cyan) data for four dithering methods (top to bottom: UD, UD 3 ms, SHIFT, OSHIFT). The probabilities at the discrete coincidence counts are connected by straight lines and the sums of neighboring probabilities are indicated by colored areas for clarity. For UD with dither width reduced to 3 ms (second row) from 20 ms the surrogate distribution for correlated data (black) is substantially shifted to the right with respect to the one for independent data (cyan). The left column shows results for a fixed significance level of α = 0.01. The coincidence count n_α (vertical bar) is the largest count with respect to the surrogate distributions for which the sum of the probabilities of increasing counts starting at this value is smaller or equal α. This defines the fraction of FP as the area under the independent distribution (blue) for counts $\geq n_\alpha$ and the fraction of FN as the area under the correlated distribution (red) for counts $\leq n_\alpha$. For the four methods the n_α are located at different counts and for UD 3 ms also the n_α of the two surrogate distributions differ (visible cyan vertical bar). The right column shows results for FP levels of 0.01 achieved by choosing a corresponding significance level α_m (values in panel titles). The count $n_{FP=1\%}$ (blue vertical bar) is the largest count with respect to the independent distribution (blue curve) for which the sum of the probabilities of increasing counts starting at this value is smaller or equal FP = 1% (blue area). This α_m applied to the surrogate distribution of correlated data defines the threshold count for the fraction of FN (red area). For UD 3 ms the threshold (visible black vertical bar) is to the right of $n_{FP=1\%}$. $\Delta\lambda$ = 90 Hz, other parameters as in **Figure 6**.

In conclusion, we now understand why the performance of OSHIFT cannot be achieved by reducing the dither width of UD as studied in **Figure 7**. Reducing the dither width reduces the fraction of FP and increases the fraction of FN because the surrogate distribution better resembles the independent distribution. Eventually, however, the dither width is so low that a substantial fraction of injected coincidences remains intact and the fraction of FN surpasses the one for OSHIFT at a larger dither width. A good surrogate method is characterized by the congruence of three distributions: the distribution of coincidence counts of independent data, the surrogate distribution of independent data, and the surrogate distribution of correlated data.

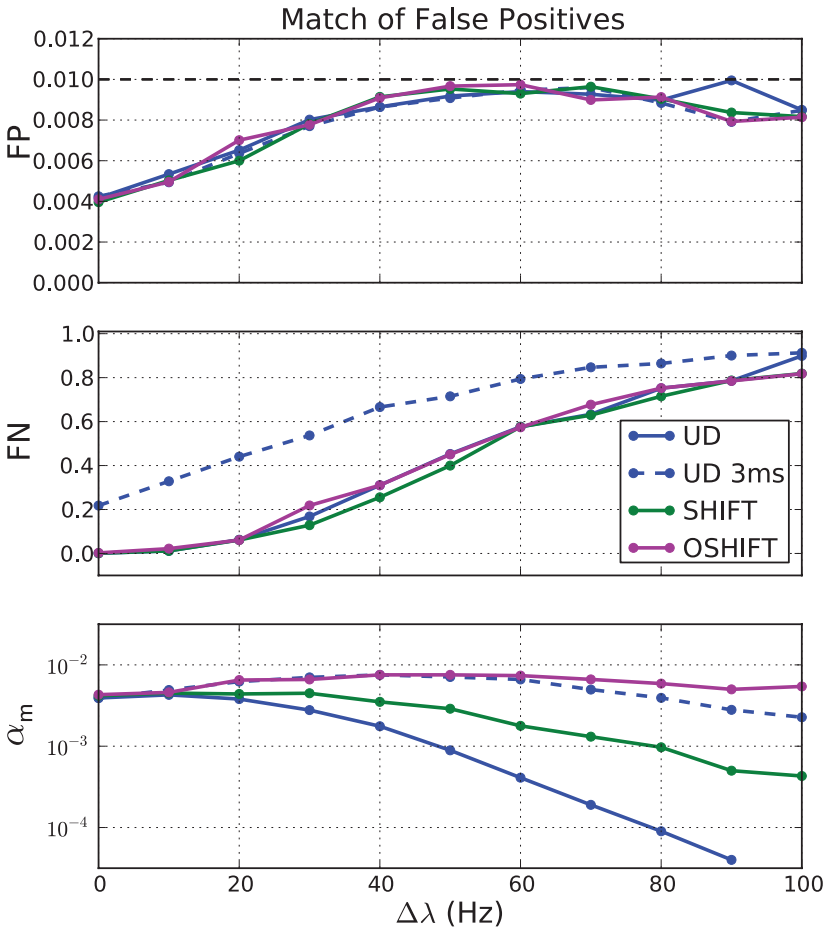


FIGURE 12 | False negatives (FN, middle panel) at matched rates of false positives (FP, top panel) for four surrogate methods (UD, UD 3 ms, SHIFT, OSHIFT). The top panel shows the optimal approximations to the target fraction of FP = 1% given the discreteness of the coincidence count distributions at the magnitude of the rate step $\Delta\lambda$. The middle panel shows the corresponding fractions of FN. The bottom panel shows the significance levels α_m of the four surrogate distributions realizing the matched FP rate of 1% on a log-scaled axis. Other parameters as in **Figure 11.**

It appears tempting to calibrate α_m on the surrogate distribution for correlated data instead of independent data to compensate for the incomplete destruction of coincidences at small dither widths. This, however, is a conceptual error in the context of our null-hypothesis because α_m then depends on the amount of synchrony originally contained in the data.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 08 June 2011
doi: 10.3389/fncom.2011.00023

Higher order spike synchrony in prefrontal cortex during visual memory

Gordon Pipa[1,2,3] and Matthias H. J. Munk[4]*

[1] *Department of Neurophysiology, Max-Planck-Institute for Brain Research, Frankfurt/Main, Germany*
[2] *Institute of Cognitive Science, University of Osnabrueck, Osnabrueck, Germany*
[3] *Frankfurt Institute for Advanced Studies, Frankfurt, Germany*
[4] *Department of Physiology of cognitive Processes, Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany*

Edited by:
Israel Nelken, Hebrew University, Israel

Reviewed by:
Elad Schneidman, Weizmann Institute of Science, Israel
Moshe Abeles, Bar-Ilan University, Israel

***Correspondence:**
Matthias H. J. Munk, Department Physiology of Cognitive Processes, Max-Planck-Institute for Biological Cybernetics, Tübingen 72076, Germany.
e-mail: matthias.munk@tuebingen.mpg.de

Precise temporal synchrony of spike firing has been postulated as an important neuronal mechanism for signal integration and the induction of plasticity in neocortex. As prefrontal cortex plays an important role in organizing memory and executive functions, the convergence of multiple visual pathways onto PFC predicts that neurons should preferentially synchronize their spiking when stimulus information is processed. Furthermore, synchronous spike firing should intensify if memory processes require the induction of neuronal plasticity, even if this is only for short-term. Here we show with multiple simultaneously recorded units in ventral prefrontal cortex that neurons participate in 3 ms precise synchronous discharges distributed across multiple sites separated by at least 500 μm. The frequency of synchronous firing is modulated by behavioral performance and is specific for the memorized visual stimuli. In particular, during the memory period in which activity is not stimulus driven, larger groups of up to seven sites exhibit performance dependent modulation of their spike synchronization.

Keywords: visual short-term memory, primate prefrontal cortex, spike synchrony, multi-unit activity, behavioral performance, stimulus coding, joint-spike events, joint-spike patterns

INTRODUCTION

Synchrony of neuronal spike firing has originally been proposed as a fundamental property of neocortical function (Delage, 1919; Hebb, 1949; Abeles, 1982, 1991) and has been observed under various conditions in numerous areas of the cerebral cortex. Early evidence was provided by studies of primary visual cortex (Gray et al., 1989; reviewed in Singer and Gray, 1995), later synchrony was observed in extrastriate (Kreiter and Singer, 1996) and other sensory areas like A1 (Ahissar et al., 1992; deCharms and Merzenich, 1996) and executive areas including frontal cortex (Vaadia et al., 1995), primary motor cortex (Murthy and Fetz, 1996; Riehle et al., 1997; Pipa et al., 2007). However, the nature of synchronous firing has nurtured a long standing debate whether synchrony serves the integration of signals distributed over large neuronal populations (Singer, 1999 versus Shadlen and Movshon, 1999). One interesting problem in this discussion was that studies in which attention was explicitly or implicitly modulated, synchrony could either change as predicted by properties of the sensory stimuli (Kreiter and Singer, 1996; Maldonado et al., 2000; Steinmetz et al., 2000) or in a counterintuitive way, which was not related to properties of the stimuli (de Oliveira et al., 1997; Thiele and Stoner, 2003). Two recent studies (Dong et al., 2008; Lima et al., 2010) have once more investigated whether the "binding-by-synchronization" hypothesis can predict spike synchrony in area V1 of behaving macaques. Both studies found synchrony which did show some degree of stimulus dependence, but reflected more spatial properties of the underlying connectivity as had been shown before for correlated firing of neurons in V1 and V2 (Nowak et al., 1999; Kohn and Smith, 2005) rather than direct evidence for figural binding. In higher cortical areas like inferotemporal cortex, neurons can synchronize their spiking when monkeys successfully solve visual recognition tasks (Gochin et al., 1994; Anderson et al., 2006) or processes features of faces (Hirabayashi and Miyashita, 2005), but evidence for stimulus dependent or even object-specific synchronized firing in higher visual areas remains sparse. Despite of this unresolved issues, cortical areas beyond sensory pathways express temporally precise spike firing which has been related to prediction of go signals (Riehle et al., 1997), decision making (Dudkin et al., 1995; Thiele and Hoffmann, 2008), spatial (Constantinidis and Goldman-Rakic, 2002), as well as working memory for temporal intervals and color (Sakurai and Takahashi, 2006).

Does millisecond precise neuronal firing have any relevance for cortical information processing? Evidence for the behavioral relevance of precise neuronal timing beyond mere covariation with behavior was recently provided by electrical stimulation experiments in auditory cortex which showed that rats can detect interstimulus-intervals of 3 ms (Yang et al., 2008). However, there is growing evidence that precise neuronal activity patterns across different spatiotemporal scales are highly relevant for information coding in sensory and associational areas of the cortex (Kayser et al., 2009). Another piece of evidence that points to the relevance of precise neuronal timing is the observation that during attention, the variance of spike responses is reduced (Mitchell et al., 2007), which may be related to the occurrence of stabilizing gamma oscillations (Rodriguez et al., 2010).

However, there are other observations of cortical synchrony which suggest that precise spike timing is a much more general principle of cortical function than serving the encoding of

behaviorally relevant information provided by sensory input. One of the prominent properties of corticocortical networks is their massive divergence and convergence (Salin and Bullier, 1995) and the very low number of synaptic contacts between individual cells (Douglas and Martin, 2004) combined with small unitary synaptic potentials (Sjöström et al., 2008). As a consequence, signal propagation along cortical pathways depends on cooperativity of a large number of converging presynaptic neurons (Sjöström et al., 2008). But, beyond feed-forward processing of sensory information, cortical networks are continuously active (Arieli et al., 1996; De Luca et al., 2006), which may be the consequence of reverberating synfire chains (Abeles et al., 1993; Prut et al., 1998) and is most likely the basis for ongoing brain processes like thinking and dreaming. Regulating the general fluidity of neuronal interactions on large spatial scales are likely to reflect general capabilities of the cortical network which can be addressed more empirically as general factor of intelligence (van den Heuvel et al., 2009). Beyond these putative cognitive functions of precise neuronal timing, synchronous cortical activity is involved in the organization of cortical circuits as abundant evidence for spike timing dependent plasticity suggests (Caporale and Dan, 2008).

Why could synchronous spiking be useful for in the organization of short-term memory in prefrontal cortex? (1) Synchrony might sustain endogenous activity during the memory delay for maintaining stimulus information without depending on further sensory drive, (2) Synchrony may support sensory coding of feature conjunctions as hypothesized in the binding hypothesis (see however Dong et al., 2008), (3) Synchronous activity could drive downstream neurons in premotor cortex to prepare and execute the behavioral responses, (4) Synchrony may reconnect more abstract representations to sensory representations during rehearsal as has been shown for locking of theta oscillations across areas with dual micro electrode recordings in ventral PFC and V4 (Liebe et al., 2009; Hoerzer et al., 2010), (5) Synchrony might structure executive processes underlying task performance by driving circuits that serve different subtasks in the memory process. We set out to determine whether we can find synchronous spiking in our multi-site prefrontal recordings and if confirmed, to test whether this synchrony is task and/or stimulus dependent.

MATERIALS AND METHODS

We therefore trained two female monkeys (*M. mulatta*) to perform a visual short-term memory task which consisted of a 0.5-s sample presentation, followed by a 3-s delay and a 2-s test presentation (**Figure 1**). Sample stimuli were randomly drawn from a set of 20 familiar stimuli and test stimuli were drawn from the same set excluding the sample of this trial in half of the trials in which non-matching test stimuli were used. Match and Non-match trials were presented in random order. When the test stimulus was shown, the monkey had to decide whether the stimulus was matching the sample and respond by pressing the left of two buttons while in case of a non-match, the monkey had to press the right button. By requiring behavioral responses for both types of test stimuli we made sure that all trials are homogenous with respect to response preparation and motor activity. Stimulus presentation and behavioral control were provided by a custom-made program. The monkeys did not have to fixate, but we measured eye movements at high resolution with the double magnetic induction method (Bour et al., 1984). The percentage of correct behavioral responses ranged between 71 and 87% across sessions. Anatomical MRI scans (T1-flash, 1 mm^3 isovoxel, 1.5 T) were used to guide implantation of recording chambers and to reconstruct recording positions. All procedures were approved by the local authorities (Regierungspräsidium) and are in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals.

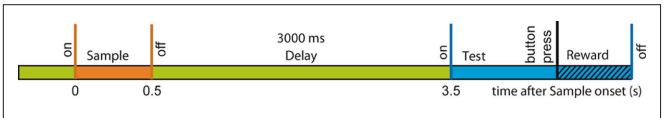


FIGURE 1 | Time structure of the visual memory task.

Simultaneous recording of multi-unit activity was performed with up to 16 platinum–tungsten-in-quartz fiber microelectrodes (Thomas RECORDING, Giessen, Germany) from ventral PFC. Electrodes had been arranged in a square shaped 4×4 grid with a distance between nearest neighbors of 500 µm. Signals were filtered (0.5–5 kHz, 3 dB/octave), digitized at 32 kHz, and saved as time stamp with attached waveform. Preprocessing included the rejection of artifacts (movements, licking) and removing line noise at 50 ± 0.5 Hz. Spike pattern analyses were performed for sets of trials constructed from the stimulus and behavioral protocol using the NeuronMeter software package (http://neuronmeter.convis.info). Data will become available online at the German Neuroinformatics Node (http://www.neuroinf.de/).

ANALYSIS OF SYNCHRONOUS FIRING

To identify differences in neuronal coupling expressed by modulation of spike synchrony we used a bivariate and multivariate extension of NeuroXidence (Pipa et al., 2008; Wu et al., submitted; see also http://www.NeuroXidence.com). In the present article, each incidence of a synchronized firing event is referred to as a joint-spike event (JSE), while the identity of a JSE is referred to as a joint-spike pattern (JS-pattern). Or, with other words JSE are realizations of a JS-pattern.

In a first step, the frequency $f_t^{k,p}(\mathrm{org})$ of JSEs of a certain JS-pattern (p) was determined by the bivariate and multivariate extension of NeuroXidence for each trial (t) and for each factor (k) of an experiment. To account for the stochasticity of spike times, a JS-pattern is defined by a millisecond wide temporal window, which accounts for the maximal uncertainty of synchronous firing (**Figure 2A**). In this paper, this uncertainty (t_c) was set to 3 ms. Note that the detection of a JSE is not based on binned spike trains, but uses the exact experimental spike times which were sampled at a precision of 32 kHz, i.e., times of threshold crossing were initially recorded as multiples of 31.25 µs. For illustrative purposes, **Figures 2B,C** demonstrate how significant JS-patterns (**Figure 2B**) are destroyed when spike trains are randomly jittered by $t_r = 15$ ms (**Figure 2C**). In the original data, the total number of significant joint-spike patterns consisted of patterns with complexities 2–6, 59% of JSP with complexity 2, 13% complexity 3, 17% complexity 4, 10% complexity 5 and 1% complexity 6. After jitter the number of significant patterns was reduced for all complexities. Compared to the frequency of significant joint-spike pattern in the original data, the percentage in respect to the complexity dropped from 59 to 54% (c2), from 13 to 3% (c3), from 17 to 4%, from 10 and 1 to

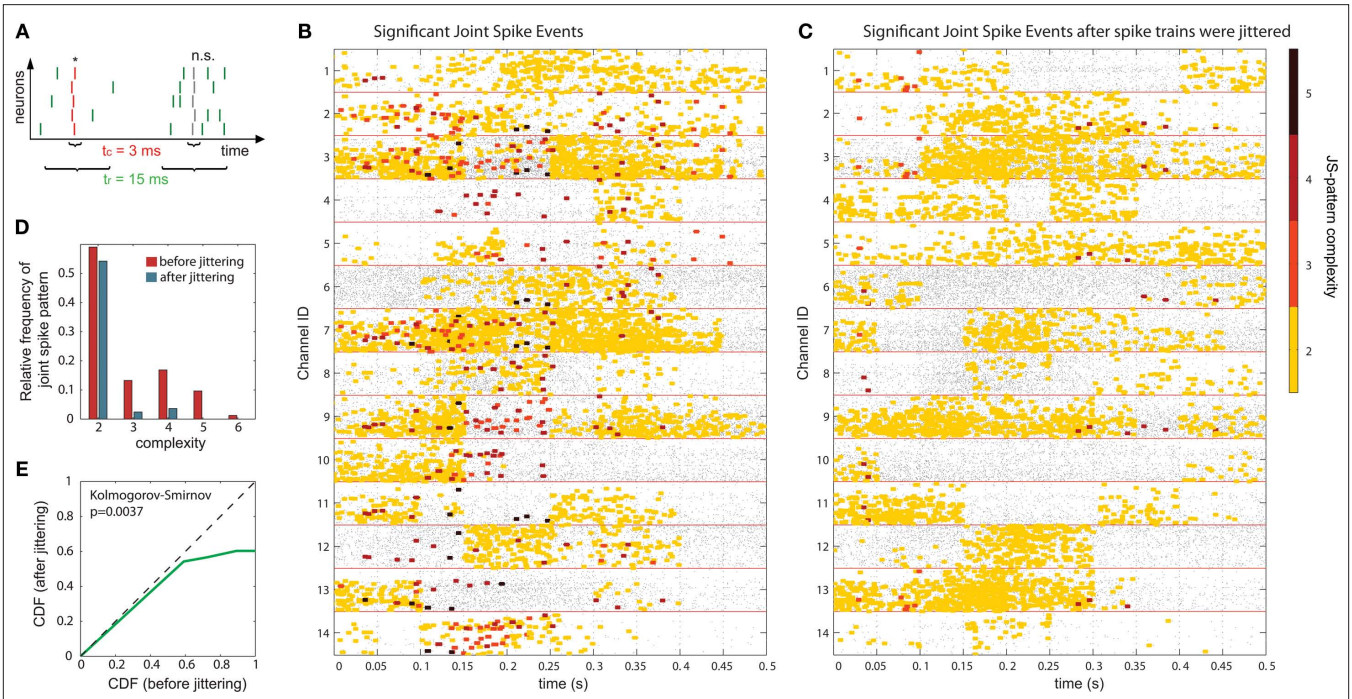


FIGURE 2 | Demonstration of time windows, coincidences before and after significance estimation and after introducing artificial jitter. (A) Sketch to illustrate the idea of testing joint-spike events (JSE) by comparing time scales. Joint-spike patterns (JSP) are considered significant if there are more or less JSE than in a dataset in which spike timing of neurons is synchronized and spike rate covariations are slower than t_r = 15 ms. To define this criterion, two time scales were introduced. The first time scale (t_c) defines the temporal precision with which spikes have to coincide in order to be detected as JSE. The second time scale (t_r) defines the maximum speed of rate covariation. Using these two time scales, JSEs are considered as significant if there are significantly more JSEs, i.e., coincidence firing with an uncertainty less than t_c, in the original data than in a surrogate data set obtained by jittering the original spike trains with ±7.5 ms. **(B)** Raster plot of all significant synchronous patterns in 14 sites after correction with t_r. **(C)** Raster plot of the same data after destroying JSPs: In order to demonstrate that the JSPs shown in **(B)** are real, we performed a control for which both – original and surrogate – data were jittered with the same t_r = 15 ms. This control corresponds to a data set that does not have any fine temporal spike synchronization across neurons, but may have spike rate covariation slower than t_r. **(D)** Quantitative comparison of the number of significant JSPs in original and control data set, shown in **(B,C)**, respectively. Red bars are normalized to 1, and the blue bars are normalized to the total numbers of non-jittered patterns (red = number of patterns of complexity c in non-jittered data / number of patterns for all complexities in non-jittered data; blue = number of patterns of complexity c in jittered data / number of patterns for all complexities in non-jittered data). **(E)** Cumulative Distribution Function of a two sample KS-test. The *p*-value of 0.0037 reflects that the number of JSPs is highly significantly decreased.

0% for complexity 5 and 6. This effect is summarized in **Figure 2D** as frequency distribution of JS-patterns as a function of their complexity and a Cumulative Distribution Function based on a two sample KS-test is plotted in **Figure 2E**.

In a second step, the frequencies $f_t^{k,p}(\text{sur})$ of chance JSEs were estimated for JS-pattern (p), trial (t), and experimental factor (k) from surrogate data which were derived from the original data by jittering each individual spike train under the assumption that neuronal spike discharge is not coupled on a fine temporal scale. We generated exactly one surrogate trial for each original trial to prevent a sampling bias. For setting the amount of jitter applied to the original data when generating the surrogates, a second slower time scale t_r was defined which was set to t_r = 15 ms. The slower time scale t_r sets the minimal interval during which rate covariation may explain coincident firing. Therefore, t_r defines the maximal extent of the jittering, which is applied to destroy any fine temporal cross-structure that may exist between different spike trains. Because spike trains are shifted as a whole against each other within t_r, the auto-structure, rate covariations across neurons as well as rate variation and other features of each individual spike train, which are slower than the time scale t_r, are preserved.

In a third step, we determined for each trial (t), each experimental factor (k), and each JS-pattern (p), the difference $\Delta f_t^{k,p} = f_t^{k,p}(\text{org}) - f_t^{k,p}(\text{sur})$ of JSE frequencies in original and surrogate data sets. Ultimately, this difference $\Delta f_t^{k,p}$ is used to test whether the strength of synchrony of a certain JS-pattern differs across experimental conditions. To this end, the bi- and multivariate versions of NeuroXidence test whether the mean or median of the delta frequencies $\Delta f_t^{k,p}$ of JSEs is significantly different across experimental factors. For the bivariate case, mean and median were compared by unpaired *t*-test and *Mann–Whitney U*-test, respectively. For the multivariate case, an ANOVA or a Kruskal–Wallis test were used. This comparison yields exactly one *p*-value per JS-pattern tested across experimental factors and trials. To prevent any sampling bias, only those JS-patterns were tested for which JSEs occurred at least once for every individual factor. For each experiment between hundreds and many thousands of JSE patterns were tested.

In order to summarize the results across all tested JS-patterns detected in each experiment, we grouped JS-patterns based on their complexity (c), which is given by the number of sites participating in a synchronous event. JS-complexity ranges from c = 2 (pairs), in

which at least two sites have fired in synchrony during a temporal window of $t_c = 3$ ms. If $c = 3$, at least three sites fired synchronously, and so on. We analyzed JS-patterns with a complexity of up to $c = 8$. To summarize results for each complexity we derived the frequency σ of JS-patterns that each expressed a significant difference in $\Delta f_t^{k,p}$ across the factors k.

In order to account for dynamic modulations of spike coupling throughout the different periods of each trial, we performed the joint-spike analysis outlined above by using sliding windows. The sliding window length was chosen to fit the assumed time scale of changes of neuronal coupling given the underlying processes that encode, maintain, or decode information. Given the sometimes very transient rate responses we used a sliding window of 100 ms length during the sample and test stimulus presentation periods. The delay period could be analyzed with longer windows of 400 ms because of much slower rate modulations. Note that this choice of the sliding window length is independent of the spike rate modulation *per se*. NeuroXidence allows for an unconstrained choice of sliding window length, because it accounts for auto-structure and rate covariation slower than t_r. This distinguishes NeuroXidence from other methods like for example the unitary event method (Grün et al., 1999, 2002, 2003) which all require stationary data.

BEHAVIORAL AND STIMULUS SPECIFIC MODULATION OF SPIKE SYNCHRONIZATION

First we used the bivariate NeuroXidence method to detect modulation of spike synchronization depending on the behavioral success of the monkeys, comparing trials with correct or incorrect behavioral responses. On average, performance was ~80%, trials with correct responses were four times more frequent than trials with behavioral errors. To prevent any bias, we balanced the number of correct and incorrect trials for each session by selecting subsets of correct trials which were close in time to the error trials. With the bivariate version of NeuroXidence we tested whether synchrony was modulated by the performance of the monkey and derived the direction of modulation, i.e., tested whether synchronous firing compared to chance occurred more often in correct trials than in incorrect (relative increase for correct), or whether synchronous firing occurred more often in incorrect trials than in correct (relative increase for incorrect). In a second step we derived the frequency ρ of JS-patterns of complexity c that expressed a significant increase of spike synchrony for correct responses $\rho(t)_c^{\text{correct}}$, and the frequency of JS-patterns of complexity c that expressed a significant increase of spike synchrony for incorrect responses $\rho(t)_c^{\text{false}}$.

The multivariate version of NeuroXidence was used to detect stimulus specific modulations of spike synchronization. Here, the k-experimental factors were all the 20 different visual stimuli presented during the sample period. These were tested for significant differences $\Delta f_t^{k,p}$ across stimuli. A significant difference indicates that the strength of spike synchrony is modulated in a stimulus specific manner. As for changes related to behavioral performance, we next summarized results by computing the frequency of JS-patterns for each complexity c that expressed a stimulus specific modulation of spike synchrony $\rho(t)_c^{\text{specific}}$.

TEMPORAL MODULATION OF JS-PATTERN COMPLEXITY

The frequencies $\rho(t)_c^{\text{specific}}$ for stimulus specific as well as $\rho(t)_c^{\text{false}}$ and $\rho(t)_c^{\text{correct}}$ for performance related modulations of spike synchrony had been computed for each JS-pattern complexity and each sliding window. Note that each JS-pattern that contributes to any of the three frequencies $\rho(t)_c$ can be considered as significant. To test whether the frequency of JS-patterns also has been significantly modulated over time, we compared $\rho(t)_c^{\text{specific}}$ and the difference $\Delta\rho_c^{\text{perf}} = \rho(t)_c^{\text{correct}} - \rho(t)_c^{\text{false}}$ to analogous results obtained from analyses of the same data, but based on permuted trials. Trial permutation exchanged trials between experimental factors while the simultaneity and the auto-structure of all recorded spike trains was preserved. This way we could destroy any performance or stimulus specific modulations, while keeping all other properties of the joint-spike trains intact so that the analysis of spike synchrony and temporal modulation of neuronal coupling is not compromised (**Figure 3A**). Therefore trial permutation serves as an ideal estimate of the frequency $\rho(t)_c$ and its variability under the null hypothesis that synchrony is unchanged between experimental factors. Trial permutation was performed independently for each sliding window, giving exactly one p-value per JS-pattern for the original trial structure and for the trial permuted data (**Figure 3B**). As for the original data, we then computed the frequencies $\Delta\rho_{c,\text{perm}}^{\text{perf}} = \rho(t)_{c,\text{perm}}^{\text{correct}} - \overline{\rho}(t)_{c,\text{perm}}^{\text{false}}$, and $\rho(t)_{c,\text{perm}}^{\text{specific}}$. Using the average $\Delta\overline{\rho}_{c,\text{perm}}^{\text{perf}}$, $\Delta\overline{\rho}_{c,\text{perm}}^{\text{specific}}$ and the SD $\text{std}(\Delta\overline{\rho}_{c,\text{perm}}^{\text{perf}})$, $\text{std}(\Delta\overline{\rho}_{c,\text{perm}}^{\text{specific}})$ of both frequencies over time for the same complexity, we expressed the modulations of JS-pattern frequency as time course of the trial for each complexity as a z-scores: $z(t)_c^{\text{perf}} = (\Delta\rho(t)_c^{\text{perf}} - \Delta\overline{\rho}(t)_{c,\text{perm}}^{\text{perf}})/\text{std}(\Delta\rho_{c,\text{perm}}^{\text{perf}})$ for behavioral performance (**Figure 3C**), and $z(t)_c^{\text{specific}} = (\Delta\rho(t)_c^{\text{specific}} - \Delta\overline{\rho}(t)_{c,\text{perm}}^{\text{specific}})/\text{std}(\Delta\rho_{c,\text{perm}}^{\text{specific}})$ for stimulus specificity (not shown). In a last step, we compared the modulation of z-scores for performance and stimulus specific modulations of spike synchrony based on a critical z-score accounting for multiple comparisons of all sliding windows and all complexities. Note that the distribution of $\Delta\rho_{c,\text{perm}}^{\text{perf}}$ is not expected to be normal given that $\Delta\rho_{c,\text{perm}}^{\text{perf}}$ is a difference of counts which are rather low. Therefore using the z-score may not be appropriate. We validate the modulation of z-scores for performance and stimulus specific modulations based on a rank order statistic. To this end we performed the permutation analysis three times, yielding in total 1368 estimates of $\Delta\rho_{\text{perm}}^{\text{perf}}$ (three times 97 estimates across time and 8 pattern complexities). We then determined the largest absolute value $\Delta\rho_{\text{crit}}$ out of all 1368 estimates and used this as the critical value for a minimal significant difference from zero (corresponding test level is $p < 0.001$). This latter rank test is independent of the underlying distribution of chance deviations from zero. Using both methods we found mostly the same time complexity pattern to be significant.

MEDIAN VERSUS MEAN

The entire analysis was performed for both, mean pattern frequency, based on t-test and ANOVA, as well as for median pattern frequency based on Mann–Whitney U-test and Kruskal–Wallis test. ANOVA and Kruskal–Wallis test were used for multivariate analyses. For both tests we obtained qualitatively and quantitatively very similar results. In particular, comparison of z-scores yielded the same significant modulation across time and for the same complexities.

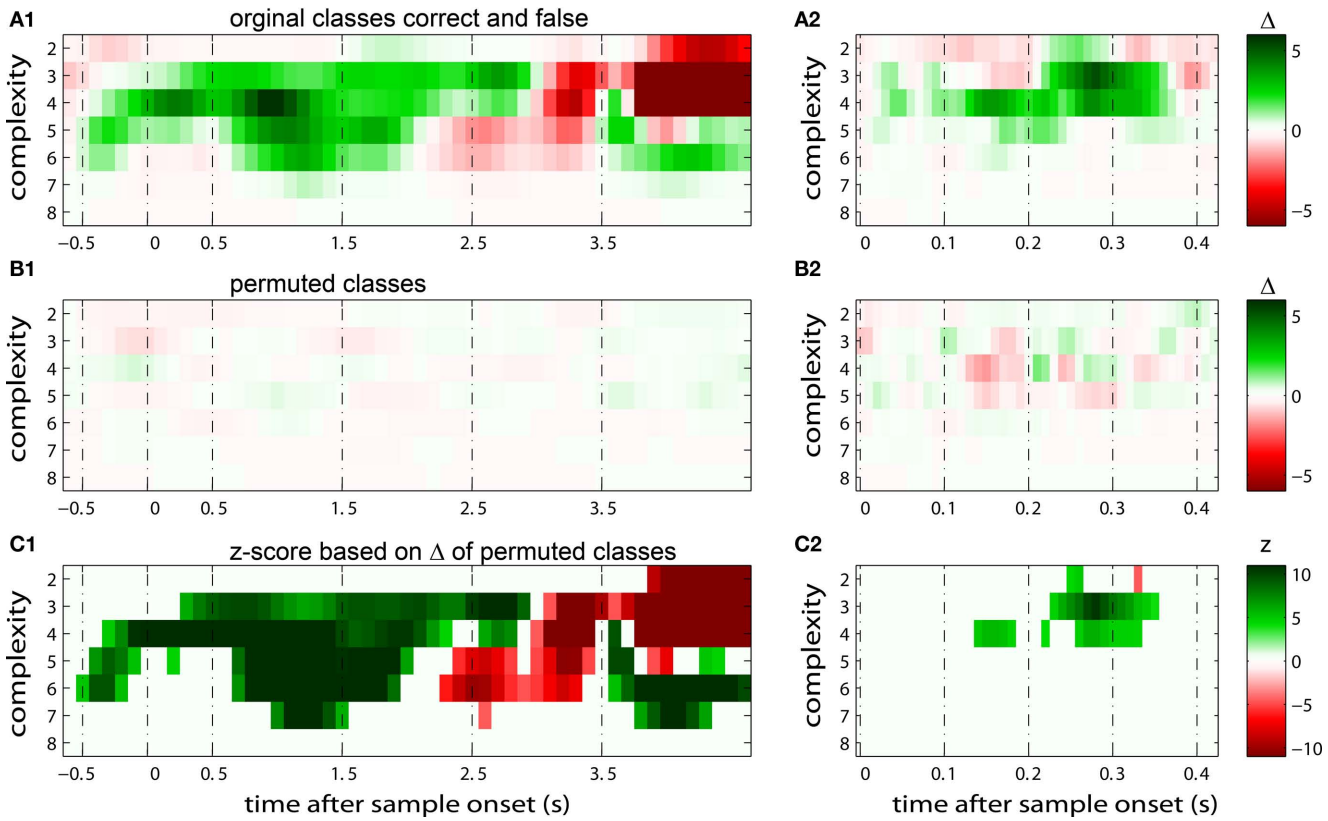


FIGURE 3 | Effects of permuting trials on pattern incidence and as basis of a significance estimate. The *x*-axis represents time after sample stimulus onset in seconds. The *y*-axis shows pattern complexity and the *z*-axis (color) is explained for each figure below. Sliding windows span 400 ms in **(A1–C1)**, while they are 100 ms long in **(A2–C2)**. **(A)** Shows the difference between the frequencies of JS-patterns with increases of synchrony for trials with correct and false behavioral responses. **(B)** Shows the results for the identical dataset, but with randomly permuted trials across the classes "correct" and "false." Based on the average level of differences in **(B)** and the SD of these differences, differences in **(A)** were expressed as *z*-scores in **(C)**. We subtracted the average difference per complexity, but evaluated significance across the entire time course of the trial, starting at –0.5 and ending at 4.3 s after sample stimulus onset. **(C)** We then divided this difference, derived for each sliding window and complexity, by the SD of the respective measurement in the case that classes are permuted [shown in **(B)**].

However, the results based on evaluation of means revealed slightly higher values. Therefore we present the more conservative results based on median testing in this paper.

NUMBER OF SURROGATES

In the presented approach we derived exactly one surrogate trial from each original trial by shifting all spike trains individually by a random time smaller t_r. A small number of surrogates prevents a sampling bias, because if original and surrogate data have exactly the same number of samples, they also have the same degrees of freedom. Increasing the number of samples of the surrogates could be achieved by more than one jitter configuration of the same original trial. This, however, would have two effects. First, the number of different patterns would be larger in the surrogate data, since the probability for individual patterns to occur – at least once – scales with the number of samples. The second effect is that computing the difference $\Delta f_t^{k,p}$ of spike pattern frequencies is non-trivial: in order to compute this difference one can use the average $\overline{\Delta f}_t^{k,p}(\mathrm{sur})$ computed across surrogates for the same trial. This again gives as many surrogate samples as original frequencies such that the same paired test $\overline{\Delta f}_t^{k,p} = f_t^{k,p}(\mathrm{org}) - \overline{f}_t^{k,p}(\mathrm{sur})$ can be used. However, using more than one surrogate causes the distribution of the difference to approach a normal distribution. This in turn changes the median compared to the mean. Since this change is stronger for skewed distributions, the amount of change is also a function of the frequency of patterns. A true null hypothesis implies that the amount of change of the median compared to the mean, which is induced by using more than one surrogate, is a function of the firing rate. Using more than one surrogate per trial can falsify the significance estimation (a detailed discussion on these effects and the choice of number of surrogates per trial including numerical calibrations can be found in Pipa et al., 2008 and Wu et al., submitted). Thus, using only one surrogate per trial is the most conservative approach. Since our results indicate that the test power with just one surrogate is still sufficiently high, we decided to use a single surrogate per trial.

NULL HYPOTHESIS

The null hypothesis (HØ) of this study assumes that synchronization of spike discharge is not different across different conditions of the experiment. Synchronization of spiking activity across sites

is measured by comparing the frequency of a certain JS-pattern with the expected frequency if neurons are not synchronized. More specifically, here synchronization is defined as coordinated firing on a time scale faster than t_c. Slower effects on a time scale larger t_r such as rate covariation across neurons, are not considered as synchronization. Testing HØ is therefore based on, first, a spike rate and spike train auto-structure corrected measure of synchronization, and second, a test that checks whether an experimental excess or deficit of spike synchrony compared to chance levels is the same or different across conditions. The latter test is based on a mean or median test for each JS-pattern (p) and uses the spike rate and spike train auto-structure corrected measure of synchronization $\Delta f_t^{k,p} = f_t^{k,p}(\text{org}) - f_t^{k,p}(\text{sur})$. HØ is rejected, in case of testing the mean, if the average of $\Delta f_t^{k,p}$ across trials for a certain JS-pattern (p) is different across the factors k. The median testing rejects HØ, if the median of $\Delta f_t^{k,p}$ across trials for a certain JS-pattern (p) is different across the factors k. The median test is more strict, because HØ is only rejected if the difference of $\Delta f_t^{k,p}$ is consistent across trials. Due to the rate and auto-structure correction, based on surrogate data, any source of changes of $\Delta f_t^{k,p}$ other than fine temporal changes on a time scale faster than t_c can be excluded (analytical and numerical demonstration for this can be found in Pipa et al., 2008 and Wu et al., submitted).

PATTERN COMPLEXITY

We investigated JS-pattern complexity ranging from 2 to 8. In order to estimate the impact of JS-pattern complexity on global cortical cooperativity, we distinguish between sub- and supra-patterns. A sub-pattern is a pattern that is embedded in a more complex JS-pattern. Thus, the complexity of a sub-pattern is always smaller than the complexity of the embedding JS-pattern. As an example, any complexity 3 pattern contains three sub-patterns of complexity 2. The more complex embedding pattern is called supra-pattern. It is not straight forward to predict the significance of a given JS-pattern, i.e., whether its sub- and supra-patterns are significantly different from chance level. For a sub-pattern, we know that it occurs at least as often as its supra-pattern. This, however, is not sufficient for qualifying a sub-pattern as significant JSE, even if the embedding supra-pattern has been proven to be significant. The reason is that the expected frequency of chance occurrences of a pattern usually increases with decreasing complexity. Thus, the frequency of a supra-pattern may be larger than the critical minimal frequency of patterns of large complexity, but below the critical frequency for low complexity patterns. In this case, significant high complexity JS-patterns occur, while sub-patterns may not be significant, as can be observed in the data presented here (e.g., **Figures 3C1,C2**). In the opposite case, low complexity JS-patterns are significant, but not their embedding supra-patterns. The simplest explanation is that the supra-pattern does not occur often enough.

RESULTS

We report results based on the analysis of neuronal spiking of 133 multi-units recorded during 12 experimental sessions of a visual memory task (**Figure 1**). The two monkeys performed a total of 9830 trials with on average 80% correct responses. In these data we identified differences in neuronal coupling expressed by modulation of spike synchrony by using a bivariate and multivariate extension of the NeuroXidence method. As detailed in the method section, we first identified synchronous firing on a time scale of 3 ms and corrected for rate modulation on a time scale of 15 ms and slower (**Figure 2**). To test whether joint-spike events were modulated for different experimental factors, we tested whether the rate and spike train auto-structure corrected synchrony differed across experimental conditions. The 3803 JS-patterns visible in **Figure 4C** were composed of 329 JS-pattern with complexity $c = 2$, 1455 with $c = 3$, 1581 with $c = 4$, 407 with $c = 5$, 30 with $c = 6$, and 1 JS-pattern of complexity $c = 7$. For each JS-pattern we determined whether spike synchrony was stronger in correct or in error trials. The highest number of significant JS-patterns was found between 150 and 250 ms after sample onset (**Figures 4A2,B2**). During this period, 808 JS-patterns were significant ($c2$: 91, $c3$: 308, $c4$: 325, $c5$: 73, $c6$:10, $c7$:1). This amounts to 21.2% of the identified JS-patterns and is therefore much higher than the expected number of false positives given by the test level of 1%. As shown in **Figure 4D**, the firing rates as determined in 20 ms intervals cannot explain the difference in JS-patterns. However, the incidence and temporal profile is different for different sites. Some sites participated only in low complexity JS-patterns, while others started with strong modulation of high complexity JS-patterns, which were particularly pronounced during the later phase of the sample presentation (see units 1, 6, and 12 in **Figure 4A2** and units 3, 7, and 10 in **Figure 4B2**).

Integrating across all experiments, we obtained a total of 18150 different JS-patterns (**Table 1**) that changed the level of synchrony in a performance related way (all test levels 1%) and which involved up to 8 units (corresponds to complexity = 7) simultaneously (**Figure 5**). To summarize the results, we determined the frequencies $\rho(t)_c^{\text{specific}}$ for stimulus specific, or $\rho(t)_c^{\text{false}}$ and $\rho(t)_c^{\text{correct}}$ for performance related modulations of spike synchrony per complexity and per sliding window, and expressed this as a rate (s^{-1}). Note that each JS-pattern that contributes to any of the three frequencies $\rho(t)_c$ is in excess of all patterns sampled across all experimental conditions and thus can be considered significant. **Figure 5A** shows $\rho(t)_c^{\text{correct}}$, which is the rate of JS-patterns in sliding windows of 400 ms duration for each complexity, reflecting more synchrony during trials with correct compared to incorrect responses. While **Figure 5A1** represents the entire time course starting 700 ms before sample stimulus presentation and ending after test stimulus processing, **Figure 5A2** features the sample response epoch with higher temporal resolution (sliding window of 100 ms duration). In analogy, **Figure 5B1,B2** show the rate of $\rho(t)_c^{\text{false}}$, that is the rate per second of JS-patterns which reflect more synchrony during trials with incorrect compared to correct responses. These analyses show that across all experiments, the highest rate of performance dependent JS-patterns can reach up to 120 patterns per second which was observed for complexity 4 and during error trials also 3, but not in pairs. High rates of $\rho(t)_c^{\text{correct}}$ and $\rho(t)_c^{\text{false}}$ occurred during all behaviorally relevant epochs: during sample stimulus processing, during early delay and during test stimulus processing. Remarkably, rates $\rho(t)_c^{\text{correct}}$ were particularly high during the delay period of correct trials during which visual memory was required to generate an appropriate response.

To compare changes of frequencies $\rho(t)_c^{\text{correct}}$ and $\rho(t)_c^{\text{false}}$, we computed $\Delta\rho_c^{\text{perf}} = \rho(t)_c^{\text{correct}} - \rho(t)_c^{\text{false}}$ and derived a z-score $z(t)_c^{\text{perf}}$ based on a permutation test, that randomized class labels for correct and incorrect trials, to test whether observed differences can

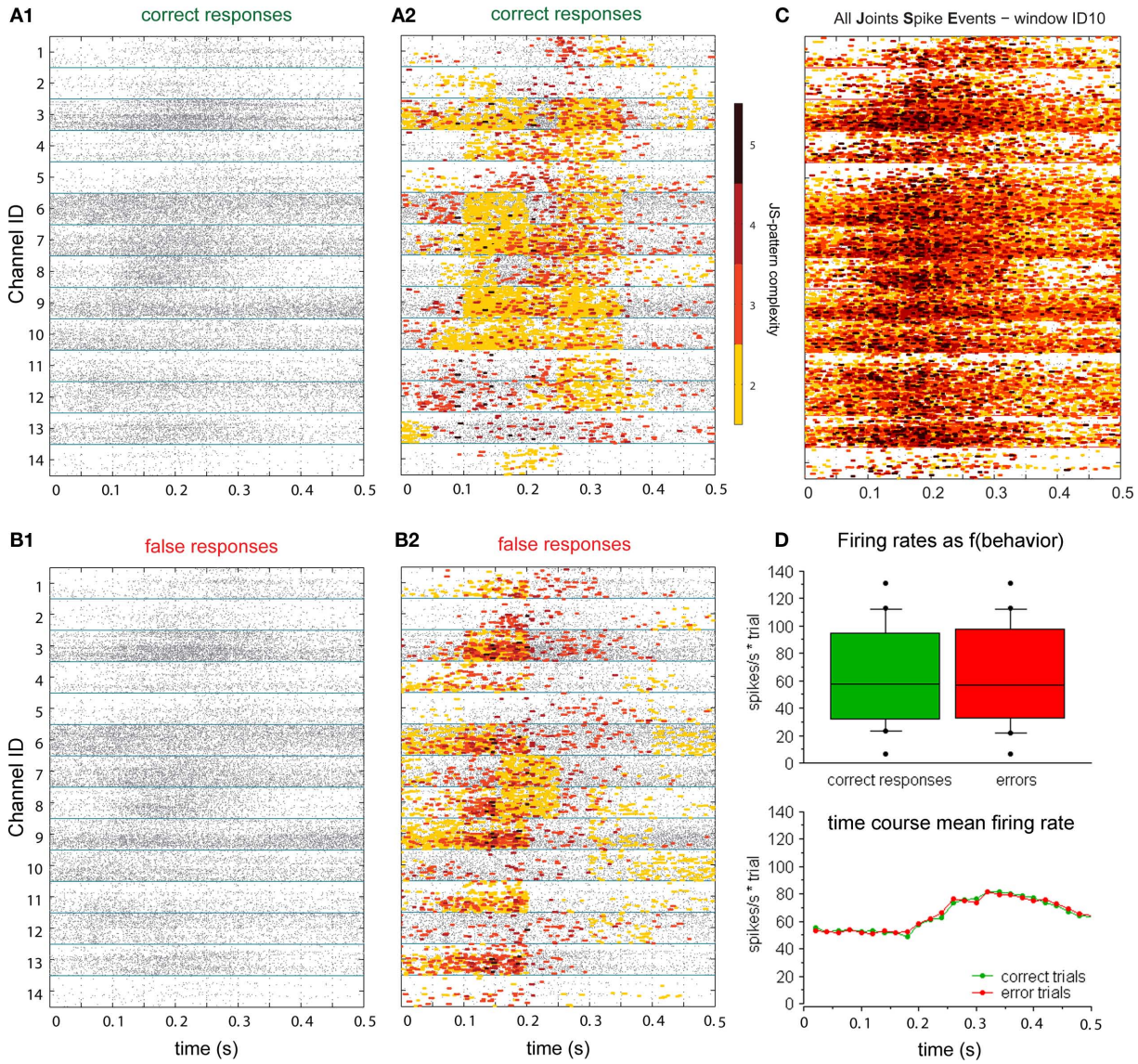


FIGURE 4 | Millisecond precise joint spike events (JSE) in prefrontal multi-units. (A1, B1) Raster plots of 14 simultaneously recorded multi-unit signals (dots) for the sample stimulus period in trials with correct **(A1)** and erroneous behavioral responses **(B1)**. **(A2, B2)** The same raster plots with superimposed JSEs marked by colored squares, each covering an interval of maximal 3 milliseconds; colors reflect JSE complexity, i.e. how many units participated in a pattern. Yellow indicates JSE of pairs of units (c = 2), red and darker colors represent higher complexities ranging from c = 3 to 5. **(A2)** shows only JSEs which occurred in sets of sites that synchronized more often during trials with correct behavioral responses, while **(B2)** shows those JSE that occurred more often during error trials. **(C)** All JS-patterns combined from **(B2, C2)**, irrespective of their modulation by behavioural performance. **(D)** Distributions of firing rates across all units, split for trials with correct and erroneous behavioral responses and below as average time course.

be explained by chance. Based on critical *z*-scores, which were corrected for multiple comparison across different complexities and different sliding windows, we identified periods and complexities ("time complexity bins") with significant modulations of the frequency of JS-patterns that each showed a significant and performance related modulation of spike synchrony (**Figure 5C1**). In other words, **Figure 5C** measures how significant the overall increase of spike synchrony was in correct compared to error trials. Strongest modulation of spike synchrony was observed for JS-patterns of higher complexities and during the delay period. While the maximum complexity with significant modulations of $z(t)_c^{\mathrm{perf}}$ reached 4 during the sample presentation (**Figure 5C2**), the complexity reached up to 7 during the delay. Surprisingly, pairwise synchrony measured by $z(t)_c^{\mathrm{perf}}$ was not significantly modulated. In general, behaviorally relevant periods are dominated by increases of synchronous activity in correct trials. Only during late delay and test stimulus presentation, spike synchrony of low complexities was stronger during error trials (**Figure 5C1**).

In stark contrast to the results for trials with correct behavioral responses, lower numbers of JSE were observed during error trials (**Figure 5B**). Most notably, this observation is not caused by some scaling or signal-to-noise problem, because during test stimulus

Table 1 | Number of joint-spike events (JSE) detected to be modulated by behavioral performance and stimulus specificity.

# JSE	Performance	Specificity
COMPLEXITY		
2	743	792
3	3283	3225
4	7239	5850
5	4733	3808
6	1443	1598
7	595	–
8	114	–
Sum	18150	15273
Units	122	118
Sessions	13	12

processing, when the monkey had to retrieve memory content and compare this to the test stimulus, an increase of JSE frequency was observed which is compatible with the JSE frequency observed during sample stimulus processing and the early delay in correct trials (compare **Figures 5A1,B1**). An interesting feature of JSE increases during test stimulus processing is that complexity during correct trials is higher (4–6) compared to the complexity of JSE during error trials (2–5). Finally, we computed a contrast for JSE modulation in trials with correct and incorrect behavioral responses after *z*-transformation using the variance obtained from permuted data in each experiment (**Figure 5C**). The main finding of this analysis is that during the first half of the delay, JSE increases during successful trials outbalance JSE during error trials while during late delay the converse is true. This is not a shaky effect, because these effects hold for many hundred milliseconds and are consistent across numerous adjacent complexities (**Figure 5D**).

We then also tested whether the occurrence of synchronous spike patterns was stimulus specific (**Figure 6**). Stimulus specific modulation of synchrony could be identified in 15273 patterns involving up to six units (**Table 1**). When comparing **Figures 5 and 6**, it is evident that stimulus specific modulation of synchronous firing is much more confined to stimulus response epochs than performance dependent modulation with two interesting exceptions: during early and mid delay. First, during early delay, JSE of complexity 3 occurred in a stimulus specific fashion for 800 ms, supporting the idea that synchronous neuronal activity during early delay is involved in encoding and stabilizing memory related activity. Second, well over a second into the delay, a short burst of JSE of complexity 5 discharged highly significant stimulus specific synchronous spikes (**Figure 6B**) which is reminiscent of the elevated rate of JSE $c = 5$ observed during correct trials in the performance analysis (**Figure 5A1**). Another interesting relation between performance dependent synchrony increases and stimulus specific synchrony increases can be observed during test stimulus processing when the monkey has to perform a comparison between arriving sensory information and memory content: First, the time complexity pattern of stimulus specific JSE peaks at around 300 ms after test stimulus onset at complexity 4 which matches the peak of performance dependent JSE modulation in correct trials. This is, of course, expected, because the analysis of stimulus specific JSE was exclusively performed on trials with correct behavioral responses. Note the different JSE pattern during error trials. Second, stimulus specific JSE modulation during test stimulus processing was more than twice as strong as during sample stimulus processing (compare, e.g., JSE of complexity 4 during "S" and "T" periods in **Figure 6A**), which was not the case for correct trials in the performance dependent modulation (**Figure 5A1**).

The temporal precision of JS-patterns is an important parameter of this study. We have chosen t_c to be 3 ms. Other studies used less precise patterns that may extend from 5 ms to even more imprecise recurrences. To select the appropriate temporal scale for our analysis, we performed the same analysis procedure for four different t_c windows with ($t_c = 2$, 3, 5, and 7 ms). The lower bound of rate responses were scaled in the same way with $t_r = \alpha^* t_c$, and $\alpha = 3$ leading to t_r values of 6, 9, 15, and 21 ms. We found that effects across the four scales were compatible, but strongest modulation of performance and stimulus related modulations of spike synchrony were observed for $t_c = 3$ ms. This finding suggests that the experimental data reported here were dominated by JS-patterns with a temporal precision of 3 ms. For smaller t_c, i.e., $t_c = 2$ ms, much less JSE were detected since most of the observed JSEs had an imprecision larger than 2 ms. For longer t_c, i.e., 5 and 7 ms, the rate correction was effectively stronger because the model imprecision t_c was too large for the dominating imprecision in the data.

DISCUSSION

The main finding of this study is that patterns of precise spike synchrony (≤3 ms), here referred to as joint-spike events (JSE), change their frequency of occurrence and their complexity in a dynamic way which depends on the behavioral success of the monkey and the stimuli in the memory task. These JSE are no rare events, nor do they occur by chance. JSE with performance related changes occur more than 20 times more often than expected by chance. This raises the question how relevant synchronous firing may be for cortical processing (Herrmann et al., 2004; Uhlhaas et al., 2009).

ANALYSIS APPROACH

Synchronous patterns of higher complexity had been observed in behaving monkeys before, but there have been and still are intensive discussions about whether such events might occur just by chance (Baker and Lemon, 2000). First and foremost, complex and usually time varying structures of spike trains caused the fear that model based analyses, which assume either stationary firing rates or idealized spike density distributions like following a Poisson distribution or reflecting a renewal process, cause false positive findings. To avoid such assumptions, we have chosen a non-parametric approach that estimates the amount of chance JSE per trial based on permuted data from the exact same experiment thus preserving the auto-structure. Therefore any kind of complex structure, but also any kind of rate modulation slower than t_r, the time scale for rate changes considered by NeuroXidence. At the same time the method stays very sensitive, on a level that is compatible to other methods, like standard pairwise cross correlation, factorial recoding of synchronous spike trains derived from data compression algorithms (Schnitzer and Meister, 2003) or the Unitary Event method (Pipa et al., 2008).

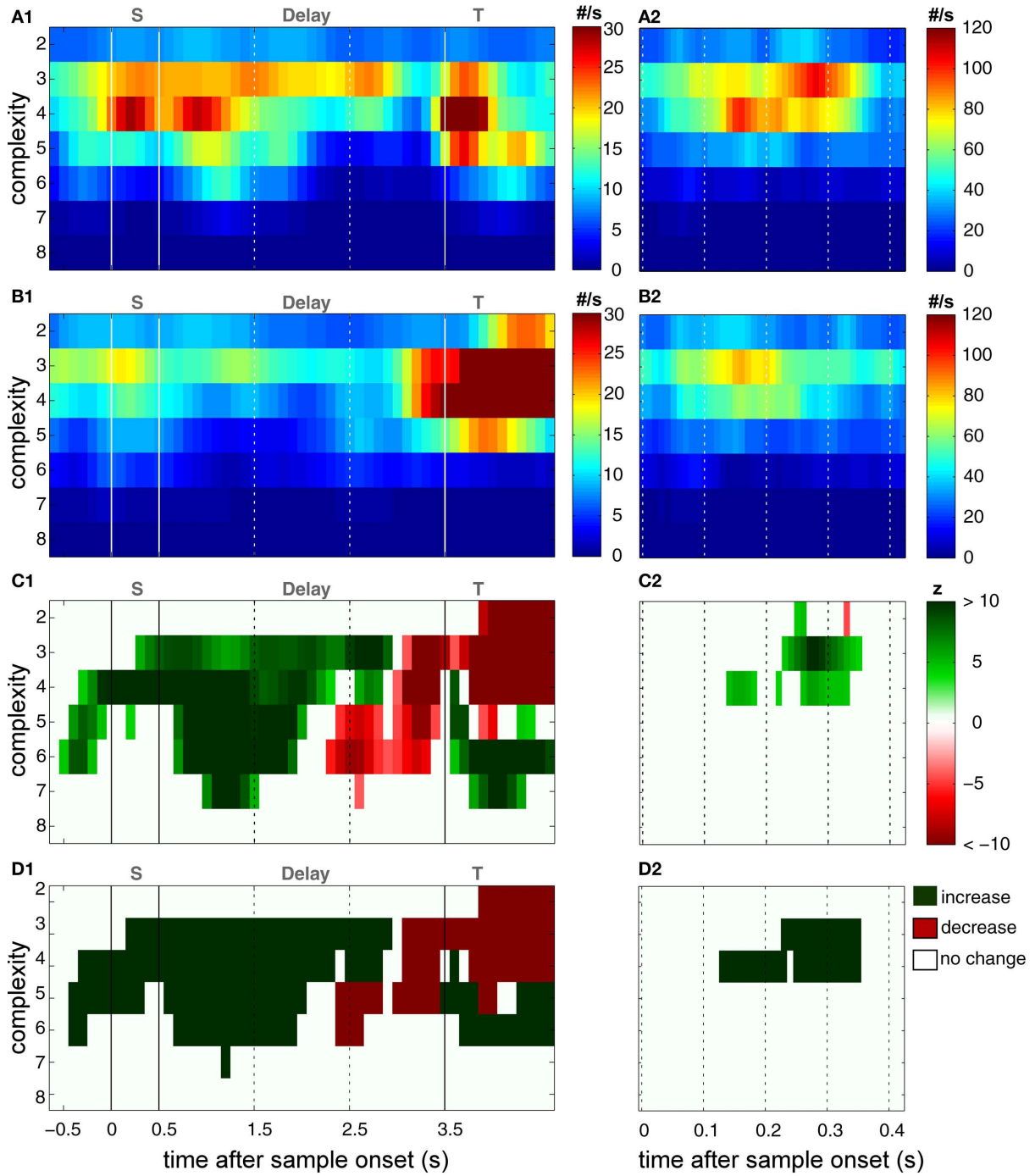


FIGURE 5 | Time course and performance dependence of joint-spike event complexity. Rate of JSE (*z*-axis/color scale) with complexity (*y*-axis) ranging from 2 to 8 which was significantly modulated by behavioral performance, displayed as a function of time with respect to the onset of sample stimuli (*x*-axis). **(A,B)** JSE rate plotted for sets of recoding sites which expressed more JSE during trials with correct behavioral responses **(A)** and during error trials **(B)**. **(C)** Time resolved contrast of the rates plotted in **(A,B)** after *z*-transformation. *z*-scores were computed by taking the absolute difference between values in **(A,B)**, divided by the SD of values obtained in permuted trials with correct and erroneous behavioral responses, thus referencing to the variance of the same experiment. The critical *z*-value was 4.2, given a test level of 1% and a Bonferroni correction for 48 sliding windows and 7 complexities. **(D)** Colored time complexity bins mark periods of significant differences of performance dependent joint-spike events at a test level of 0.1%. Significance was evaluated using a rank order test of the original differences shown in **(C1,C2)** compared with results obtained based on permuted trials (compare to **Figures 3B1,B2**). On the left **(A1–D1)**, pattern incidence is shown for the entire duration of the task based on analyses with sliding windows of 400 ms duration, while on the right **(A2–D2)**, pattern incidence was analyzed with sliding windows of 100 ms duration and plotted exclusively for the first 400 ms of sample stimulus presentation.

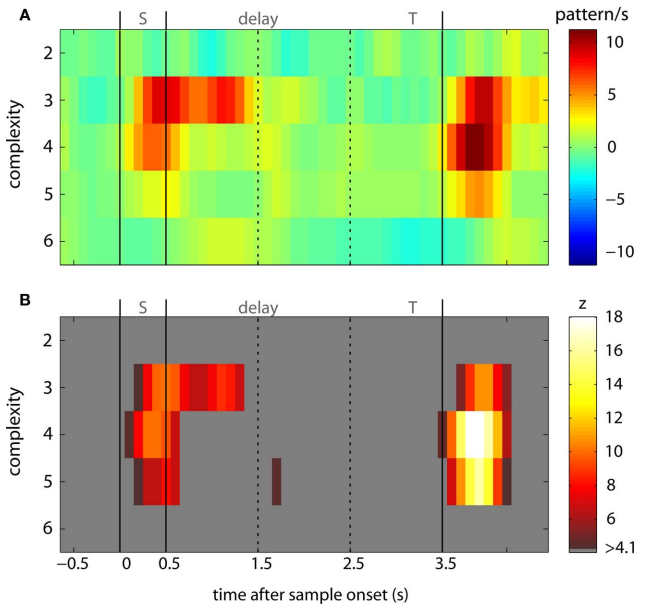


FIGURE 6 | Time course of stimulus specific joint-spike events. (A) Incidence of JSE of complexities 2–6 which exhibited a stimulus specific modulation during 3 ms short intervals in patterns per second. **(B)** Statistical evaluation of stimulus specific JSE incidence was performed in analogy to the analysis of performance dependent JSE incidence. *z*-sores were computed by dividing the number of JSE in trials in which a specific visual object was memorized divided by the SD of JSE in trials from the same recording session, but after permutation for memorized stimuli. To this end we used the identical data and the identical number of trails per stimulus, but permuted stimuli randomly across all trails. *z*-transforms were performed for each individual complexity and based on the SD derived from the entire time course starting at the beginning of the baseline and including all other epochs until after button press of the monkey. The critical *z*-value was 4.08 given a test level of 1% and a Bonferroni correction for 48 sliding windows and 5 complexities.

A complication that arises when dealing with the activity of a large number of neurons is that the number of JS-patterns grows so large that standard approaches based on single JS-patterns are no longer applicable and the amount of information is overwhelming and may even become confusing. To overcome this problem, we chose a simple strategy which consists of computing the frequency of JS-patterns that have before been shown to be significantly modulated by experimental factors. Using this simplification, we lost the identity of JS-patterns, but we were able to condense observations to a very handy low dimensional set of numbers: frequency and pattern complexity for each time window. However, this reduction requires a second level of hypothesis testing, since, even though each JS-pattern that is included in the statistics is significant, the expected level of significant JS-patterns is unknown. Therefore we used another robust non-parametrical test, based on permutations. This test compares the frequency of patterns observed in trials selected for original experimental factors (i.e., trials with correct versus incorrect behavioral responses, or the different stimuli the monkey had to memorize) and, a second set of JS-pattern frequencies derived from permuted class labels. This test preserved the simultaneity of recorded neuronal activity, but swapped trials between different experimental conditions in order to destroy any differences in neuronal synchrony between conditions. Using this, we derived a *z*-sore and applied a Bonferroni correction. This second step is robust against changes of rates and particular auto-structures of neuronal activity across conditions, because NeuroXidence uses surrogate data which maintain this structure. Since a permutation test is used, no assumption is made regarding the distribution of any parameter. Last, but not least, testing median and mean JSE frequency at the level of individual JS-patterns provided very similar results for the modulation of spike synchrony.

To avoid that synchronous activity could have escaped our attention due to shallow significance levels we set the criteria for detecting JS-patterns as conservative as possible. We chose the required temporal precision t_c of a synchronous firing pattern to be equal or less than 3 ms (see also Pipa et al., 2007 for further discussion on time scale separation). This parameter confines the analysis to very precise patterns, in particular if more than two multi-units were involved. On the one hand, the interval t_c can also be seen as a necessary upper bound of time scales which define a very fast increase of firing rate covariation across all units participating in a JS-pattern. On the other hand, the second interval t_r is important to contrast synchrony to all other kinds of rate covariation, in particular, on slower time scales. We chose t_r = 15 ms which implies that any covariation of firing rates occurring within more than 15 ms (or slower as 66 Hz) is considered as rate. Any covariation of firing rates occurring within less than 3 ms (or faster than 333 Hz) is considered to be a JS-pattern. It is important to note that 15 ms as an upper bound of rate covariations is very conservative given that firing rate changes are typically observed with bin sizes of several tens of milliseconds.

METHODOLOGICAL LIMITATIONS

A first limitation of the current approach is that the analysis does not consider the nature of the analyzed JS-patterns, e.g., their spatial structure. This implies that for example information about the similarity of patterns accounted for different experimental conditions could not be analyzed. The spatial structure, however, might be very relevant for the neuronal processes. Furthermore, this limitation implies that similarity and stability of JS-patterns over time across different sliding windows were not analyzed. This might be very relevant, as a stable increase of JSE during the delay period which lasted for more than 2 s, may have been composed of very different sets of JS-patterns over time. Knowledge of this stability might allow to distinguish between the two hypotheses which either assume that information is encoded in stable and rather small subpopulations, or, that information is encoded on the sequences of many and very rich transitions of different neuronal states. However, technically this tracking of stability seems very demanding if not even impossible at the time.

A second limitation of the current analysis is that we cannot determine the actual size of neuronal assemblies which are involved in the encoding and maintenance of behaviorally relevant information. Given our finding that up to 8 multi-units out of a population of 10–24 can be involved in behaviorally relevant JS-patterns, one may conclude that assemblies can be very large, maybe involving a third or even half of the neurons. From a theoretical perspective, such a code may appear very attractive, because the coding space becomes really large, if on average a third or half of the neurons in a population engage in JS-patterns.

A third limitation of the current approach being restricted to multi-unit signals implies that we have investigated JS-patterns among several small, spatially separated neuronal populations and *not* single, well isolated units which are generally considered to reflect the activity of a single neuron. As we have recorded all wave forms at sufficient spectrotemporal resolution we have tried to sort spikes, but with limited success, the major obstacle being that most of the signals recorded for this study were not recorded with tetrodes, but with single-ended fiber microelectrodes. Therefore, despite good S/N, synchronous spikes occurring at individual sites were most of the time misclassified as deformed rare spike waveforms, which were discarded. Thus we restricted the analysis and interpretation of this study to multi-unit signals. As a consequence, estimates of the assembly size as discussed above are even further hampered by concluding about synchronous firing only for several groups of neurons. However, this is a conservative approach for answering the question whether synchronous firing exists above chance, because each individual locally observed spike might already represent a synchronous event. Since we do not rely on any statistical assumption for the distribution of JS-patterns occurring by chance, but simply permute the existing time series, there is no trivial explanation for false positive events.

Our finding that JS-patterns with a precision of 3 ms and rate corrected at a time scale of 9 ms were more modulated by experimental variables like behavioral performance than patterns at more precise or less precise scales suggests that the precision of 3 ms is biologically meaningful. Compared to other studies, which reported JS-patterns of 5 ms precision, these time scales appear to be too short (Riehle et al., 1997). This difference, however, can be partly attributed to the chosen analysis techniques. For example in the paper by Riehle et al. (1997) the unitary event method was used. This method detects JSE based on binned spike trains with a bin size corresponding to the assumed temporal precision of the JS-pattern. Binning however requires that the window must be larger than the actual precision of the pattern, because JS-patterns close to the boarder of two bins have an effectively much smaller window than JS-patterns which are centered on a bin. Therefore, the detectability of a JS-pattern with methods using binning depends on the relative position of JS-patterns within the bins. In contrast, NeuroXidence describes a JS-pattern by the exact preset imprecision, given by t_c. Results presented in Pipa et al. (2007) demonstrate this link between two time scales for exactly the same data as presented in Riehle et al. (1997). By applying NeuroXidence to these data, the authors confirmed the previous results based on the binning UE method to contain JSEs on the same time scale of 5 ms. However, reducing the preset time scale from 5 to 3 ms for the NeuroXidence method, resulted in an even stronger deviation from the chance level, while the number of patterns detected by binning decreased significantly. This indicates that the NeuroXidence method is more sensitive to find the lower bound of spiking precision of JS-patterns.

An earlier study that has successfully dealt with higher order spike patterns extracted from simultaneous recordings of retinal ganglion cells has used a factorial recoding of synchronous spike trains that was derived from data compression algorithms (Schnitzer and Meister, 2003). This method is also very efficient in detecting and storing synchronous spike trains, but unfortunately also involved time binning which has been shown to miss coincidences (Pipa et al., 2007). The advantage of this method is that one can preserve the identity of each unit and of all groups of units involved in synchronous firing which is certainly a feature we want to include in future versions of our analysis technique.

FUNCTIONAL IMPLICATIONS

For the analysis of synchronous firing patterns one can distinguish the complexity from the order of a JS-pattern. While the complexity just gives the number of neurons or sites involved in a JS-pattern, the order determines the real underlying correlation structure. The latter is necessary to distinguish between the chance level and the occurrence of sub-patterns of a more complex JS-pattern (Martignon et al., 2000; Nakahara and Amari, 2002; Schneider and Grün, 2003). However, the latter is also more of theoretical nature than of any practical relevance. It had been demonstrated that the amount of data necessary to distinguish between a certain set of orders is gigantic compared to the amount of data which is usually available in real experiments, but also with respect to the amount of information a neuron in the cortex would have to decode if sub-patterns would be able to carry relevant information. Both arguments are in favor of using the much simpler JS-pattern complexity. Pattern complexity can be simply interpreted as a kind of input saliency for downstream neurons. The higher the complexity, the more neurons participate, the more salient the input pattern is, because the more numerous simultaneous or nearly simultaneous inputs are, the more they can draw from spatial summation properties of postsynaptic membranes resulting in more rapid depolarization or even nonlinear amplification of the postsynaptic membrane potential.

What precisely synchronous spike firing of distributed populations of cortical neurons really means for information processing and generating appropriate behavior is not yet well understood. The observation that JSE incidence was massively increased even before the predictable end of the memory delay is reminiscent of JSE in motor cortex due to sensorimotor expectancy (Riehle et al., 1997), suggesting that spike synchrony in PFC could reflect a mechanism for the temporal organization of executive processes. However, the finding that synchronous spiking is modulated by, both, behavioral performance and the memorized visual stimuli, suggests that synchrony is a very fundamental processing mechanism of the cortex. How well synchronous spike signals can be used in the future to actually decode information processed and maintained in distributed cortical circuits remains to be seen. The well established fact that coincident neuronal activity is a potent trigger for synaptic plasticity suggests that synchronous activity may be a better predictor for what cortical circuits need to be adapted for rather than an expression of their current performance.

ACKNOWLEDGMENTS

We are grateful to Hanka Klon-Lipok, Michaela Klinkmann, Urda Franzius and Ellen Städtler for their enduring long-term efforts to train our animals, prepare recording equipment and help us during experiments, Thomas Maurer for excellent technical support, Christiane Kiefert and Clemens Sommer for dedicated animal care. Funding: VolkswagenStiftung I/77 142 (Matthias H. J. Munk), EU NEST-Pathfinder GABA Project 043309. In part financed by the EC Project PHOCUS Grant 240763 (Gordon Pipa).

REFERENCES

Abeles, M. (1982). *Local Cortical Circuits, An Electrophysiological Study*. Heidelberg: Springer.

Abeles, M. (1991). *Corticonics, Neural Circuits of the Cerebral Cortex*. New york: Cambridge University Press.

Abeles, M., Bergman, H., Margalit, E., and Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol.* 70, 1629–1638.

Ahissar, E., Vaadia, E., Ahissar, M., Bergman, H., Arieli, A., and Abeles, M. (1992). Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context. *Science* 257, 1412–1415.

Anderson, B., Harrison, M., and Sheinberg, D. L. (2006). A multielectrode study of the inferotemporal cortex in the monkey: effects of grouping on spike rates and synchrony. *Neuroreport* 17, 407–411.

Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* 273, 1868–1871.

Baker, S. N., and Lemon, R. N. (2000). Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J. Neurophysiol.* 84, 1770–1780.

Bour, L. J., van Gisbergen, J. A., Bruijns, J., and Ottes, F. P. (1984). The double magnetic induction method for measuring eye movement – results in monkey and man. *IEEE Trans. Biomed. Eng.* 31, 419–427.

Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian learning rule. *Annu. Rev. Neurosci.* 31, 25–46.

Constantinidis, C., and Goldman-Rakic, P. S. (2002). Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J. Neurophysiol.* 88, 3487–3497.

De Luca, M., Beckmann, C. F., De, S. N., Matthews, P. M., and Smith, S. M. (2006). fMRI resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage* 29, 1359–1367.

de Oliveira, S. C., Thiele, A., and Hoffmann, K. P. (1997). Synchronization of neuronal activity during stimulus expectation in a direction discrimination task. *J. Neurosci.* 17, 9248–9260.

deCharms, R. C., and Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* 381, 610–613.

Delage, Y. (1919). *Le rêve. Étude Psychologique, Philosophique et Littéraire*. Paris: Presses Universitaires de France.

Dong, Y., Mihalas, S., Qiu, F., von der, H. R., and Niebur, E. (2008). Synchrony and the binding problem in macaque visual cortex. *J. Vis.* 8, 30–16.

Douglas, R. J., and Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451.

Dudkin, K. N., Kruchinin, V. K., and Chueva, I. V. (1995). Neurophysiologic correlates of the decision-making processes in the cerebral cortex of monkeys during visual recognition. *Neurosci. Behav. Physiol.* 25, 348–356.

Gochin, P. M., Colombo, M., Dorfman, G. A., Gerstein, G. L., and Gross, C. G. (1994). Neural ensemble coding in inferior temporal cortex. *J. Neurophysiol.* 71, 2325–2337.

Gray, C. M., Konig, P., Engel, A. K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338, 334–337.

Grün, S., Diesmann, M., and Aertsen, A. (2002). Unitary events in multiple single-neuron spiking activity: 1. Detection and significance. *Neural Comput.* 14, 43–80.

Grün, S., Diesmann, M., Grammont, F., Riehle, A., and Aertsen, A. (1999). Detecting unitary events without discretization of time. *J. Neurosci. Methods* 94, 67–79.

Grün, S., Riehle, A., and Diesmann, M. (2003). Effect of cross-trial nonstationarity on joint-spike events. *Biol. Cybern.* 88, 335–351.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Herrmann, C. S., Munk, M. H., and Engel, A. K. (2004). Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn. Sci. (Regul. Ed.)* 8, 347–355.

Hirabayashi, T., and Miyashita, Y. (2005). Dynamically modulated spike correlation in monkey inferior temporal cortex depending on the feature configuration within a whole object. *J. Neurosci.* 25, 10299–10307.

Hoerzer, G. M., Liebe, S., Schloegl, A., Logothetis, N. K., and Rainer, G. (2010). Directed coupling in local field potentials of macaque V4 during visual short-term memory revealed by multivariate autoregressive models. *Front. Comput. Neurosci.* 4:14. doi: 10.3389/fncom.2010.00014

Kayser, C., Montemurro, M. A., Logothetis, N. K., and Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron* 61, 597–608.

Kohn, A., and Smith, M. A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* 25, 3661–3673.

Kreiter, A. K., and Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *J. Neurosci.* 16, 2381–2396.

Liebe, S., Hoerzer, G., Logothetis, N. K., Maass, W., and Rainer, G. (2009). "Long range coupling between V4 and PF in theta band during visual short-term memory," in *39th Annual Conference of the Society for Neuroscience, Program 652.20.*

Lima, B., Singer, W., Chen, N. H., and Neuenschwander, S. (2010). Synchronization dynamics in response to plaid stimuli in monkey V1. *Cereb. Cortex.* 20, 1556–1573.

Maldonado, P. E., Friedman-Hill, S., and Gray, C. M. (2000). Dynamics of striate cortical activity in the alert macaque: II. Fast time scale synchronization. *Cereb. Cortex* 10, 1117–1131.

Martignon, L., Deco, G., Laskey, K., Diamond, M., Freiwald, W., and Vaadia, E. (2000). Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies. *Neural Comput.* 12, 2621–2653.

Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55, 131–141.

Murthy, V. N., and Fetz, E. E. (1996). Synchronization of neurons during local field potential oscillations in sensorimotor cortex of awake monkeys. *J. Neurophysiol.* 76, 3968–3982.

Nakahara, H., and Amari, S. (2002). Information-geometric measure for neural spikes. *Neural Comput.* 14, 2269–2316.

Nowak, L. G., Munk, M. H., James, A. C., Girard, P., and Bullier, J. (1999). Cross-correlation study of the temporal interactions between areas V1 and V2 of the macaque monkey. *J. Neurophysiol.* 81, 1057–1074.

Pipa, G., Riehle, A., and Grün, S. (2007). Validation of task-related excess of spike coincidences based on NeuroXidence. *Neurocomputing* 70, 2064–2068.

Pipa, G., Wheeler, D. W., Singer, W., and Nikolic, D. (2008). NeuroXidence: reliable and efficient analysis of an excess or deficiency of joint-spike events. *J. Comput. Neurosci.* 25, 64–88.

Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Slovin, H., and Abeles, M. (1998). Spatiotemporal structure of cortical activity: properties and behavioral relevance. *J. Neurophysiol.* 79, 2857–2874.

Riehle, A., Grün, S., Diesmann, M., and Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278, 1950–1953.

Rodriguez, R., Kallenbach, U., Singer, W., and Munk, M. H. (2010). Stabilization of visual responses through cholinergic activation. *Neuroscience.* 165, 944–954.

Sakurai, Y., and Takahashi, S. (2006). Dynamic synchrony of firing in the monkey prefrontal cortex during working-memory tasks. *J. Neurosci.* 26, 10141–10153.

Salin, P. A., and Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiol. Rev.* 75, 107–154.

Schneider, G., and Grün, S. (2003). Analysis of higher-order correlations in multiple parallel processes. *Neurocomputing* 52–54, 771–777.

Schnitzer, M. J., and Meister, M. (2003). Multineuronal firing patterns in the signal from eye to brain. *Neuron* 37, 499–511.

Shadlen, M. N., and Movshon, J. A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* 24, 67–25.

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–25.

Singer, W., and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586.

Sjöström, P. J., Rancz, E. A., Roth, A., and Hausser, M. (2008). Dendritic excitability and synaptic plasticity. *Physiol. Rev.* 88, 769–840.

Steinmetz, P. N., Roy, A., Fitzgerald, P. J., Hsiao, S. S., Johnson, K. O., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* 404, 187–190.

Thiele, A., and Hoffmann, K. P. (2008). Neuronal firing rate, inter-neuron correlation and synchrony in area MT are correlated with directional choices during stimulus and reward expectation. *Exp. Brain Res.* 188, 559–577.

Thiele, A., and Stoner, G. (2003). Neuronal synchrony does not correlate with motion coherence in cortical area MT. *Nature* 421, 366–370.

Uhlhaas, P. J., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolic, D., and Singer, W. (2009). Neural synchrony in cortical networks: history, concept and current status. *Front. Integr. Neurosci.* 3:17. doi: 10.3389/neuro.07.017.2009.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., and Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* 373, 515–518.

van den Heuvel, M. P., Stam, C. J., Kahn, R. S., and Hulshoff Pol, H. E. (2009) Efficiency of functional brain networks and intellectual performance. *J. Neurosci.* 29, 7619–7624.

Yang, Y., DeWeese, M. R., Otazu, G. H., and Zador, A. M. (2008). Millisecond-scale differences in neural activity in auditory cortex can drive decisions. *Nat. Neurosci.* 11, 1262–1263.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 December 2009; paper pending published: 20 December 2009; accepted: 08 May 2011; published online: 08 June 2011.

Citation: Pipa G and Munk MHJ (2011) Higher order spike synchrony in prefrontal cortex during visual memory. Front. Comput. Neurosci. ***5****:23. doi: 10.3389/fncom.2011.00023*

frontiers in
COMPUTATIONAL NEUROSCIENCE

METHODS ARTICLE
published: 28 May 2010
doi: 10.3389/fncom.2010.00012

Bayesian inference for generalized linear models for spiking neurons

Sebastian Gerwinn[1,2]*, Jakob H. Macke[1,2,3] and Matthias Bethge[1,2]

[1] *Computational Vision and Neuroscience, Max Planck Institute for Biological Cybernetics, Tübingen, Germany*
[2] *Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Tübingen, Germany*
[3] *Gatsby Computational Neuroscience Unit, University College London, London, UK*

Edited by:
Peter Dayan,
University College London, UK

Reviewed by:
Jonathan Pillow, University of Texas, USA
Fabrizio Gabbiani, Baylor College of Medicine, USA

***Correspondence:**
Sebastian Gerwinn, Computational Vision and Neuroscience, Max Planck Institute for Biological Cybernetics, Spemannstrasse 41, 72076 Tübingen, Germany.
e-mail: sgerwinn@tuebingen.mpg.de

Generalized Linear Models (GLMs) are commonly used statistical methods for modelling the relationship between neural population activity and presented stimuli. When the dimension of the parameter space is large, strong regularization has to be used in order to fit GLMs to datasets of realistic size without overfitting. By imposing properly chosen priors over parameters, Bayesian inference provides an effective and principled approach for achieving regularization. Here we show how the posterior distribution over model parameters of GLMs can be approximated by a Gaussian using the Expectation Propagation algorithm. In this way, we obtain an estimate of the posterior mean and posterior covariance, allowing us to calculate Bayesian confidence intervals that characterize the uncertainty about the optimal solution. From the posterior we also obtain a different point estimate, namely the posterior mean as opposed to the commonly used maximum *a posteriori* estimate. We systematically compare the different inference techniques on simulated as well as on multi-electrode recordings of retinal ganglion cells, and explore the effects of the chosen prior and the performance measure used. We find that good performance can be achieved by choosing an Laplace prior together with the posterior mean estimate.

Keywords: spiking neurons, Bayesian inference, population coding, sparsity, multielectrode recordings, receptive field, GLM, functional connectivity

INTRODUCTION

A common problem in system neuroscience is to understand how information about the sensory stimulus is encoded in sequences of action potentials (spikes) of sensory neurons. Given any stimulus, the goal is to predict the neural response as well as possible, as this can give insights into the computations carried out by the neural ensemble. To this end, we want to have flexible generative models of the neural responses which can still be fit to observed data. The difficulty in choosing a model is to find the right trade-off between flexibility and tractability. Adding more parameters or features to the model makes it more flexible but also harder to fit, as it is more prone to overfitting. The Bayesian framework allows one to control for the model complexity even if the model parameters are underconstrained by the data, as imposing a prior distribution over the parameters allows regularizing the fitting procedure (Lewicki and Olshausen, 1999; Ng, 2004; Steinke et al., 2007; Mineault et al., 2009).

From a statistical point of view, building a predictive model for neural responses constitutes a regression problem. Linear least squares regression is the simplest and most commonly used regression technique. It provides a unique set of regression parameters, but one that is derived under the assumption that neural responses in a time bin are Gaussian distributed. This assumption, however, is clearly not appropriate for the spiking nature of neural responses. Generalized Linear Models (GLMs) provide a flexible extension of ordinary least squares regression which allows one to describe the neural response as a point process (Brillinger, 1988; Chornoboy et al., 1988) without losing the possibility of finding a unique best fit to the data (McCullagh and Nelder, 1989; Paninski, 2004).

The simplest example of the generalized linear spiking neuron model is the linear-nonlinear Poisson (LNP) cascade model (Chichilnisky, 2001; Simoncelli et al., 2004). In this model, one first convolves the stimulus with a linear filter, subsequently transforms the resulting one-dimensional signal by a pointwise nonlinearity into a non-negative time-varying firing rate, and finally generates spikes according to an inhomogeneous Poisson process. Importantly, the GLM model is not limited to noisy Poisson spike generation: analogous to the stimulus signal, one can also convolve the recent history of the spike train with a feedback filter and transform the superposition of both stimulus and spike history filter outputs through the pointwise nonlinearity into an instantaneous firing rate in order to generate the spike output. In this way one can mimic dynamical properties such as bursts, refractory periods and rate adaptation. Finally, it is possible to add further input signals originating from the convolution of a filter kernel with spike trains generated by other neurons (Borisyuk et al., 1985; Brillinger, 1988; Chornoboy et al., 1988). This makes it possible to account for couplings between neurons, and to model data which exhibit so called noise correlations, i.e., correlations which can not be explained by shared stimulus selectivity. Although the GLM only gives a phenomenological description of the neurons' properties, it has been shown to perform well for the prediction of spike trains in the retina (Pillow et al., 2005, 2008), in the hippocampus (Harris et al., 2003) and in the motor cortex (Truccolo et al., 2010).

In this paper we seek to explore the potential uses and limitations of the framework for approximate Bayesian inference for GLMs based on the Expectation Propagation algorithm (Minka, 2001). With this framework, we can not only approximate the

posterior mean but also the posterior covariance and hence compute confidence intervals for the inferred parameter values. Furthermore, the posterior mean is an alternative to the commonly used point estimators, maximum *a posteriori* (MAP) or maximum likelihood. Like the MAP also the posterior mean can be used with a Gaussian or a Laplacian prior leading to an L2 or an L1-norm regularization. To establish the approximate inference framework, we compare these point estimates on the basis of two different quality measures: prediction performance and filter reconstruction error. In addition, we investigate different binning schemes and their impact on the different inference procedures. Along with the paper we publish a MATLAB (the code is available at http://www.kyb.tuebingen.mpg.de/bethge/code/glmtoolbox/) toolbox in order to support researchers in the field to do Bayesian inference over the parameters of the GLM spiking neuron model.

The paper is organized as follows. In Section "Generalized Linear Modeling for Spiking Neurons", we review the definition of the Generalized Linear Model and present the expansion into a high-dimensional feature space. We explain how a Laplace prior can improve the prediction performance in this setting and how different loss functions can be used to rate different quality aspects. In Section "Approximating the Posterior Distribution Using EP", we present how the posterior distribution for observed data in the GLM setting can be approximated via the Expectation Propagation algorithm. Finally in Section "Potential Uses and Limitations" we systematically compare the MAP estimator to the posterior mean assuming Gaussian versus a Laplacian prior. In addition we apply the GLM framework to multi-electrode recordings from a population of retinal ganglion cells and discuss the potential differences of discretizing time directly or discretizing the features.

GENERALIZED LINEAR MODELING FOR SPIKING NEURONS

SPECIFYING THE LIKELIHOOD

The Generalized Linear Model (GLM) of spiking neurons describes how a stimulus $\mathbf{s}(t)$ is encoded into a set of spike trains $\{t_j^i\}$ generated by neurons $i = 1,\ldots,N, j = 1,\ldots,N_i$ (Brillinger, 1988; Chornoboy et al., 1988; Paninski, 2004; Okatan et al., 2005; Truccolo et al., 2005) (See Stevenson et al., 2008 for a recent review). More precisely, $\mathbf{s}(t)$ is a vector of dimensionality n, which describes the history of the stimulus signal up to time t according to a suitable parametrization. For example, in Section "Potential Uses and Limitations" where we apply the GLM to retinal ganglion cell data, the vector $\mathbf{s}(t)$ contains the light intensities of the full-field flicker stimulus for the last n frames up to time t. The GLM assumes that an observed spike train $\{t_j\}$ is generated by a Poisson process with a time-varying rate $\lambda(t)$. In its simplest form the rate $\lambda(t)$ depends only on the stimulus vector $\mathbf{s}(t)$. This special case of the GLM is also known as the LNP model (Simoncelli et al., 2004). Specifically, the rate can be written as a Linear-Nonlinear cascade:

$$\lambda(t) = f\left(\mathbf{s}(t)^\top \mathbf{w}_s\right) \tag{1}$$

First, the stimulus is filtered with a linear filter $\mathbf{w}_s$ which is referred to as the *receptive field* of the neuron. Subsequently, the pointwise monotonic nonlinearity f transforms the real-valued output of the linear filtering into a non-negative instantaneous firing rate. If the current stimulus has a strong overlap with the receptive field, that is if $\mathbf{s}(t)^\top \mathbf{w}_s$ is large, this will yield a large probability of firing. If it is strongly negative, the probability of firing will be zero or close to zero.

In the classical GLM framework (McCullagh and Nelder, 1989), f^{-1} is also called "link function". For the Poisson process noise model, the link function must be both convex and log-concave in order to preserve concavity of the log-posterior (Paninski, 2004). Thus it must grow at least linearly and at most exponentially. Typical choices of this nonlinearity are the exponential or a threshold linear function,

$$f(x) = \begin{cases} 0, & \text{if} \quad x < 0 \\ x, & \text{if} \quad x \geq 0 \end{cases}.$$

As the spikes are assumed to be generated by a Poisson process, the log-likelihood of observing a spike train $\{t_j\}$ is given by

$$\begin{aligned} \log p\left(\{t_j\} \mid \mathbf{w}_s, \mathbf{s}(t)\right) &= \sum_j \log \lambda(t_j) - \int_0^T \lambda(\tau) \mathrm{d}\tau \\ &= \sum_j \log f\left(\mathbf{s}\left(t_j\right)^\top \mathbf{w}_s\right) - \int_0^T f\left(\mathbf{s}(\tau)^\top \mathbf{w}_s\right) \mathrm{d}\tau. \end{aligned} \tag{2}$$

In this simple form, the GLM ignores some commonly observed properties of spike trains, such as refractory periods or bursting effects. In order to address this problem, we want to make the firing rate $\lambda(t)$ dependent not only on the stimulus but also on the history of spikes generated by the neuron. To this purpose, an additional linear filtering term can be added into Eq. 1. For example, by convolving the spikes generated in the past with a negative-valued kernel, we can account for the refractory period. The instantaneous firing rate of the GLM then results from a superposition of two terms, a stimulus and a spike feedback term:

$$\lambda(t) = f\left(\mathbf{s}(t)^\top \mathbf{w}_s + \psi_h(t)^\top \mathbf{w}_h\right). \tag{3}$$

The m-dimensional vector $\psi_h(t)$ describes the spiking history of the neuron up to time t according to a suitable parametrization. A simple parametrization is a *spike histogram vector* whose components contain the number of spikes in a set of preceding time windows. That is, the k-th component $(\psi_h(t))_k$ contains the number of spikes in the time window $(t-\Delta_{k+1}, t-\Delta_k]$ with $\Delta_0 < \Delta_1 < \cdots < \Delta_m$. The linear weights $\mathbf{w}_h$ can then be fit empirically to model the specific dynamic properties of the neuron such as its refractory period or bursting behavior. The encoding scheme is illustrated in **Figure 1**.

Analogous to the spike feedback just described, the encoding can readily be extended to the population case, if the vector $\psi_h(t)$ for each neuron not only describes its own spiking history, but includes the spiking history of all other neurons as well. Taken together, the log-likelihood of observing the spike times $\{t_j^i\}$ for a population of $i = 1,\ldots,N$ neurons is given by

$$\begin{aligned} \log p\left(\left\{t_j^i\right\} \mid \mathbf{w}_s^i, \mathbf{w}_h^i\right) &= \sum_{i,j} \log \lambda^i\left(t_j^i\right) - \int_0^T \lambda^i(s) \mathrm{d}s \\ &= \sum_{i,j} \log f\left(\mathbf{s}\left(t_j^i\right)^\top \mathbf{w}_s^i + \psi_h\left(t_j^i\right)^\top \mathbf{w}_h^i\right) \\ &\quad - \int_0^T f\left(\mathbf{s}(\tau)^\top \mathbf{w}_s^i + \psi_h(\tau)^\top \mathbf{w}_h^i\right) \mathrm{d}\tau. \end{aligned} \tag{4}$$

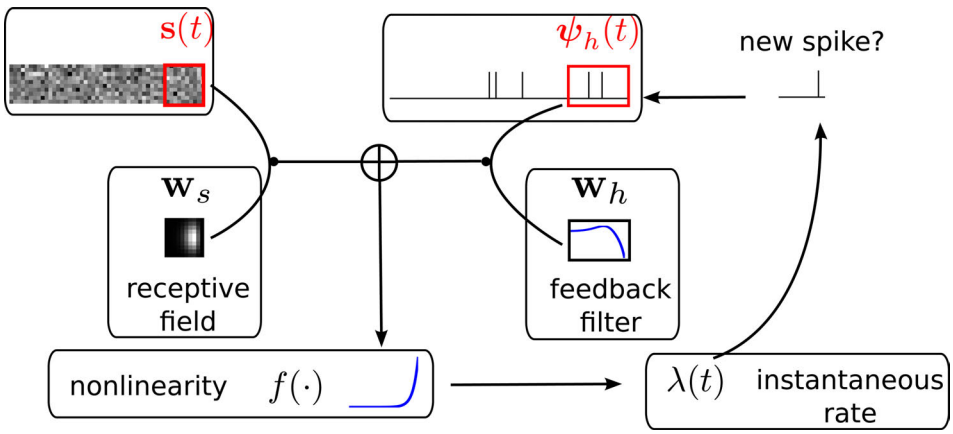


FIGURE 1 | Illustration of the generative encoding model associated with a GLM: the stimulus s(*t*) as well as the spiking history $\psi_h(t)$ are filtered with their corresponding receptive fields $\mathbf{w}_s$ and $\mathbf{w}_h$. A nonlinearity *f* is applied to the sum of the outputs to produce an instantaneous rate, which then is used to generate new spikes.

Although the likelihood factorizes over different neurons *i*, this does not imply that the neurons fire independently. In fact, every neuron can affect any other neuron *i* via the spiking history term $\psi_h(t)$. Thus, by fitting the weighting term $\mathbf{w}_h^i$ to the data we can also infer effective couplings between the neurons.

In order to evaluate Eq. 4 we have to calculate the integral $\int_0^T f(\mathbf{s}(\tau)^\top \mathbf{w}_s^i + \psi_h(\tau)^\top \mathbf{w}_h^i)\mathrm{d}\tau$ numerically. In terms of computation time, this easily becomes a dominating factor when the recording time *T* is large. Many artificial stimuli used for probing sensory neurons such as white noise can be described as piecewise constant functions. For example, the stimulus used for the retinal ganglion cells in Section "Population of Retinal Ganglion Cells" had a refresh rate of 180 Hz. In this case, the stimulus $\mathbf{s}(t)$ only changes at particular points in time. Further, if we use the spike histogram vector mentioned above to describe the spiking history of the neurons, then also $\psi_h(\tau)$ is a piecewise constant function. Thus, we can find time points $\tau_1,\ldots,\tau_z$ between which neither the stimulus nor the vector describing the spiking history changes. We call the τ_i "discretization-points". Also in cases in which the features are not piecewise constant such a discretization can be approximately obtained in a data-dependent manner, which we show in Section "Data-Dependent Discretization of the Time-Axis". By decomposing the integral over $(0, T)$ into a sum of integrals over the intervals $[\tau_k, \tau_{k+1})$ within which the integrand stays constant, the log-likelihood can be simplified to:

$$\log p\left(\{t_j^i\} \mid \mathbf{w}_s^i, \mathbf{w}_h^i\right) = \sum_{i,j} \log f\left(\mathbf{s}\left(t_j^i\right)^\top \mathbf{w}_s^i + \psi_h\left(t_j^i\right)^\top \mathbf{w}_h^i\right) - \sum_{k,i} \left(\tau_{k+1} - \tau_k\right) f\left(\mathbf{s}(\tau_k)^\top \mathbf{w}_s^i + \psi_h(\tau_k)^\top \mathbf{w}_h^i\right) \quad (5)$$

Note that $\psi_h(\tau_k)$ and $\psi_s(\tau_k)$ are constant, since the features do not change in the interval $[\tau_k, \tau_{k+1})$.

EXTENDING THE COMPUTATIONAL POWER OF GLMs

To increase the flexibility of a GLM, several extensions are possible. For example, one can add hidden variables (Kulkarni and Paninski, 2007; Nykamp, 2008) or weaken the Poisson assumption to a more general renewal process (Pillow, 2009). By adding only a few extra parameters to the model these extensions can be very effective in increasing the computational power of the neural response model. The downside of this approach is that most of these extensions do not yield a log-concave and hence unimodal posterior anymore. Another option for increasing the flexibility of the GLM which preserves the desirable property of concave log-posterior is to add more and more linearly independent parameters for the description of the stimulus and spike history that are promising candidates for improving the prediction of spike generation. For example, in addition to the original stimulus components $\mathbf{s}(t)_i$ we can also include their quadratic interactions $\mathbf{s}(t)_i \mathbf{s}(t)_j$. In this way, we can obtain an estimate of the computations of nonlinear neurons such as complex cells. This is similar to the spike-triggered covariance method (Van Steveninck and Bialek, 1988; Rieke et al., 1997; Rust et al., 2005; Pillow and Simoncelli, 2006) but more general, as we can still include the effect of the spike history. In principle, one can add arbitrary features to the description of both the stimulus as well as the spiking history. As a consequence, it is possible to approximate any arbitrary point process under mild regularity assumptions (see Daley and Vere-Jones, 2008). Like in standard least squares regression the actual merit of the Bayesian fitting procedure described in this paper is to have mechanisms for finding linear combinations of these features that provide a good description of the data. Therefore, it often makes sense to use a set of basis functions whose span defines the space of candidate functions (Pillow et al., 2005). We should choose a sufficiently rich ensemble of basis functions such that any plausible kind of stimulus or history dependence can be realized within this ensemble. We denote the feature space for the spiking history by ψ_h and the feature space for the stimulus by ψ_s. The concatenation of both feature vectors is denoted by $\psi_{s,h}$. Together we can write down the log-likelihood of observing a spike train $\{t_j^i\}_{j,i}$:

$$\log p\left(\{t_j^i\} \mid \mathbf{w}_s, \mathbf{w}_h\right) = \sum_{i,j} \log \lambda^i\left(t_j^i\right) - \sum_i \int_0^T \lambda^i(s)\mathrm{d}s \quad (6)$$

$$= \sum_{i,j} \log f\left(\psi_h\left(t_j^i\right)^\top \mathbf{w}_h^i + \psi_s\left(t_j^i\right)^\top \mathbf{w}_s^i\right) - \sum_i \int_0^T f\left(\psi_h(\tau)^\top \mathbf{w}_h^i + \psi_s(\tau)^\top \mathbf{w}_s^i\right) d\tau \quad (7)$$

DATA-DEPENDENT DISCRETIZATION OF THE TIME AXIS

If we choose the features ψ_h, ψ_s such that they do not change between distinct discretization-points τ_k, i.e., $\psi_{s,h}$ is constant in the interval $[\tau_k, \tau_{k+1})$ the likelihood can be simplified to:

$$p\left(\left\{t_j^i\right\} \mid \mathbf{w}_s, \mathbf{w}_h\right) = \sum_{i,j} \log f\left(\psi_h\left(t_j^i\right)^\top \mathbf{w}_h^i + \psi_s\left(t_j^i\right)^\top \mathbf{w}_s^i\right) - \sum_{i,k} (\tau_{k+1} - \tau_k) f\left(\psi_h(\tau_k)^\top \mathbf{w}_h^i + \psi_s(\tau_k)^\top \mathbf{w}_s^i\right) \quad (8)$$

When approximating the features by describing the spike history dependence with a piecewise constant function, this yields a finite number of discretization-points in time between which, the resulting conditional rate, given the spiking history, does not change. In order to illustrate this process, consider the following simple scenario illustrated in **Figure 2**. Suppose there is only one neuron, which receives a constant input. Accordingly, the feature describing the stimulus is constant $\psi_s(t) \equiv 1$, which appear as the last entry in the combined feature vectors $\psi_{h,s}(t)$ in the figure. The spiking history H_t up to time t is represented by two dimensions, which are approximated by piecewise constant functions, changing only at 2 and 10 ms. Note, that the time axis, labeled with time-parameter s in **Figure 2** is pointing into the past and centered at the current time point t. As long as we did not observe a spike, the feature values of the two basis functions are zero, i.e., $\psi_h(t)_1 = \psi_h(t)_2 = 0$ for $t < t_1$. Once we have observed a spike, this enters in both features via the first constant value. Hence in this example $\psi_h(t)_1 = 5$, $\psi_h(t)_2 = 1$ for $\tau_1 = t_1 \le t < \tau_2 = \tau_1 + 2$ ms. When the observed spike leaves the 2 ms window and enters the second time window of the basis functions the feature values change to $\psi_h(t)_1 = 1$, $\psi_h(t)_2 = 2$ for $\tau_2 \le t < \tau_3 = \tau_2 + 8$ ms. In order to calculate the conditional rate, we have to evaluate $f(\psi_h(t)^\top \mathbf{w}_h + \psi_s(t)^\top \mathbf{w}_s)$. For the weights in **Figure 2**, this gives the qualitative time course of the conditional rate $\lambda(t|H_t, \mathbf{s}(t))$ as depicted in **Figure 2**.

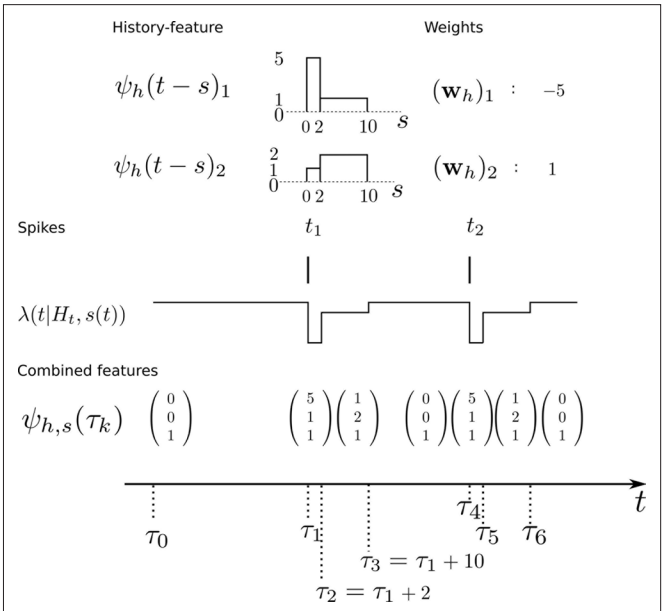


FIGURE 2 | Illustration of the data-dependent time discretization. Two spikes from one neuron have been observed at time points t_1 and t_2. Since we assume a constant input ($\mathbf{s}(t) \equiv 1$) the last entry in the combined feature vector $\psi_{h,s}(t)$ is always 1. The spiking history up to time t, denoted with H_t is described with two basis functions, $(\psi_h(t))_1$, $(\psi_h(t))_2$. Each of these could have its own discretization, but here both have the same, namely at 0, 2 and 10 ms. That is, the basis functions are approximated with a piecewise constant function with jumps at 0, 2 and 10 ms. Each spiking history feature has its own weight, as has the stimulus. Thus, the feature vector describing both, the stimulus as well as the spiking history $\psi_{s,h}(t)$ is a three-dimensional vector, changing its value at discretization-points τ_k. In each interval $[\tau_k, \tau_{k+1})$ the rate $\lambda(\tau_k|H_t, \mathbf{s}(\tau_k))$ can be calculated. In this specific case, it only assumes three different values, $\exp(\psi_{s,h}(t)^\top \mathbf{w}_{h,s}) = \exp(-27), \exp(-2), \exp(1)$, assuming that the weight for the stimulus is $\mathbf{w}_s = 1$.

USING LAPLACE PRIORS FOR BETTER REGULARIZATION

The expansion of the stimulus and the spiking history in high-dimensional feature spaces comes at the cost of having a large number of parameters to deal with. As we only have access to a limited amount of data, regularization is necessary to avoid overfitting. In the Bayesian framework, this can be done by choosing a prior distribution $p(\mathbf{w}) = p((\mathbf{w}_s, \mathbf{w}_h))$ over the linear weights $\mathbf{w}_s$ and $\mathbf{w}_h$. As these parameters enter the log-likelihood linearly, the prior distribution can be interpreted as specifying how likely we think that a particular feature is active, or necessary for explaining a typical data set. The prior distribution becomes more important as we increase the number of parameters.

Two commonly used priors are the Gaussian,

$$p(\mathbf{w}) = \frac{1}{2\sqrt{\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma} \| \mathbf{w} \|_2^2\right) = \frac{1}{2\sqrt{\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \mathbf{w}^\top \mathbf{w}\right) \quad (9)$$

and the Laplace prior,

$$p(\mathbf{w}) = \left(\frac{2}{\tau}\right)^n \exp\left(-\tau \| \mathbf{w} \|_1\right) = \prod_{k=1}^n \frac{2}{\tau} \exp\left(-\tau |w_k|\right). \quad (10)$$

Given a prior distribution, one can write down the posterior distribution,

$$p(\mathbf{w} \mid D) \propto p(\mathbf{w}) p(D \mid \mathbf{w})$$

which specifies how likely a set of weights $\mathbf{w}$ is, given the observed data D and the prior belief over the weights. The data D contains both, observed spike trains as well as stimuli.

To obtain a particular choice of parameter values a popular point estimate is MAP estimate, that is the point of maximal posterior density $\mathtt{argmax}_{\mathbf{w}}\, p(\mathbf{w}|D)$. The MAP estimate is equivalent to the maximum likelihood estimate regularized with the log-prior. As mentioned above, the use of Laplace priors can yield advantageous regularization properties (Tibshirani, 1996; Lewicki and Olshausen, 1999; Ng, 2004; Steinke et al., 2007; Mineault et al., 2009). For a sparse prior, most of the features are likely to have zero weight, but if they have a non-zero weight, the amplitude is less constrained. In order to favor sparse solutions, the direct approach would be to penalize the number of non-zero parameter entries. The number

of non-zero entries is sometimes referred to as the "L0-norm" of the parameter vector (despite the fact that it is not a proper norm). Unfortunately, finding the L0-norm regularized weights is a hard problem. Using the L1-norm however, is a useful relaxation which in some cases even gives an equivalent solution (Donoho and Stodden, 2006). The log of the Laplace prior-probability (see Eq. 10) of a given parameter vector is proportional to the L1-norm of this vector. Therefore, using a Laplace prior is equivalent to penalizing the L1-norm of the parameters. Finally using a Gaussian prior is equivalent to penalizing the L2-norm of the parameter vector (see Eq. 9).

From a practical point of view, log-concavity is another desirable property of the prior distribution as it here ensures that the posterior $p(\mathbf{w}|D) \propto p(\mathbf{w})p(D|\mathbf{w})$ is also log-concave and therefore finding the maximum of the posterior (i.e., computing the MAP estimator) is a convex optimization problem (Paninski et al., 2004). For the GLM, log-concavity and convexity of the link function f is also required to guarantee log-concavity of the posterior. Both priors, the Gaussian as well as the Laplacian are log-concave. Although the posterior is log-concave when a Laplace prior is used, calculating the MAP is still a non-trivial problem. As the Laplace prior is non-differentiable at zero, the gradient at any point containing a zero in at least one component cannot be calculated. Thus standard techniques like conjugate gradient or iterative reweighted least squares fail. For the case of a Gaussian likelihood and Laplace prior the LASSO algorithm (Tibshirani, 1996) can be used. For the case of a likelihood originating from a GLM, the posterior is differentiable in each orthant, and hence subgradients can be calculated. In our implementation, we use the algorithm of Andrew and Gao (2007).

PERFORMANCE MEASURES

After we have obtained an estimate of the parameters of a GLM, we would like to evaluate the quality of the estimate.

Prediction performance

To measure the performance of an estimate, we calculated the difference between the estimated model and the ground truth model with respect to the log-likelihoods on a test set. The test set was generated with the same weights for each trial. In this way we can assess how likely a previously unseen spike train sampled from the ground truth model is under the estimated model. The difference between the average log-likelihoods can be seen as an approximation to the Kullback–Leibler distance of the estimated model from ground truth.

$$\begin{aligned} l(\mathbf{w},\hat{\mathbf{w}}) &= \frac{1}{N}\sum_{i=1}^{N} \log p\left(D_i \,|\, \mathbf{w}\right) - \log p\left(D_i \,|\, \hat{\mathbf{w}}\right) \\ &\approx \int \log\left(\frac{p\left(D \,|\, \mathbf{w}\right)}{p\left(D \,|\, \hat{\mathbf{w}}\right)}\right) p\left(D \,|\, \mathbf{w}\right) \mathrm{d}D \\ &= D_{\mathrm{KL}}\left[p\left(\cdot \,|\, \mathbf{w}\right) \| \, p\left(\cdot \,|\, \hat{\mathbf{w}}\right)\right] \end{aligned} \tag{11}$$

Here D_i is a spike train in the i-th of N trials generated with the true weights $\mathbf{w}$ whereas the estimated weights are $\hat{\mathbf{w}}$. The more likely the spike trains are, the better is the weight estimate, which specifies the estimated model. Therefore, the difference in log-likelihood of the different models measures how well the estimated model does at predicting spike times from the ground truth model.

Mean squared error reconstruction

A different way of quantifying the performance of an estimation algorithm for synthetic data would be to check how closely the estimated parameters ($\hat{\mathbf{w}}$) match those that were put into the model as ground truth ($\mathbf{w}$). In particular for judging the quality of the reconstructed filter shapes a popular choice is to look at the mean square error between the true and estimated parameters:

$$l(\mathbf{w},\hat{\mathbf{w}}) = \sum_j \left|\mathbf{w}_j - \hat{\mathbf{w}}_j\right|^2 \tag{12}$$

APPROXIMATING THE POSTERIOR DISTRIBUTION USING EP

It has been shown that the MAP yield a good prediction performance (Pillow et al., 2008) but there are a couple of reasons why one would like to know more about the posterior than just its maximum. For example the posterior mean is known to be the optimal point estimate with respect to the mean squared error (Eq. 12). Furthermore, in many cases we are not only interested in a point estimate of the parameters, but we also want to know the dispersion of the posterior. In other words, we want to have confidence intervals indicating how strongly the parameters of a model are constrained by the observed data.

The resulting uncertainty estimate in turn can be used for optimal design (Lewi et al., 2008; Seeger, 2008), that is we can decide which stimulus to present next, in order to maximally reduce our uncertainty about the parameters. Furthermore, a distribution of the full posterior distribution gives rise to the marginal likelihood, which is the likelihood of the data under the model, without assuming specific linear filters. The marginal likelihood can be used to optimize the parameters of the prior without performing a crossvalidation (Chib, 1995; Seeger, 2008). Mathematically, the uncertainty is encoded in the dispersion of the posterior distribution over parameters $\mathbf{w}$ given observed data D:

$$p(\mathbf{w} \,|\, D) = \frac{1}{Z} p(D \,|\, \mathbf{w}) p(\mathbf{w}) \tag{13}$$

where

$$Z = \int p(D \,|\, \mathbf{w}) p(\mathbf{w}) \mathrm{d}\mathbf{w}.$$

Taken together there are strong arguments why it is useful to investigate the information conveyed by the posterior other than just the location of its maximum. The posterior is really the summary of all we can learn from the data about the given model.

Unfortunately, exact Bayesian inference (calculation of the normalization constant Z) is intractable in our case. Therefore, we are interested in finding a good approximation to the full posterior. If we can determine the posterior mean and covariance, this naturally leads to a Gaussian approximation of the posterior. Furthermore, we note that the true posterior in our case is unimodal, as both likelihood and prior are log-concave (Paninski, 2004). We employ the Expectation Propagation (EP) algorithm in order to compute a Gaussian approximation to the full posterior (Opper and Winther, 2000, 2005; Minka, 2001; Seeger, 2005) (see Nickisch

and Rasmussen, 2008 for alternative approximations schemes). The key observation is that the likelihood as well as the Laplace prior factorizes over simple terms, each of which is intrinsically one-dimensional. We have three types of factors:

$$f_1(u_i) = \exp\left(\log\left(f(u_i)\right) - \Delta\tau_i f(u_i)\right) = f(u_i)\exp\left(-\Delta\tau_i f(u_i)\right) \quad (14)$$

$$f_2(u_i) = \exp\left(-\tau_i f(u_i)\right) \quad (15)$$

$$f_3(u_i) = \exp(-\tau\,|u_i|) \quad (16)$$

where, $u_i := \psi_{s,h}(\tau_i)^{\mathrm{T}}\mathbf{w}_{s,h}$ defines the one-dimensional direction for each of these factors. $\psi_{s,h}$ and $\mathbf{w}_{s,h}$ denote the concatenation of the feature vectors describing the spiking history and the stimulus history respectively. Equation 14 corresponds to a factor or individual term in the sum of the log-likelihood (Chichilnisky, 2001) if there was a spike at τ_{i+1} and no spike in the interval (τ_i, τ_{i+1}) of length $\Delta\tau_i := (\tau_{i+1} - \tau_i)$. Equation 15 corresponds to a factor if there was no spike at time τ_{i+1}. Finally, Eq. 16 represents the Laplace terms for the prior in the product for the posterior distribution. The Expectation Propagation algorithm approximates each of those factors with a Gaussian factor:

$$f_i(u_i) \approx \exp\left(-\frac{1}{2}\pi_i u_i^2 + b_i u_i\right) \quad (17)$$

Thus, if we multiply all of these approximating factors, we obtain a Gaussian distribution, which is straightforward to normalize:

$$p(\mathbf{w}\,|\,D) \approx \frac{1}{Z}\prod_i \exp\left(-\frac{1}{2}\pi_i u_i^2 + b_i u_i\right) \quad (18)$$

$$= \frac{1}{Z}\exp\left(-\frac{1}{2}\mathbf{w}^{\top}\sum_i \pi_i \psi_{s,h}(\tau_i)\psi_{s,h}(\tau_i)^{\top}\mathbf{w} + \sum_i b_i \psi_{s,h}(\tau_i)^{\top}\mathbf{w}\right) =: Q(\mathbf{w}) \quad (19)$$

$$= \frac{1}{(2\pi)^{n/2}\det\mathbf{C}^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{w}-\mu)^{\top}\mathbf{C}^{-1}(\mathbf{w}-\mu)\right), \quad (20)$$

with

$$\mathbf{C} = \left(\sum_i \pi_i \psi_{s,h}(\tau_i)\psi_{s,h}(\tau_i)^{\top}\right)^{-1} \quad (21)$$

$$\mu = \mathbf{C}\left(\sum_i b_i \psi_{s,h}(\tau_i)\right) \quad (22)$$

The task now is to update the parameters π_i, b_i for the approximating factors such that the moments of the resulting approximation are as close to the true moments as possible. The crucial consistency equation which the EP algorithm tries to attain is given by Opper and Winther (2005):

$$D_{\mathrm{KL}}\left[f_i(u_i)\frac{Q(u_i)}{\exp\left(-\frac{1}{2}\pi_i u_i^2 + b_i u_i\right)}\middle\|\, Q(u_i)\right] \stackrel{!}{=} 0, \quad (23)$$

where D_{KL} denotes the Kullback–Leibler divergence or relative entropy. $Q(u_i)$ is the marginal Gaussian distribution in the direction of $\psi_{s,h}(\tau_i)$. It is the Gaussian distribution one obtains, when taking the complete approximation $Q(\mathbf{w})$ and projects it on $\psi_{s,h}(\tau_i)$. In other words, we require the approximation to be consistent in the sense that, if we replace the approximating factor $\exp(-1/2\pi_i u_i^2 + b_i u_i)$ with the true factor $f_i(u_i)$, the marginal moments in the direction of $\psi_{s,h}(\tau_i)$ should not change. To achieve this consistency, EP cycles through the factors and updates the parameters of each approximating factor such that Eq. 23 holds. For Eq. 23 to hold, only moments of a one-dimensional distributions have to be calculated. This can efficiently be done using numerical integration (Piessens et al., 1983). We omit the details of this updating scheme here and refer to the Appendix. The interested reader is referred to our MATLAb code and to further literature (Heskes et al., 2002; Qi et al., 2004; Seeger et al., 2007). The computational cost of EP is quadratic in the number of parameters (as the posterior covariance has to be estimated) and linear in the number of factors (in the GLM setting this is the same as the number of discretization-points) per cycle through the factors. In our simulations 30 iterations through all factors were sufficient for convergence.

Another frequently used way of approximating the posterior distribution with a Gaussian, is the so called Laplace approximation or Laplace's method (MacKay, 2003; Rasmussen and Williams, 2006; Lewi et al., 2008). A second-order Taylor expansion is calculated around the MAP. As the posterior is unimodal, the MAP can be found efficiently. Calculating the Hessian at a particular point can also be obtained analytically, given the posterior is differentiable at that point. The Laplace prior we use, however, is non-differentiable at zero. Therefore, the posterior is not differentiable at any point which contains at least one zero in one component. As we expect the MAP to assign many components zero weight, we cannot calculate the Hessian at that point. Furthermore, in a different setting it has been shown that the quality of the Laplace approximation is inferior to the one achieved by the EP approximation (Kuss and Rasmussen, 2005; Koyama and Paninski, 2009). The Laplace approximation is only sensitive to the local curvature at the point of maximal posterior density. As the EP approximation is based on moment matching it is influenced by the shape of the full posterior distribution.

POTENTIAL USES AND LIMITATIONS

In the following, we systematically compare the different point estimates, posterior mean and MAP. We vary the assumed prior distribution as well as the loss function in terms of which the performance is measured. In particular, we also investigate cases in which the assumed prior distribution differs from the "true" distribution used to generate the parameters. Finally, we also look at the possible effects of data discretization.

MAP VERSUS POSTERIOR MEAN

Tibshirani (1996) showed that for Gaussian likelihood and Laplace priors, the MAP gives sparse solutions and performs best, given the true underlying weights are sparse. If the data is assumed to be distributed according to a logistic likelihood, a similar result has been found by Ng (2004). Here, for the case of data generated by a GLM, we would like to see whether the same holds true, and also compare the MAP to the posterior mean.

To illustrate the effect of a Laplace prior when increasing the number of features in the GLM of spiking neurons, we considered the following examples. We made a series of simulations with GLM neurons for which the space of possible features was successively increased from 10 to 230 dimensions. The stimulus was Gaussian white noise discretized into 10 ms bins. The stimulus history $\mathbf{s}(t)$ was set to contain the stimulus values of the last 20 bins describing the stimulus history for a period of 200 ms. From the 20 dimensional stimulus history $\mathbf{s}(t)$ we constructed the full 230 dimensional quadratic feature space:

$$\begin{aligned}\psi_s(t) := (&s(t),\ldots,s(t-20\Delta),\\ &s(t)^2, s(t)s(t-\Delta),\ldots,s(t)s(t-20\Delta),\\ &s(t-\Delta)^2, s(t-\Delta)s(t-2\Delta),\ldots,\\ &\ldots,s(t-20\Delta)^2)\end{aligned}$$

with $\Delta = 10$ ms, similar as in Rust et al. (2005). From this basis of the 230 dimensional feature space a subset of increasing size was selected. That is, the dimensionality of the weight vector increased from 10 to 230, too. For all simulations, a GLM neuron was simulated until the likelihood consisted of 400 factors, i.e., 400 τ_k in the sum in Eq. 8 (alternatively one could also fix the time-duration of a trial or the number of spikes per trial).

We compared three different choices of priors, and use models which either had matching priors, or different ones:

1. Gaussian weights: Each weight was sampled independently from a Gaussian distribution. The variance was set to $20/\dim(\psi_s)$.
2. Laplacian weights: Each weight was sampled independently from a Laplace distribution. The variance was set to $20/\dim(\psi_s)$.
3. Sparse weights: A subset of only 10 dimensions was assigned with non-zero weights. For the assignment of the 10 weights, we draw 10 samples from a Laplace distribution with variance 2 and zero mean.

In **Figure 3** the Kullback–Leibler distance is plotted as a function of the dimensionality of the feature space for each of the generating distributions. In **Figure 3A** the weights of the ground truth model are sampled from a Gaussian distribution. Analogously, **Figure 3B** shows the results for the Laplace distribution and **Figure 3C** for the strongly sparse weights. We plot the average KL-divergence over 5000 trials ± 1 SD. As can be seen, the EP estimate for the Laplace (L1) prior performs best, if the true underlying weights are sparse. If the weights are sampled from a Laplace or a Gaussian distributions, the parameter vector of the true model is non-sparse and the L2 regularized MAP performs best. Interestingly, even for the case in which the weights are sampled from a Laplace distribution, the MAP performs best when using an L2-penalty term. Since we know the prior variance that was used to generate the weights, we did not perform a crossvalidation to set the regularization parameter, neither for the MAP estimates, nor for the posterior mean estimates (EPL1, EPL2). (Note that, in cases where the true distribution of weights is different to the prior used, it is possible that the prediction performance could be increased by picking a variance which is different to the "true" one).

In cases, in which the parameters are really drawn from the prior distribution, the posterior mean estimate can be shown to be the optimal parameter estimate, as it will minimize the mean squared error. Thus, in the two cases, in which we sampled the weights according to a Gaussian and a Laplacian distribution respectively, we expect the EP approximation to be superior to the MAP estimate in terms of the mean squared error. In the situation where the weights are actually sparse the performance is less clear, as the EP estimates assume a prior which is different to the one used to generate the weights. Therefore, it is not guaranteed in this case, that the posterior mean will be the optimal parameter estimate with respect to the mean squared error.

In general, we expect the MAP estimate to give a sparser solution than the posterior mean. If we have not seen much data, we expect the prior to dominate the posterior. In this case the maximum of the posterior will be at zero, resulting in a zero weight for the MAP. However, as the likelihood factors are not symmetric, the posterior is also not symmetric in general. Thus, even for weights for which the MAP is at zero, the probability mass is not symmetrically distributed around that maximum. Hence, the posterior mean in this

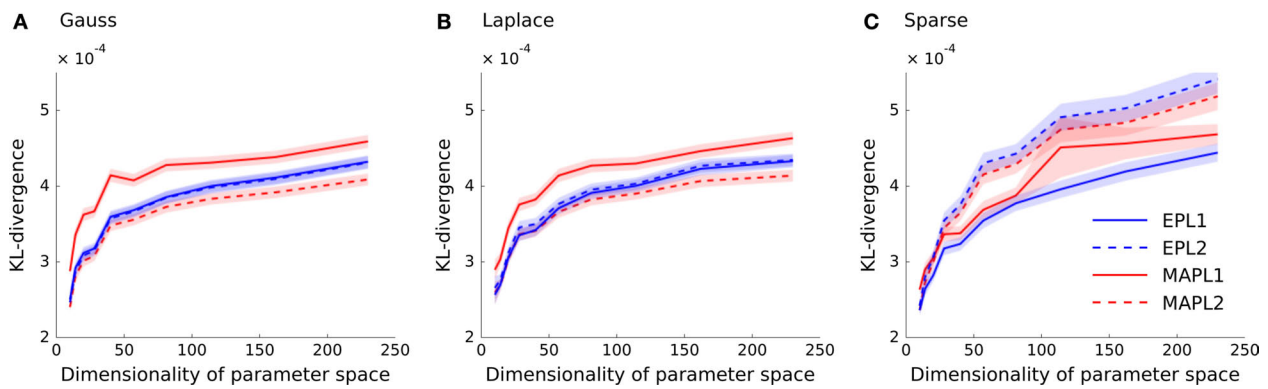


FIGURE 3 | Prediction performance in high-dimensional feature spaces of increasing size. The mean across 5000 trials of the differences in the log-likelihoods is plotted as a function of increasing stimulus dimension. The different point estimates are MAP with Laplace regularization (MAPL1, solid red), MAP with a Gaussian prior (MAPL2, dashed red) and the posterior mean approximated with EP for the Laplace (solid blue) as well as for the Gaussian prior (dashed blue). Confidence intervals indicate standard error of the mean difference. Panel **(A)** shows the performance when a Gaussian distribution is used for sampling the weights and **(B)** for a Laplace distribution. **(C)** Shows the prediction performance if the weights are actually sparse, that is the true dimensionality is constantly 10. The overall variance for the generation of weights in panels **(A)** and **(B)** were kept fix to the same value as in **(C)**.

case will be non-zero and the solution less sparse. In **Figure 4** we plotted the mean squared reconstruction error for the different estimators. As can be seen the EP approximation to the posterior mean performs better than the MAP. This is also true for the sparse setting, however the effect gets less prominent if the dimensionality of the parameter space is increased.

The quality of the different point estimates, quantified by the mean squared error and by the prediction performance are summarized in **Table 1**. To obtain a single number for the overall performance, we summed the errors for each individual dimension of parameter space (integral over each curve in **Figures 3 and 4**). The posterior mean gives a good estimate in all settings when a Laplacian prior is used. For the prediction performance the MAP with the L2 prior can lead to better results if the true prior is Gaussian or Laplacian.

BINNING AND IDENTIFIABILITY

In Section "Generalized Linear Modeling for Spiking Neurons" we specified the log-likelihood in terms of time-discretized features. This results in a binning with not necessarily equidistant discretization-points τ_j. Another popular way to simplify the log-likelihood is to bin the time axis directly. In this section we would like to illustrate the possible effects of the two discretizations by means of a simple example. For some areas, for example in the auditory cortex, the precise timing of spikes is important (Carr and Konishi, 1990; Wightman and Kistler, 1992). By binning spikes into a discrete set of bins, one might lose this precise timing. If one discretizes the time axis directly and wants to keep the precise timing, one needs to specify very small time bins. This leads to a large number of discretization-points and hence very many factors for the likelihood. Alternatively, if one discretizes the features, the discretization is adapted to the spike times and thus could lead to possibly fewer discretization-points while still achieving a high temporal resolution. However, if a lot of spike times have been observed, discretization of the basis functions for the features could lead to a time discretization which is too fine for optimization purposes. A compromise would be to adaptively add discretization-points when needed, but constrain the minimal inter discretization-point interval. In general, the discretization of the features allows one to specify the resolution and (given that resolution) produces then the minimal number of discretization-points.

To illustrate possible differences between a discretization of features versus a discretization of the time axis, we considered the following example: two GLM neurons were simulated. One of them had a stimulus filter, while the other one was only dependent on the spikes from the first neuron. The filters for the stimulus as well

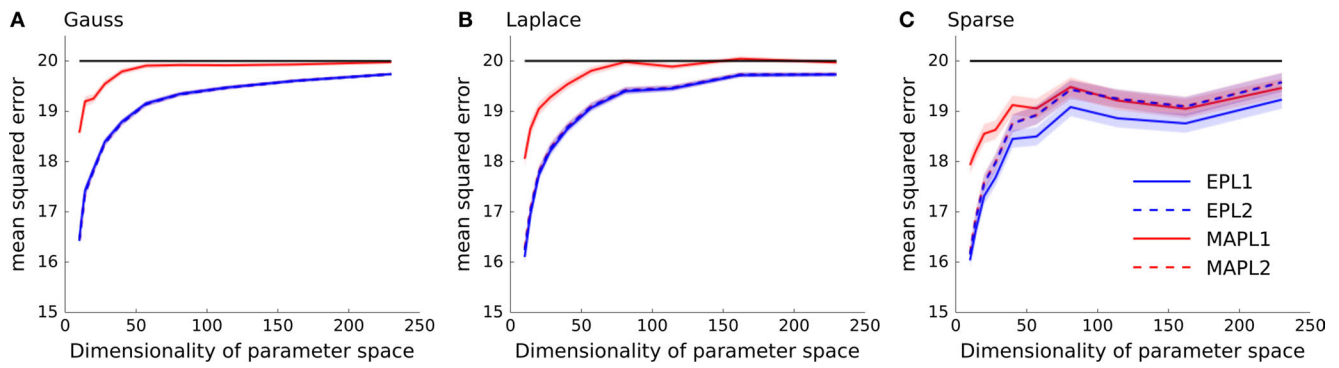


FIGURE 4 | Mean squared error as a function of increasing dimensionality of the parameter space. The same data as in **Figure 3** is plotted, but instead the performance is measured in mean squared error between the estimated weights and the true underlying weights as opposed to the differences in log-likelihoods shown in **Figure 3**. **(A)** Shows the performance if the underlying weights are sparse, in panel **(B)** a Laplace distribution is used to sample the weights and in **(C)** a Gaussian distribution is used. In each panel the mean across 5000 trials is plotted ± standard error of mean. In solid black the prior variance is plotted, which is the expected mean squared error of the constant estimator.

Table 1 | Comparison of different quality measures and point estimates. In the left table integrated KL-divergence is shown for the MAP and the posterior mean point estimates when either a Laplace or a Gaussian prior is assumed. Each row corresponds to a ground truth prior which was used to sample the weights. Each number corresponds to an integral of a curve in **Figure 3**. The right table reports the same when the mean squared error is used as a loss function. Thus, each number is the integral over one curve in **Figure 4** and therefore reports the overall performance of the different estimators. For each ground truth model and loss function the best overall estimator is colored in red.

	Integrated KL-divergence				Integrated MSE			
	MAP with		EP-mean with		MAP with		EP-mean with	
	Laplace	Gauss	Laplace	Gauss	Laplace	Gauss	Laplace	Gauss
GROUND TRUTH								
Gauss	3.93×10^{-3}	**3.39×10^{-3}**	3.532×10^{-3}	3.5×10^{-3}	195.996	186.095	186.248	**185.992**
Laplace	3.87×10^{-3}	**3.46×10^{-3}**	3.52×10^{-3}	3.58×10^{-3}	194.246	185.52	**184.99**	185.391
Sparse	3.66×10^{-3}	3.83×10^{-3}	**3.41×10^{-3}**	3.96×10^{-3}	188.698	183.685	**180.536**	183.542

as the spiking history filters are illustrated in **Figure 5** (black lines). Because the second neuron was positively coupled to the first one with a small latency, we expect it to produce spikes which have a small temporal offset with respect to the spikes of the first neuron. Intuitively, the observed spikes trains could be explained by two different settings:

1. The weights are exactly as the ones used for simulating the spike trains.
2. The second neuron is not coupled to the first neuron at all, but has the same stimulus filter as the first one, however, with a small latency. Therefore it responds to the same stimulus but at later times.

If spikes were generated deterministically, these two setting cannot be distinguished. In the noisy case, however, given a sufficient amount of data, one should be able to disentangle the two scenarios, as finding the maximum likelihood point is a convex problem. However, for finite amount of training data and in the presence of binning noise, the situation is less clear. Therefore, we sampled 3 s of spike trains and estimated the parameters from the data, once when the features are discretized and once when the time axis is discretized. The time bins were chosen such that at most one spike fell into a bin.

The estimate for the approximated posterior mean are plotted in **Figure 5**. If the features are discretized the filter could be recovered. If we discretize the time directly, we see indeed a slight shift toward the second scenario. That is, the stimulus filter for the second neuron in that case is slightly elevated, whereas the strength of the coupling filter is diminished.

POPULATION OF RETINAL GANGLION CELLS

To compare the different methods for the analysis of real data, we applied the algorithms to multi-electrode recordings of seven salamander retinal ganglion cells. Our goal was to describe the stimulus selectivity of the population by fitting a GLM with history terms and cross-neuron terms to the recorded data. We used multi-electrode recordings of salamander retinal ganglion cells generously provided by Michael J Berry II. The dataset has been published in Fairhall et al. (2006), where all recording details are described. We selected a recording of seven neurons, which had an average firing rate of 1.1 spikes per second and a minimal interspike-interval of 2.8 ms. The stimulus used in the experiments consisted of 20 min white noise full-field flicker with a refresh rate of 180 Hz. To illustrate the ability of the model to also infer population models from small data sets, we fitted the population recording to the first 2 min of the recording.

For the features describing the spiking history, we used the density function of the Γ-distribution with different parameters as basis functions:

$$f_i(t) = t^{\alpha_i - 1} \exp(-\beta_i t) \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}, \tag{24}$$

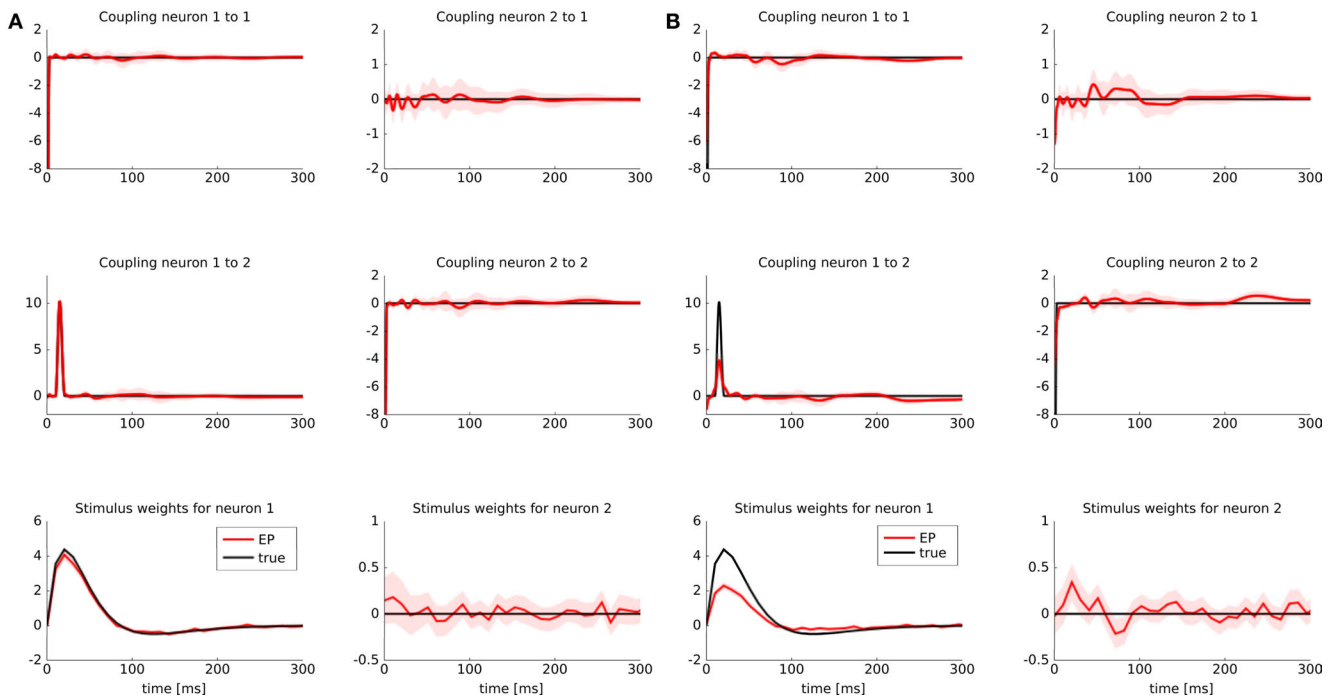


FIGURE 5 | Identifyability in the presence of binning noise. (A) Estimated filters, when the features are discretized (approximated with a piecewise constant function, see **Figure 2**). **(B)** Estimated filters when the spike times are binned. The binning was performed such that at most one spike fell into one bin. All spikes were aligned to the right hand side of their corresponding bins. When the time axis is binned directly and hence the precise timing of a spike is lost, the estimated filter for the spiking history are slightly weaker than the true ones (black), whereas the stimulus filters are slightly positive at a small latency. For the sake of readability we only plotted the approximated posterior mean (±2σ).

where the means α_i/β_i as well as the variances α_i/β_i^2 were logarithmically spaced between 1 and 700 ms and 1 and 1000 respectively (A similar basis consisting of raised cosines was also used in Pillow et al. (2005, 2008). Due to the logarithmic spacing, we have a finer resolution for small time-lags and coarser resolution for long time-lags. For example, we expect the first basis function, which has a sharp peak at zero to be mainly active or associated with the refractory period. As we discretize the basis functions rather than directly the time axis, each spike generates as many discretization-points τ_j as there are discretization-points for the basis functions (see Section "Generalized Linear Modeling for Spiking Neurons"). For the stimulus we used the same basis function set. As for the spike history dependence these functions were approximated with a piecewise constant function. The discretization for the basis-function time axis in this case was the same as for the original stimulus and therefore slightly coarser than the one for the spike history features. The basis functions are plotted in **Figure 6**.

For this setup we computed the different point estimates and posterior approximations for the weights corresponding to the features describing the spike history dependence (**Figure 7**) as well as for the weights corresponding to the stimulus filters (**Figure 8**).

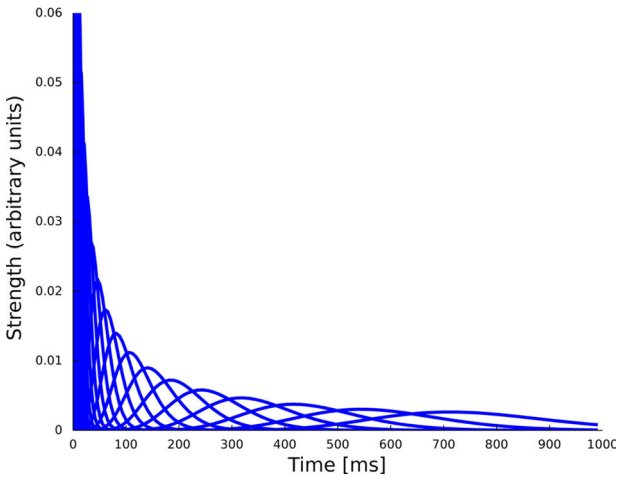


FIGURE 6 | Set of 23 basis functions to span the spiking history as well as the stimulus dependence. Each function is a density function of a Γ-distribution with different means and variances, see Eq. 24. The time axis for the features describing the spiking history was logarithmically discretized up to 1000 ms.

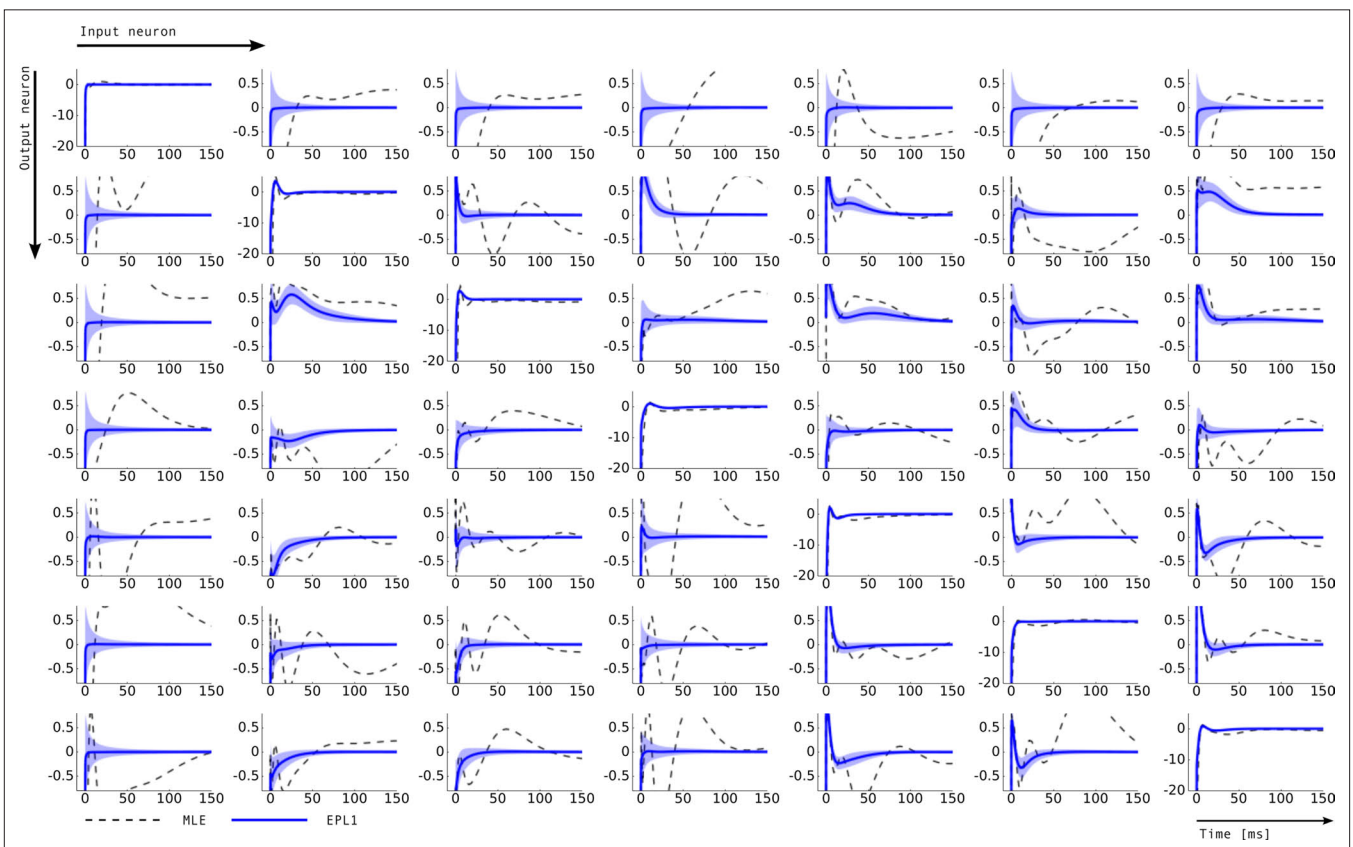


FIGURE 7 | Inferred connectivity in the network of seven retinal ganglion cells. Plotted are the induced dependencies by the weights, that is the superposition of basis functions, weighted by the inferred weights from two different estimators: maximum likelihood (MLE) and approximated posterior when a Laplace prior is used (EPL1). For the EP approximation the posterior mean together with 2 SD is plotted. Each row corresponds to one output neuron and each column corresponds to a input neuron. Thus, the entry (*i*, *j*) describes the influence of a spike of neuron *j* on the firing rate of neuron *i*. For example on the diagonal a strong negative coupling on a short time-scale can be observed, representing the refractory period of a neuron. The maximum likelihood estimate as well as thee posterior mean agree on the self-feedback but exhibit a large difference on some couplings, e.g., neuron $1 \rightarrow 4$. In general, neuron 1 seems to be less constrained than other neurons, which is also indicated by the large uncertainty intervals for the connections from and to neuron 1.

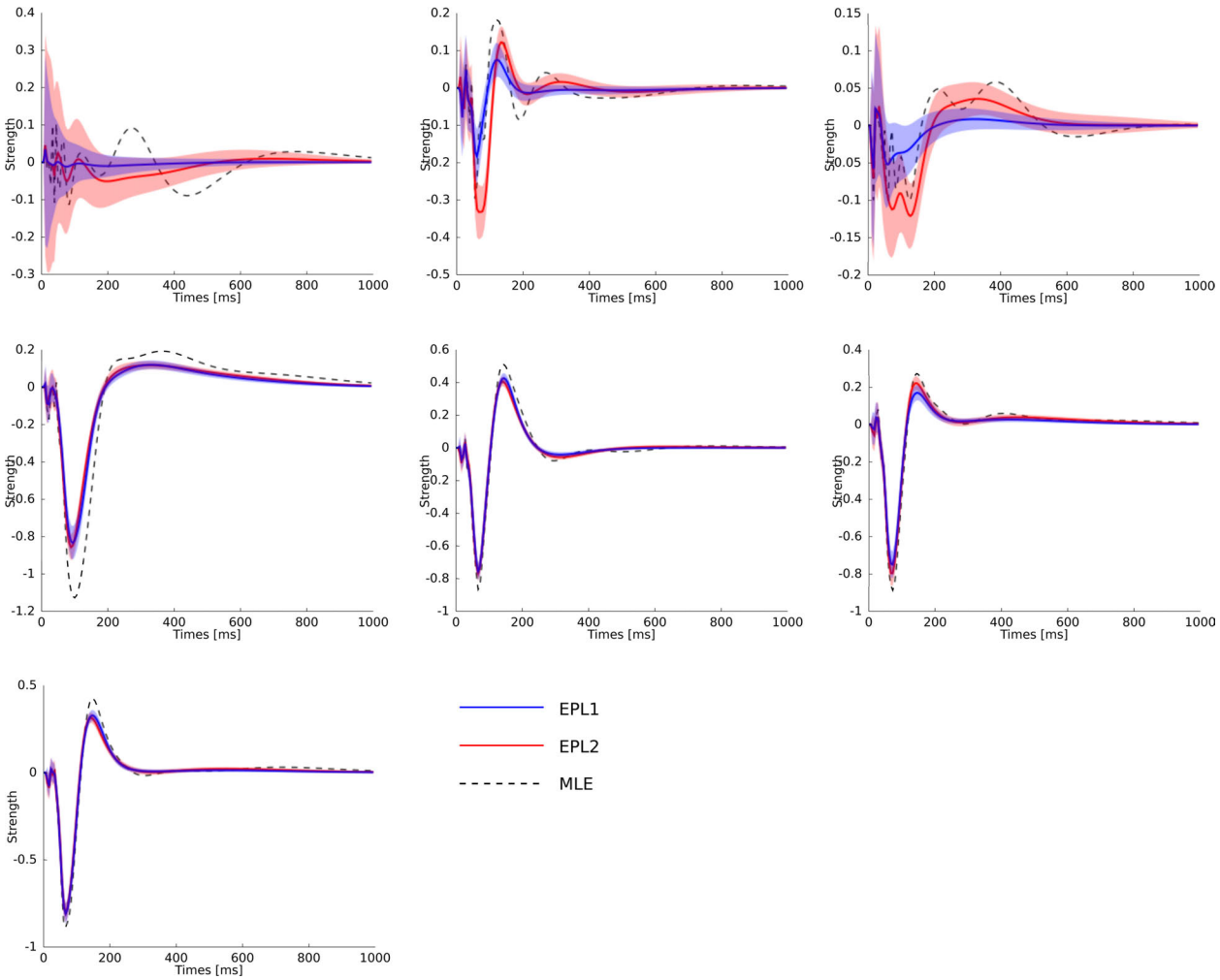


FIGURE 8 | Statistical dependence of the neural activity of seven neurons on the stimulus specified by the superposition of the basis functions plotted in Figure 6 weighted by the estimated weights. The same colors for the different estimators as in **Figure 7** are used. Additionally the posterior mean (±2σ confidence intervals) for the EP approximation with a Gaussian prior is plotted in red. Each plot corresponds to one neuron in the same order as in **Figure 7**. As can be seen, the maximum likelihood estimator is overfitting. One sees, that the posterior uncertainty for neuron 1 and also for neuron 3 are much larger as for the other neurons analog to **Figure 7**.

For training, only 2 min out of the 20 min of recording were used. Another 2 min were used for setting hyperparameters, i.e., prior variances. Given the posterior variances for each of the weights and the basis functions, we can calculate errorbars on the time course of the coupling and stimulus filters. The filters are defined as the weighted sum of the basis functions. For example, the gamma-functions f_i in Eq. 24 are weighted by the weights, corresponding to the entry in the feature vector ψ_h. Errorbars on the coupling filter $f(t)$ can then be estimated using the marginal variances:

$$\begin{aligned}\mathrm{Var}[f(t)|D] &= \mathrm{Var}\left[\mathbf{f}(t)^{\top}\mathbf{w}|D\right] \\ &= \mathbf{f}(t)^{\top}\mathrm{Cov}[\mathbf{w}|D]\mathbf{f}(t),\end{aligned} \tag{25}$$

where $\mathbf{f}(t)$ is a vector of the corresponding basis functions $f_i(t)$ and $\mathrm{Cov}[\mathbf{w}|D]$ is part of the posterior covariance matrix corresponding to the weights for the features described by $f_i(t)$. In the above equation D represents the dataset used for training, containing both, stimulus and spike trains. To illustrate this, we also plotted confidence regions of 2 SD for the coupling parameters of the population. The confidence intervals for the Gaussian approximation are plotted in red when a Laplacian prior is used and in gray when a Gaussian prior is used. Based on the confidence intervals for the coupling filters, only a few of the connections are actually significant, as can be seen in **Figure 7**. This cannot be concluded from the couplings estimated via MAP or MLE. For example, we see that connections to neuron 1 (first column in **Figure 7**) as well as connections from neuron 1 to any other neuron (first row) are underconstrained by the data, indicated by the large uncertainty for those connections compared to those for others. Consequently, the connections are set to zero by the prior and hence effectively excluded from the model. The strong negative self-feedback coupling, indicating the refractory period can be estimated with a much higher degree of certainty. We also find some significant couplings between neurons, both negatively coupled (e.g., neuron

$2 \rightarrow 5$) and positively coupled (e.g., $7 \rightarrow 2$). The maximum likelihood estimator assigns a non-zero filter to almost every coupling between neurons. The EP-mean, however, forces most of the filters to be zero. To quantify the difference in the estimated filters, we calculated the squared difference between the maximum likelihood and the EP-mean weights. This squared difference is 1.5 times larger than the average squared norm of the individual parameter vectors, which indicates that not only the absolute value of the maximum likelihood estimator is larger but also the qualitative shape is different. On the other hand the differences in prediction performance as measured by the likelihood is rather small (see **Table 2**). Thus, proximity in terms of one quality measure need not necessarily imply proximity in terms of the other as well. If the posterior uncertainty is small, the parameter vectors are much more constrained by the data and the filters estimated by the maximum likelihood estimator are closer w.r.t. the mean squared distance to the EP-mean. For example this is true for most of the stimulus filters (see **Figure 8**). In contrast, if the posterior uncertainty is rather large, for example for the stimulus filters of neuron 1 and neuron 3, the estimated weights differ more. This suggests, that we do not have sufficient information to estimate *all* parameters, but we are able to extract *some* weights from the given data.

To compare the different estimators quantitatively, we used the same performance measure as for **Figure 3**, namely the negative log-likelihood on a test set. To obtain confidence intervals on the performance measure we split the part of the dataset, neither used for training nor for validation into 16 different test sets (10%, i.e., 2 min for training, 10% for validation and 80% for testing, split into 16 sets of 1 min length). The performance values are summarized in **Table 2**. By this performance measure the EP estimate with a Laplacian prior performs significantly better than the MAP estimate with the same prior. The performance difference to the maximum likelihood estimator is not huge, this indicates, that the weights are not sufficiently constrained by 1 min slices of the data. Especially the coupling terms not well constrained as can be seen by the difference in the estimated filter by the maximum likelihood and the posterior mean, see **Figure 7**. By judging from the data, we do not know if the couplings are needed, hence excluding them from the model, i.e., setting the corresponding weights to zero, seems to be a safe choice. This can be achieved by using a strong prior distribution. The difference between a Gaussian and a Laplace prior is not large for the coupling terms (not shown), for the stimulus filters we see a small difference for the first three neurons, see **Figure 8**. Note, that in cases where there is a significant coupling between neurons, the EP and the maximum likelihood fit agree.

Table 2 | Mean prediction performance of different point estimates averaged over 16 test sets of 1 min length. As we do not have access to the true underlying model, the prediction performance here is measured in negative log-likelihood score not in differences in likelihoods.

Estimate	Negative log likelihood $\pm 2\sigma$
MLE	$3.609 \times 10^{-2} \pm 3.665 \times 10^{-4}$
MAPL1	$3.521 \times 10^{-2} \pm 2.836 \times 10^{-4}$
MAPL2	$3.497 \times 10^{-2} \pm 2.592 \times 10^{-4}$
EPL1	$\mathbf{3.461 \times 10^{-2} \pm 2.459 \times 10^{-4}}$
EPL2	$3.716 \times 10^{-2} \pm 2.973 \times 10^{-4}$

p-value for EPL1< MAPL2: 0.0219.

DISCUSSION

Bayesian inference methods are particularly useful for system identification tasks where a large number of parameters need to be estimated. By specifying a prior over the parameters a full probabilistic model is obtained that provides a principled framework for regularizing the model complexity. Furthermore, knowledge of the posterior distribution allows one both to derive point estimators that are optimized for loss functions that are suitable to the problem at hand and to quantify the uncertainty about such estimates.

A major hurdle for using a Bayesian approach is that computing the posterior distribution is often intractable. Even for numerical approximation techniques of the posterior distribution there is usually – *a priori* – no guarantee how well they work. Therefore, it is important to perform careful quality control studies if such methods are to be applied to a new estimation problem. In this paper, we presented such control studies for approximate Bayesian inference in the GLMs of spiking neurons using Expectation Propagation (EP) and compared it to standard methods like maximum likelihood and MAP estimates. Expectation Propagation provides both a posterior mean and a posterior covariance approximation. These first and second-order moments are sufficient to obtain a rough sketch of the location and dispersion of the posterior distribution. The posterior mean, in particular, can be used as a point estimator which is known to minimize the mean squared error loss. This loss function is an expedient choice if one aims at reconstructing the filter shapes. As we have shown in this work, the posterior mean estimate obtained with EP yields a smaller mean squared reconstruction error of the parameters than maximum likelihood or MAP estimation.

It should be noted, however, that the filter shapes represent statistical couplings only. Clearly, the existence of a statistical coupling does not necessarily imply the existence of a physical coupling as well. Statistical dependence could, for example, also be a consequence of common input, or other indirect couplings. In fact, it is known that noise correlations between retinal ganglion cells are mainly due to common input, and not direct synaptic couplings (Trong and Rieke, 2008). In the model an inferred coupling simply indicates that there is a dependence between the neurons which cannot be explained by the stimulus filters or the neural self-couplings.

Receptive field estimation aims at a functional characterization of neural response properties. Therefore, it is natural to compare different estimates by asking how well they can predict spike trains generated in response to new test data. Evaluating the performance of predicting a particular spike train is often based on the use of a spike train metric (Victor and Purpura, 1997), as the predicted spike trains have to be compared to the observed spike trains. In general, one wants to compare models, and not only particular spike trains, and therefore averages the prediction performance across very many samples from the two models one wants to compare.

The Bayesian framework offers a principled way to obtain an optimal point estimate which minimizes the loss function averaged across the posterior distribution. Although it is unlikely that this optimization problem can be solved analytically, one can sample weights from the posterior and then sample several spike trains

for these given weights. In other words, we can generate samples from the predictive distribution. For the prediction performance measure specified by the loss in Eq. 11, for example, an optimal point estimate would be given by those weights which on average yield the largest likelihood for the ensemble of spike trains drawn from the predictive distribution. Neither the MAP nor the posterior mean is optimal with respect to this criterion. Theoretically, the MAP is optimized for the zero-one-loss, whereas the posterior mean is optimized for the squared error loss (Lehmann and Casella, 1998). In Appendix "Bayes-Optimal Point Estimate for Average Log-Loss", we demonstrate on a simple, concrete example (estimation of the probability of a coin flip and log-loss as loss function) that an optimized predictor will perform better (on average) than the MAP estimate, irrespective of what data was observed. Clearly, this approach is only possible if one has at least an approximate model of the posterior, as we have presented here.

For a single GLM this will yield a set of parameters which are guaranteed to be optimal on average. The optimality of course only holds if the model is correct (i.e., the observed spike trains are indeed samples from a GLM), the prior is appropriately chosen, and the posterior distribution can be calculated precisely. In practice, it is not clear how justifiable each of the the three assumptions is going to be. Therefore, it is an interesting open question of how much better point estimates which are optimized using this approach will perform when compared to other optimization methods. Empirically, we observed that the posterior mean estimate obtained with EP is always better then the MAP with respect to squared error loss. With respect to the prediction error, the MAP performed slightly better than the EP posterior mean estimate if the weights were drawn from a Gaussian or Laplacian distribution, while the EP posterior mean was better than the MAP estimator if the weights were drawn from the truly sparse distribution. Of course, one could also directly use the predictive distribution as it will in general assign higher likelihood to unseen spikes than any point estimate. However, the predictive distribution cannot be described by a single GLM as it is an average over many models.

Our study also provides some insights about the effect of different kinds of prior distributions on the estimation performance. The choice of prior in the Bayesian framework offers a principled way of regularization. Here, we compared specifically a Gaussian and a Laplacian prior. While there was almost no difference in performance between the EP posterior mean estimator for the Laplacian and the Gaussian prior if the true prior was Gaussian or Laplace, the assumption of a Laplacian prior led to a substantial advantage when the true weight vectors had only a few non-zero components. This confirms the intuition that one can profit from using a Laplacian prior if one sets up a large number of candidate features of which only a few are likely to be useful in the end. Interestingly, for the MAP estimator, the use of a Laplacian prior almost always led to a substantial impairment and resulted in a relatively small improvement only w.r.t. the prediction performance if the weights were sampled from a sparse distribution for which almost all coefficients are zero.

While the posterior mean, and even more so the MAP estimator can strongly depend on the particular choice of prior distribution, this indeterminacy is a problem only if the dispersion of the posterior distribution is not taken into account appropriately. This is a strong case for the use of EP as the MAP estimator does not provide any control to what extent the result is actually constrained by the data. By also computing the posterior covariance rather than just a point estimator, we obtain confidence intervals which can serve exactly to this purpose. For the retinal ganglion cell data analyzed in Section "Population of Retinal Ganglion Cells", for example, it allowed us to distinguish between neuronal couplings, that are significant and others which were not (see neuron 1 in **Figure 7**). One can also see that whenever the confidence intervals were large, the maximum likelihood estimator deviated substantially from the Bayesian point estimators.

APPENDIX

EXPECTATION PROPAGATION WITH GAUSSIANS

Finding the posterior moments

In the following we will explain the essentials for approximating posterior distributions with a Gaussian distribution via the Expectation Propagation algorithm.

Suppose the joint distribution of a parameter vector of interest $\mathbf{w}$ and n independent observations $D = \{x_1,\ldots,x_n\}$ factors as:

$$p(D,\mathbf{w}) = p(\mathbf{w})\prod_{i=1}^{n} p\left(x_i \mid \mathbf{w}\right), \tag{A1}$$

where $p(\mathbf{w})$ is a chosen prior distribution. Further we assume, that each of the likelihood factors depends on a linear projection of the parameters $\mathbf{w}$ only. That is, a likelihood factor can be written as:

$$p\left(x_i \mid \mathbf{w}\right) = p\left(x_i \mid \psi_i^{\top}\mathbf{w}\right). \tag{A2}$$

Hence, each likelihood factor is intrinsically one-dimensional. Next, we choose an (un-normalized) Gaussian $\tilde{t}_i$ with which we would like to approximate each of those factors:

$$p\left(x_i \mid \psi_i^{\top}\mathbf{w}\right) \approx \exp\left(-\frac{1}{2}\pi_i\left(\psi_i^{\top}\mathbf{w}\right)^2 + b_i\left(\psi_i^{\top}\mathbf{w}\right)\right) \tag{A3}$$

$$= \exp\left(-\frac{1}{2}\pi_i\mathbf{w}^{\top}\left(\psi_i\psi_i^{\top}\right)\mathbf{w} + b_i\mathbf{w}^{\top}\left(\psi_i\right)\right) =: \tilde{t}_i\left(\psi_i^{\top}w\right) \tag{A4}$$

Plugging this into Eq. A1, we obtain for the approximation $Q(\mathbf{w}|D)$ to the posterior:

$$Q(\mathbf{w} \mid D) = \exp\left(-\frac{1}{2}\mathbf{w}^{\top}\left(\sum_i^n \pi_i\psi_i\psi_i^{\top}\right)\mathbf{w} + \mathbf{w}^{\top}\left(\sum_i^n b_i\psi_i\right)\right)p(\mathbf{w}) \tag{A5}$$

The prior distribution $p(\mathbf{w})$ is allowed to have two different forms. It can either be a Gaussian in which case the inverse prior covariance has to be added to the outer products of the features ψ_i. Another option is, that the prior distribution also factorizes into intrinsic one-dimensional terms. This would be the case for example, if a Laplace prior is used.

$$p(\mathbf{w}) \propto \prod_{\kappa}\exp\left(-\tau\left|\mathbf{w}_k\right|\right)$$
$$= \prod_{\kappa} p_p\left(\psi_k^{\top}\mathbf{w}\right)$$

with

$$p_p(u) = \exp(|u|), \quad \psi_k = \left(0,\ldots,0,\underset{\check{k}}{1},0,\ldots\right)^{\top} \tag{A6}$$

In order to obtain the desired Gaussian approximation to the true posterior, the problem is now to find the parameters π_i, b_i. Once these parameters are found, we get the desired approximation via Eq. A1. If the posterior consists of a single factor, then the desired parameters π_1, b_1 are easily obtained via moment matching. The moments usually have to be calculated by a numerical one-dimensional integration along the direction ψ_1. To incorporate a new factor, we fix the parameters of the first one and try to find suitable b_2,π_2 for the second factor. More precisely, we want to minimize the Kullback–Leibler distance:

$$D_{\mathrm{KL}}\left[Q\left(\mathbf{w}\,|\,\{x_1,x_2\}\right)\|Q\left(\mathbf{w}\,|\,\{x_1\}\right)p\left(x_2\,|\,\psi_2^{\top}\mathbf{w}\right)\right] \tag{A7}$$

$$= D_{\mathrm{KL}}\left[Q\left(\mathbf{w}\,|\,\{x_1\}\right)\exp\left(-\frac{1}{2}\pi_2\left(\psi_2^{\top}\mathbf{w}\right)^2+b_2\left(\psi_2^{\top}\mathbf{w}\right)\right)\right.$$
$$\left.\|Q\left(\mathbf{w}\,|\,\{x_1\}\right)p\left(x_2\,|\,\psi_2^{\top}\mathbf{w}\right)\right] \tag{A8}$$

As both Q distributions are the same and all other factors vary only along one dimension ψ_2, the only degree of freedom we have are the moments in that direction (see Seeger, 2005). Technical speaking, we can split the integration of the Kullback–Leibler distance into two parts. One over the direction ψ_2 and one in the orthogonal direction. Now, for notational simplicity, we denote $\psi_2^{\top}\mathbf{w} =: u_2$. The moments of the Gaussian side in Eq. A8 can easily be computed by looking at the exponent. Let μ_1, σ_1 be the moments of the Q distribution in the direction of ψ_2:

$$-\frac{1}{2\sigma_1}\left(u_2-\mu_1\right)^2-\frac{1}{2}\pi_2u_2^2+b_2u_2 \tag{A9}$$

$$=-\frac{1}{2}u_2^2\left(\frac{1}{\sigma_1}+\pi_2\right)+u_2\left(\frac{\mu_1}{\sigma_1}+b_2\right)-\frac{1}{2}\frac{\mu_1^2}{\sigma_1} \tag{A10}$$

Thus the moments μ_2, σ_2 are:

$$\sigma_2=\left(\frac{1}{\sigma_1}+\pi_2\right)^{-1} \tag{A11}$$

$$\mu_2=\sigma_2\left(\frac{\mu_1}{\sigma_1}+b_2\right) \tag{A12}$$

Now, these moments have to be matched with the numerically obtained ones μ_2', σ_2' of $Q(u_2|\{x_1\})\, p(x_2|u_2)$ by adjusting π_2, b_2. This can be done, by choosing the parameters according to:

$$\pi_2=\frac{1}{\sigma_2'}-\frac{1}{\sigma_1} \tag{A13}$$

$$b_2=\mu_2'\left(\frac{1}{\sigma_1}+\pi_2\right)-\frac{\mu_1}{\sigma_1} \tag{A14}$$

In this fashion we can incorporate one likelihood factor after another. This procedure is known as assumed density filtering (see Minka, 2001). The obtained approximation to the posterior depends on the order in which we incorporate the likelihood factors. The idea of Expectation Propagation is not to stop after one such sweep over the factors. EP rather tries to fulfill the consistency (Opper and Winther, 2005):

$$\frac{Q\left(\mathbf{w}\,|\,\{x_1,\ldots,x_n\}\right)}{\exp\left(-\frac{1}{2}\pi_i\left(\psi_i^{\top}\mathbf{w}\right)^2+b_i\left(\psi_i^{\top}\mathbf{w}\right)\right)}p\left(x_i\,|\,\psi_i^{\top}\mathbf{w}\right)\overset{D_{\mathrm{KL}}}{=}Q\left(\mathbf{w}\,|\,\{x_1,\ldots,x_n\}\right) \tag{A15}$$

That is, we replace one of the approximating factors with the original one and require the moments not to change. To achieve this, one usually select an arbitrary factor i and divide it out of the current approximation. The resulting distribution is called the cavity distribution $Q^{\setminus i}(\mathbf{w})$. If we call the current moments of the approximation μ, Σ, the moments in the direction of ψ_i are given by:

$$\mu_i=\psi_i^{\top}\mu \tag{A16}$$

$$\sigma_i=\psi_i^{\top}\sum\psi_i \tag{A17}$$

Thus, we have for the cavity distribution:

$$Q^{\setminus i}\left(\psi_i^{\top}\mathbf{w}\right)=\frac{Q\left(\psi_i^{\top}\mathbf{w}\,|\,\{x_1,\ldots,x_n\}\right)}{\exp\left(-\frac{1}{2}\pi_i\left(\psi_i^{\top}\mathbf{w}\right)^2+b_i\left(\psi_i^{\top}\mathbf{w}\right)\right)} \tag{A18}$$

$$=\exp\left(-\frac{1}{2}\frac{\left(u_i-\mu_i\right)^2}{\sigma_i}+\frac{1}{2}\pi_iu_i^2-b_iu_i\right) \tag{A19}$$

Where we have abbreviated $u_i := \psi_i^{\top}\mathbf{w}$. By using the same algebra as before, we have for the moments of the cavity distribution:

$$\sigma_i^{\setminus i}=\left(\frac{1}{\sigma_i}-\pi_i\right)^{-1} \tag{A20}$$

$$\mu_i^{\setminus i}=\sigma_i^{\setminus i}\left(\frac{\mu_i}{\sigma_i}-b_i\right) \tag{A21}$$

Now, we are in the same situation as before, because we want to update the parameters π_i, b_i in order to match the moments of the approximation to the ones of the cavity distribution times the original factor. These moments have to be calculated numerically, which can efficiently be computed as the involved integrals are only one-dimensional. We call these numerical moments μ_i', σ_i':

$$\mathbf{E}_{Q^{\setminus i}(u_i)p(x_i|u_i)}\left[u_i\right]=\mu_i' \tag{A22}$$

$$\mathbf{E}_{Q^{\setminus i}(u_i)p(x_i|u_i)}\left[\left(u_i-\mu_i\right)^2\right]=\sigma_i' \tag{A23}$$

The moments have to match those of the complete approximation which gives:

$$\sigma_i'\overset{!}{=}\left(\frac{1}{\sigma_i^{\setminus i}}+\pi_i^{\mathrm{new}}\right)^{-1} \tag{A24}$$

$$\mu_i' \stackrel{!}{=} \left(\frac{1}{\sigma_i^{\setminus i}} + \pi_i^{\text{new}}\right)\left(\frac{\mu_i^{\setminus i}}{\sigma_i^{\setminus i}} + b_i^{\text{new}}\right) \tag{A25}$$

$$\Rightarrow \pi_i^{\text{new}} = \frac{1}{\sigma_i'} - \frac{1}{\sigma_i^{\setminus i}} \tag{A26}$$

$$b_i^{\text{new}} = \mu_i'\left(\frac{1}{\sigma_i^{\setminus i}} + \pi_i^{\text{new}}\right) - \frac{\mu_i^{\setminus i}}{\sigma_i^{\setminus i}} \tag{A27}$$

Now we can plug in the definition of the moments of the cavity distribution to get an update for the parameters:

$$\Delta\pi_i = \pi^{\text{new}} - \pi^{\text{old}} \tag{A28}$$

$$\Delta b_i = b^{\text{new}} - b^{\text{old}} \tag{A29}$$

Together with Eq. A5 this results in a rank one update of the full distribution over the complete parameter vector $\mathbf{w}$. More precisely we have a rank one update of the covariance matrix of the approximating Gaussian as well as an update of the mean:

$$\begin{aligned}\Sigma^{\text{new}} &= \Sigma^{\text{old}} - \psi_i \psi_i^\top \frac{\Delta\pi_i}{1+\sigma_i\Delta\pi_i}\\ \mu^{\text{new}} &= \mu^{\text{old}} + \frac{\Delta b_i - \mu_i\Delta\pi_i}{1+\sigma_i\Delta\pi_i}\psi_i\end{aligned} \tag{A30}$$

Where we have used the Woodbury identity to obtain Eq. A30. To implement these equations in a numerically stable manner, one usually represents the covariance by it Cholesky decomposition:

$$\Sigma = \mathbf{L}\mathbf{L}^\top \tag{A31}$$

where $\mathbf{L}$ is a lower triangular matrix. To calculate the moments for the Laplace factors, we used a technique by Seeger (2008) as numerical integration of Laplace factors can unstable.

Marginal likelihood

The marginal Likelihood for the hyperparameters θ is defined by:

$$\begin{aligned}L(\theta,\text{Model}) &= P(D\,|\,\theta,\text{Model})\\ &= \int P(D,\mathbf{w}\,|\,\theta,\text{Model})\,d\mathbf{w}\\ &= \int P(\mathbf{w}\,|\,\theta,\text{Model})\prod_{i=1}^n P\left(x_i\,|\,\mathbf{w},\theta,\text{Model}\right)d\mathbf{w}\end{aligned} \tag{A32}$$

When considering only the parameters π_i, b_i, EP gives us an unnormalized approximation to the likelihood factors $\tilde{t}_i(\mathbf{w})$. As long as one is interested in the posterior only, this does not matter, because:

$$\begin{aligned}P(\mathbf{w}\,|\,D) &= \frac{P(\mathbf{w}\,|\,\theta,\text{Model})\prod_{i=1}^n \tilde{t}_i(\mathbf{w})C_i}{\int P(\mathbf{w}\,|\,\theta,\text{Model})\prod_{i=1}^n \tilde{t}_i(\mathbf{w})C_i\,d\mathbf{w}}\\ &= \frac{P(\mathbf{w}\,|\,\theta,\text{Model})\prod_{i=1}^n \tilde{t}_i(\mathbf{w})}{\int P(\mathbf{w}\,|\,\theta,\text{Model})\prod_{i=1}^n \tilde{t}_i(\mathbf{w})\,d\mathbf{w}}\end{aligned} \tag{A33}$$

However, if we want to approximate the marginal likelihood we need the C_i explicitly:

$$L(\theta,\text{Model}) \approx \int P(\mathbf{w}\,|\,\theta,\text{Model})\prod_{i=1}^n C_i\tilde{t}_i(\mathbf{w}\,|\,\theta,\text{Model})\,d\mathbf{w} \tag{A34}$$

The idea is to not only match the moments but the 0th moments as well. We require the expectation of $P(x_i\,|\,\mathbf{w})$ and $\tilde{t}_i(\mathbf{w})$ under $Q^{\setminus i}(\mathbf{w})$ to be the same for all i, from which we obtain:

$$Z_i := E_{Q^{\setminus i}}\left[P\left(x_i\,|\,\mathbf{w}\right)\right] = E_{Q^{\setminus i}}\left[C_i\tilde{t}_i(\mathbf{w})\right] = \underbrace{C_i E_{Q^{\setminus i}}\left[\tilde{t}_i(\mathbf{w})\right]}_{=:\tilde{Z}_i} \tag{A35}$$

For the $\tilde{Z}_i$ we have:

$$\begin{aligned}\tilde{Z}_i &= \frac{1}{\sqrt{2\pi\sigma_{\setminus i}}}\int \exp\left(-\frac{1}{2}\pi_i u_i^2 + b_i u_i\right)\exp\left(-\frac{1}{2}\frac{\left(u_i-\mu_{\setminus i}\right)^2}{\sigma_{\setminus i}}\right)du_i\\ &= \frac{1}{\sqrt{2\pi\sigma_{\setminus i}}}\int \exp\left(-\frac{1}{2}\frac{\left(u_i - \left(\pi_i+\sigma_{\setminus i}^{-1}\right)^{-1}\left(b_i+\sigma_{\setminus i}^{-1}\mu_{\setminus i}\right)\right)^2}{\left(\pi_i+\sigma_{\setminus i}^{-1}\right)^{-1}}\right)du_i\\ &\qquad .\exp\left(-\frac{1}{2}\mu_{\setminus i}^2\sigma_{\setminus i}^{-1} + \frac{1}{2}\left(\pi_i+\sigma_{\setminus i}^{-1}\right)^{-1}\left(b_i+\sigma_{\setminus i}^{-1}\mu_{\setminus i}\right)^2\right)\\ &= \frac{\sqrt{2\pi\left(\pi_i+\sigma_{\setminus i}^{-1}\right)}}{\sqrt{2\pi\sigma_{\setminus i}}}\exp\left(-\frac{1}{2}\frac{\left(\sigma_{\setminus i}b_i^2 + 2\mu_{\setminus i}b_i - \pi_i\mu_{\setminus i}^2\right)}{\pi_i\sigma_{\setminus i}+1}\right)\end{aligned} \tag{A36}$$

Therefore, we have for the marginal likelihood:

$$\begin{aligned}\log C_i &= \log Z_i - \log\tilde{Z}_i \Rightarrow \log L(\theta,\text{Model})\\ &= \log\int\exp\left(\sum_{i=1}^n \log C_i - \frac{1}{2}\pi_i\mathbf{w}^\top\psi_i\psi_i^\top\mathbf{w} + b_i\psi_i^\top\mathbf{w}\right)d\mathbf{w}\\ &= \sum_{i=1}^n \log C_i + (2\pi)^{\frac{D}{2}}\,|\,\Sigma_p\,|^{\frac{1}{2}}\exp\left(\frac{1}{2}\mu_p^\top\Sigma_p^{-1}\mu_p\right)\quad\text{where}\\ \Sigma_p &= \left(\sum_i \pi_i\psi_i\psi_i^\top\right)^{-1}\\ \mu_p &= \Sigma_p\left(\sum_i b_i\psi_i\right)\end{aligned} \tag{A37}$$

One can also calculate gradients of the marginal likelihood with respect to hyperparameters (see Seeger, 2005).

MATLAB toolbox

Along with the paper we publish a MATLAB toolbox for inference in a generalized linear models `http://www.kyb.tuebingen.mpg.de/bethge/code/glmtoolbox/`. The code provides routines for:

1. Sampling spike trains from a GLM
2. Calculation of different point estimators: maximum likelihood, MAP, posterior mean
3. Approximation of the posterior covariance via EP.

Either a Laplacian or a Gaussian prior can be specified. For the Gaussian prior an arbitrary covariance matrix is allowed.

BAYES-OPTIMAL POINT ESTIMATE FOR AVERAGE LOG-LOSS

In the following we consider a simple example of a coin flip to illustrate the potential benefit of an optimized point estimate for the expected loss after having observed the data. Let x be Bernoulli distributed with unkown parameter $\theta \in [0, 1]$. If we observe N data points $x_i \in \{0, 1\}$ with k ones and assume a uniform prior over $\theta \sim U[0, 1]$, we can compute the posterior distribution for θ:

$$p(\theta|\{x_i\}) = \frac{1}{Z}\prod_i \theta^{x_i}(1-\theta)^{1-x_i}$$

$$Z = \int_0^1 \prod_i \theta^{x_i}(1-\theta)^{1-x_i}\,d\theta,$$

which is a Beta-distribution with parameters $\alpha = k + 1, \beta = N + 1$. The posterior mean is given by $\mu = (k + 1)/(N + 2)$. We define the average log-loss to be:

$$\text{loss}(\theta, \hat{\theta}) = \sum_{x=0,1} -p(x|\theta)\log p(x|\hat{\theta})$$

Then, we can calculate the expected average log-loss after having observed the data $\{x_i\}$:

$$\begin{aligned} F(\hat{\theta}) &=: \int\left[\sum_{x=0,1} -p(x|\theta)\log p(x|\hat{\theta})\right] p(\theta|\{x_i\})\,d\theta \\ &:= \int\left[-\log(\hat{\theta})\theta - \log(1-\hat{\theta})(1-\theta)\right] p(\theta|\{x_i\})\,d\theta \\ &= -\mu\log(\hat{\theta}) - (1-\mu)\log(1-\hat{\theta}) \end{aligned}$$

F can now be minimized with respect to the point estimate $\hat{\theta}$. The derivative with respect to $\hat{\theta}$ is given by:

$$0 \overset{!}{=} \frac{dF}{d\hat{\theta}} = -\frac{\mu}{\hat{\theta}} + \frac{1-\mu}{1-\theta} \quad \Rightarrow \hat{\theta} = \mu$$

Therefore the posterior mean optimizes the expected prediction performance as measured by the average log-loss. We can also calculate the difference in expected performance between the posterior mean and the MAP, which is given by $\theta_{\text{MAP}} = k/N$. The difference in expected performance is given by:

$$\begin{aligned} F(\theta_{\text{MAP}}) - F(\mu) &= -\mu\log(\theta_{\text{MAP}}) - (1-\mu)\log(1-\theta_{\text{MAP}}) \\ &\quad + \mu\log(\mu) + (1-\mu)\log(1-\mu) \\ &= \mu\log\left(\frac{\mu}{\theta_{\text{MAP}}}\right) + (1-\mu)\log\left(\frac{1-\mu}{1-\theta_{\text{MAP}}}\right) \end{aligned}$$

The difference in expected log-loss is the Kullback–Leibler divergence between the distribution corresponding to the optimized estimate (the posterior mean) and the distribution induced by the MAP estimate. As the Kullback–Leibler divergence is always non-negative, this shows that the loss incurred by the MAP estimate is greater than the optimized estimate, irrespective of the data (k) that was observed. In the extreme cases, i.e., $k = 0$ or $k = N$, the difference becomes infinite. This simple example shows that, in principle, an extra gain in performance can be achieved by optimizing the parameters for the expected performance over the posterior distribution.

ACKNOWLEDGMENTS

This research was funded by the German Ministry of Education, Science, Research and Technology through the Bernstein award to Matthias Bethge (BMBF, FKZ:01GQ0601), and the Max Planck Society. We would like to thank Michael Berry for generously providing us with multi-electrode recording data to illustrate the method. Additionally, we thank Fabian Sinz and Philipp Berens for discussions and comments on the manuscript. A brief version of portions of this research were previously published as "Bayesian Inference for Spiking Neuron Models with a Sparsity Prior" in the proceedings "Advances in Neural Information Processing Systems 20 (2008)" (Gerwinn et al., 2008).

REFERENCES

Andrew, G., and Gao, J. (2007). "Scalable training of L1-regularized log-linear models," in *ICML '07: Proceedings of the 24th International Conference on Machine Learning*, Corvalis, OR (New York, NY: ACM), 33–40.

Borisyuk, G. N., Borisyuk, R. M., Kirillov, A. B., Kovalenko, E. I., and Kryukov, V. I. (1985). A new statistical method for identifying interconnections between neuronal network elements. *Biol. Cybern.* 52, 301–306.

Brillinger, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cybern.* 59, 189–200.

Carr, C. E., and Konishi, M. (1990). A circuit for detection of interaural time differences in the brain stem of the barn owl. *J. Neurosci.* 10, 3227–3246.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *J. Am. Stat. Assoc.* 90, 1313–1321.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network* 12, 199–213.

Chornoboy, E. S., Schramm, L. P., and Karr, A. F. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybern.* 59, 265–275.

Daley, D. J., and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*. New York: Springer.

Donoho, D. L., and Stodden, V. (2006). "Breakdown point of model selection when the number of variables exceeds the number of observations," in *Proceedings of the International Joint Conference on Neural Networks* (Piscataway: IEEE), 16–21.

Fairhall, A. L., Burlingame, C. A., Narasimhan, R., Harris, R. A., Puchalla, J. L., and Berry, M. J. (2006). Selectivity for multiple stimulus features in retinal ganglion cells. *J. Neurophysiol.* 96, 2724–2738.

Gerwinn, S., Macke, J., Seeger, M., and Bethge, M. (2008). "Bayesian inference for spiking neuron models with a sparsity prior," in *Advances in Neural Information Processing Systems 20*, eds J. C. Platt, D. Koller, Y. Singer and S. Roweis (Cambridge, MA: MIT Press), 529–536.

Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., and Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature* 424, 552–556.

Heskes, T., Zoeter, O., Darwiche, A., and Friedman, N. (2002). "Expectation propagation for approximate inference," in *Proceedings UAI-2002*, eds A. Darwiche and N. Friedman (San Francisco: Morgan Kaufmann), 216–233.

Koyama, S., and Paninski, L. (2009). Efficient computation of the maximum *a posteriori* path and parameter estimation in integrate-and-fire and more general state-space models. *J. Comput. Neurosci.* doi: 10.1007/s10827-009-0150-x

Kulkarni, J. E., and Paninski, L. (2007). Common-input models for multiple neural spike-train data. *Network* 18, 375–407.

Kuss, M., and Rasmussen, C. E. (2005). Assessing approximate inference for binary Gaussian process classification. *J. Mach. Learn. Res.* 6, 1679–1704.

Lehmann, E. L., and Casella, G. (1998). *Theory of Point Estimation*. New York: Springer Verlag.

Lewi, J., Butera, R., and Paninski, L. (2008). Sequential optimal design of neurophysiology experiments. *Neural. Comput.* 21, 619–687.

Lewicki, M. S., and Olshausen, B. A. (1999). Probabilistic framework for the adaptation and comparison of image codes. *J. Opt. Soc. Am.* 16, 1587–1601.

MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. New York: Chapman and Hall/CRC.

Mineault, P. J., Barthelmé, S., and Pack, C. C. (2009). Improved classification images with sparse priors in a smooth basis. *J. Vis.* 9, 10–17.

Minka, T. P. (2001). *A Family of Algorithms for Approximate Bayesian Inference*. PhD thesis, Massachusetts Institute of Technology, Cambridge.

Ng, A. Y. (2004). "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proceedings of the Twenty-first International Conference on Machine Learning* (New York, NY: ACM), 78–85.

Nickisch, H., and Rasmussen, C. E. (2008). Approximations for binary gaussian process classification. *J. Mach. Learn. Res.* 9, 2035–2078.

Nykamp, D. Q. (2008). Pinpointing connectivity despite hidden nodes within stimulus-driven networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 78, 021902.

Okatan, M., Wilson, M. A., and Brown, E. N. (2005). Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural. Comput.* 17, 1927–1961.

Opper, M., and Winther, O. (2000). Gaussian processes for classification: mean-field algorithms. *Neural. Comput.* 12, 2655–2684.

Opper, M., and Winther, O. (2005). Expectation consistent approximate inference. *J. Mach. Learn. Res.* 6, 2177–2204.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15, 243–262.

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

Piessens, R., de Doncker-Kapenga, E., Uberhuber, C. W., and Kahaner, D. K. (1983). *QUADPACK: A Subroutine Package for Automatic Integration*. Berlin: Springer.

Pillow, J. (2009). "Time-rescaling methods for the estimation and assessment of non-Poisson neural encoding models," in *Advances in Neural Information Processing Systems 22*, eds Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams and A. Culotta (Cambridge, MA: MIT Press), 1473–1481.

Pillow, J., 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, 11003–11013.

Pillow, J., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999.

Pillow, J., and Simoncelli, E. P. (2006). Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis. *J. Vis.* 6, 414–428.

Qi, Y. A., Minka, T. P., Picard, R. W., and Ghahramani, Z. (2004). "Predictive automatic relevance determination by expectation propagation," in *Proceedings of the Twenty-first International Conference on Machine Learning*. New York, NY: ACM.

Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

Rieke, F., Warland, D., van Steveninck, R. R., and Bialek, W. (1997). *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press.

Rust, N. C., Schwartz, O., Movshon, J. A., and Simoncelli, E. P. (2005). Spatiotemporal elements of macaque v1 receptive fields. *Neuron* 46, 945–956.

Seeger, M. (2005). *Expectation Propagation for Exponential Families*. Technical Report, University of California at Berkeley.

Seeger, M. (2008). Bayesian inference and optimal design for the sparse linear model. *J. Mach. Learn. Res.* 9, 759–813.

Seeger, M., Gerwinn, S., and Bethge, M. (2007). Bayesian inference for sparse generalized linear models. *Lect. Notes Comput. Sci.* 4701, 298.

Simoncelli, E., Paninski, L., and Pillow, J. (2004). *The Cognitive Neurosciences*, Chapter 23. Cambridge, MA: MIT Press, 327–338.

Steinke, F., Seeger, M., and Tsuda, K. (2007). Experimental design for efficient identification of gene regulatory networks using sparse Bayesian models. *BMC Syst. Biol.* 1, 51.

Stevenson, I. H., Rebesco, J. M., Miller, L. E., and Körding, K. P. (2008). Inferring functional connections between neurons. *Curr. Opin. Neurobiol.* 18, 582–588.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* 58, 267–288.

Trong, P. K., and Rieke, F. (2008). Origin of correlated activity between parasol retinal ganglion cells. *Nat. Neurosci.* 11, 1343–1351.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J. Neurophysiol.* 93, 1074–1089.

Truccolo, W., Hochberg, L. R., and Donoghue, J. P. (2010). Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nat. Neurosci.* 13, 105–111.

Van Steveninck, R. D. R., and Bialek, W. (1988). Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc. R. Soc. Lond., B, Biol. Sci.* 234, 379–414.

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

Wightman, F. L., and Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* 91, 1648–1661.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 December 2009; paper pending published: 14 January 2010; accepted: 23 April 2010; published online: 28 May 2010.

*Citation: Gerwinn S, Macke JH and Bethge M (2010) Bayesian inference for generalized linear models for spiking neurons. Front. Comput. Neurosci. **4**:12. doi: 10.3389/fncom.2010.00012*

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 06 October 2010
doi:10.3389/fncom.2010.00126

Demixing population activity in higher cortical areas

Christian K. Machens*

Group for Neural Theory, INSERM Unité 960, Département d'Etudes Cognitives, École Normale Supérieure, Paris, France

Edited by:
Jakob H. Macke, University College London, UK

Reviewed by:
Byron Yu, Carnegie Mellon University, USA
Satish Iyengar, University of Pittsburgh, USA

****Correspondence:***
Christian K. Machens, Département d'Etudes Cognitives, École Normale Supérieure, Paris, France.
e-mail: christian.machens@ens.fr

Neural responses in higher cortical areas often display a baffling complexity. In animals performing behavioral tasks, single neurons will typically encode several parameters simultaneously, such as stimuli, rewards, decisions, etc. When dealing with this large heterogeneity of responses, cells are conventionally classified into separate response categories using various statistical tools. However, this classical approach usually fails to account for the distributed nature of representations in higher cortical areas. Alternatively, principal component analysis (PCA) or related techniques can be employed to reduce the complexity of a data set while retaining the distributional aspect of the population activity. These methods, however, fail to explicitly extract the task parameters from the neural responses. Here we suggest a coordinate transformation that seeks to ameliorate these problems by combining the advantages of both methods. Our basic insight is that variance in neural firing rates can have different origins (such as changes in a stimulus, a reward, or the passage of time), and that, instead of lumping them together, as PCA does, we need to treat these sources separately. We present a method that seeks an orthogonal coordinate transformation such that the variance captured from different sources falls into orthogonal subspaces and is maximized within these subspaces. Using simulated examples, we show how this approach can be used to demix heterogeneous neural responses. Our method may help to lift the fog of response heterogeneity in higher cortical areas.

Keywords: prefrontal cortex, population code, principal component analysis, multi-electrode recordings, blind source separation

INTRODUCTION

Higher-order cortical areas such as the prefrontal cortex receive and integrate information from many other areas of the brain. The activity of neurons in these areas often reflects this mix of influences. Typical neural responses are shaped both by the internal dynamics of these systems as well as by various external events such as the perception of a stimulus or a reward (Rao et al., 1997; Romo et al., 1999; Brody et al., 2003; Averbeck et al., 2006; Feierstein et al., 2006; Gold and Shadlen, 2007; Seo et al., 2009). As a result, neural responses are extremely complex and heterogeneous, even in animals that are performing relatively facile tasks such as simple stimulus–response associations (Gold and Shadlen, 2007).

To make sense of these data, researchers typically seek to relate the firing rate of a neuron to one of various experimentally controlled task parameters, such as a sensory stimulus, a reward, or a decision that an animal takes. To this end, a number of statistical tools are exploited such as regression (Romo et al., 2002; Brody et al., 2003; Sugrue et al., 2004; Kiani and Shadlen, 2009; Seo et al., 2009), signal detection theory (Feierstein et al., 2006; Kepecs et al., 2008), or discriminant analysis (Rao et al., 1997). The population response is then characterized by quantifying how each neuron in the population responds to a particular task parameter. Subsequently, neurons can be attributed to different (possibly overlapping) response categories, and population responses can be constructed by averaging the time-varying firing rates within such a category.

This classical, single-cell based approach to electrophysiological population data has been quite successful in clarifying what information neurons in higher-order cortical areas represent. However, the approach rarely succeeds in giving a complete account of the recorded activity on the population level. For instance, many interesting features of the population response may go unnoticed if they have not been explicitly looked for. Furthermore, the strongly distributional nature of the population response, in which individual neurons can be responsive to several task parameters at once, is often left in the shadows.

Principal component analysis (PCA) and other dimensionality reduction techniques seek to alleviate these problems by providing methods that summarize neural activity at the population level (Nicolelis et al., 1995; Friedrich and Laurent, 2001; Zacksenhouse and Nemets, 2008; Yu et al., 2009; Machens et al., 2010). However, such "unsupervised" techniques will usually neglect information about the relevant task variables. While the methods do provide a succinct and complete description of the population response, the description may yield only limited insights into how different task parameters are represented in the population of neurons.

In this paper, we propose an exploratory data analysis method that seeks to maintain the major benefits of PCA while also extracting the relevant task variables from the data. The primary goal of our method is to improve on dimensionality reduction techniques by explicitly taking knowledge about task parameters into account. The method has previously been applied to data from the prefrontal cortex to separate stimulus- from time-related activities (Machens et al., 2010). Here, we describe the method in greater detail, derive it from first principles, investigate its performance under noise, and generalize it to more than two task parameters. Our hope is that this method provides a better visualization of a given data set, thereby

yielding new insights into the function of higher-order areas. We will first explain the main ideas in the context of a simple example, then show how these ideas can be generalized, and finally discuss some caveats and limitations of our approach.

RESULTS

RESPONSE HETEROGENEITY THROUGH LINEAR MIXING

Recordings from higher-order areas in awake behaving animals often yield a large variety of neural responses (see e.g., Miller, 1999; Churchland and Shenoy, 2007; Jun et al., 2010; Machens et al., 2010). These observations at the level of individual cells could imply a complicated and intricate response at the population level for which a simplified description does not exist. Alternatively, the large heterogeneity of responses may be the result of a simple mixing procedure. For instance, response variety can come about if the responses of individual neurons are random, linear mixtures of a few generic response components (see e.g., Eliasmith and Anderson, 2003).

To illustrate this insight, we will construct a simple toy model. Imagine an animal which performs a two-alternative-forced choice task (Newsome et al., 1989; Uchida and Mainen, 2003). In each trial of such a task, the animal receives a sensory stimulus, s, and then makes a binary decision, d, based on whether s falls into one of two response categories. If the animal decides correctly, it receives a reward. We will assume that the activity of the neurons in our toy model depends only on the stimulus s and the decision d.

To obtain response heterogeneity, we construct the response of each neuron as a random, linear mixture of two underlying response components, one that represents the stimulus, $z_1(t,s)$, and one that represents the decision, $z_2(t,d)$, see **Figure 1A**. The time-varying firing rate of neuron i is then given by

$$r_i(t,s,d) = a_{i1} z_1(t,s) + a_{i2} z_2(t,d) + c_i + \eta_i(t). \quad (1)$$

Here, the parameters a_{i1} and a_{i2} are the mixing coefficients of the neuron, the bias parameter c_i describes a constant offset, and the term $\eta_i(t)$ denotes additive, white noise. We assume that the noise of different neurons can be correlated so that

$$\langle \eta_i(t)\eta_j(t+\tau)\rangle_t = \delta(\tau) H_{ij}, \quad (2)$$

where the angular brackets denote averaging over time, and H_{ij} is the noise covariance between neuron i and j. We will assume that there are N neurons and, for notational compactness, we will assemble their activities into one large vector, $\mathbf{r}(t,s,d) = (r_1(t,s,d),\ldots,r_N(t,s,d))^T$. After doing the same for the mixing coefficients, the constant offset, and the noise, we can write equivalently,

$$\mathbf{r}(t,s,d) = \mathbf{a}_1 z_1(t,s) + \mathbf{a}_2 z_2(t,d) + \mathbf{c} + \mathbf{n}(t). \quad (3)$$

Without loss of generality, we can furthermore assume that the mixing coefficients are normalized so that $\mathbf{a}_i^T \mathbf{a}_i = 1$ for $i \in \{1,2\}$. Since we assume that the mixing coefficients are drawn at random, and independently of each other, the first and second coefficient will be uncorrelated, so that on average, $\mathbf{a}_1^T \mathbf{a}_2 = 0$, implying that $\mathbf{a}_1$ and $\mathbf{a}_2$ are approximately orthogonal.

With this formulation, individual neural responses mix information about the stimulus s and the decision d, leading to a variety of responses, as shown in **Figure 1B**. While with only two underlying components, the overall heterogeneity of responses remains limited, the response heterogeneity increases strongly when more components are allowed (see **Figures 3A,B** for an example with three components).

PRINCIPAL COMPONENT ANALYSIS FAILS TO DEMIX THE RESPONSES

The standard approach to deal with such data sets is to sort cells into categories. In our example, this approach may yield two overlapping categories of cells, one for cells that respond to the stimulus and one for cells that respond to the decision. While this approach tracks down which variables are represented in the population, it will fail to quantify the exact nature of the population activity, such as the precise co-evolution of the neural population activity over time.

A common approach to address these types of problems are dimensionality reduction methods such as PCA (Nicolelis et al., 1995; Friedrich and Laurent, 2001; Hastie et al., 2001; Zacksenhouse and Nemets, 2008; Machens et al., 2010). The main aim of PCA is to find a new coordinate system in which the data can be represented in a more succinct and compact fashion. In our toy example, even though we may have many neurons with different responses ($N = 50$ in **Figure 1**, with five examples shown in **Figure 1B**), the activity of each neuron can be represented by a linear combination of only two components. In the N-dimensional space of neural activities, the two components, $z_1(t,s)$ and $z_2(t,d)$, can be viewed as two coordinates of a coordinate system whose axes are given by the vectors of mixing coefficients, $\mathbf{a}_1$ and $\mathbf{a}_2$. Since the first two coordinates capture all the relevant information, the components live in a two-dimensional subspace. Using PCA, we can retrieve the two-dimensional subspace from the data. While the method allows us to reduce the dimensionality and complexity of the data dramatically, PCA will in general only retrieve the two-dimensional subspace, but not the original coordinates, $z_1(t,s)$ and $z_2(t,d)$.

To see this, we will briefly review PCA and show what it does to the data from our toy model. PCA commences by computing the covariances of the firing rates between all pairwise combination of neurons. Let us define the mean firing rate of neuron i as the average number of spikes that this neuron emits, so that

$$r_i = \frac{1}{M_t M_s M_d} \sum_{t=1}^{M_t} \sum_{s=1}^{M_s} \sum_{d=1}^{M_d} r_i(t,s,d) \quad (4)$$

$$=: \langle r_i(t,s,d)\rangle_{t,s,d}. \quad (5)$$

We will use the angular brackets in the second line as a shorthand for averaging. The variables to be averaged over are indicated as subscript on the right bracket. Here, the average runs over all time points t, all stimuli s, and all decisions d. For the vector of mean firing rates we write $\mathbf{r} = (r_1,\ldots,r_N)^T$.

The covariance matrix of the data summarizes the second-order statistics of the data set,

$$C = \left\langle \big(\mathbf{r}(t,s,d) - \mathbf{r}\big)\big(\mathbf{r}(t,s,d) - \mathbf{r}\big)^T \right\rangle_{t,s,d}, \quad (6)$$

and has size $N \times N$ where N is the number of neurons in the data set. Given the covariance matrix, we can compute the firing rate variance that falls along arbitrary directions in state space. For instance, the variance captured by a coordinate axis given by a normalized vector $\mathbf{u}$ is simply $L = \mathbf{u}^T C \mathbf{u}$. We can then look for the axis

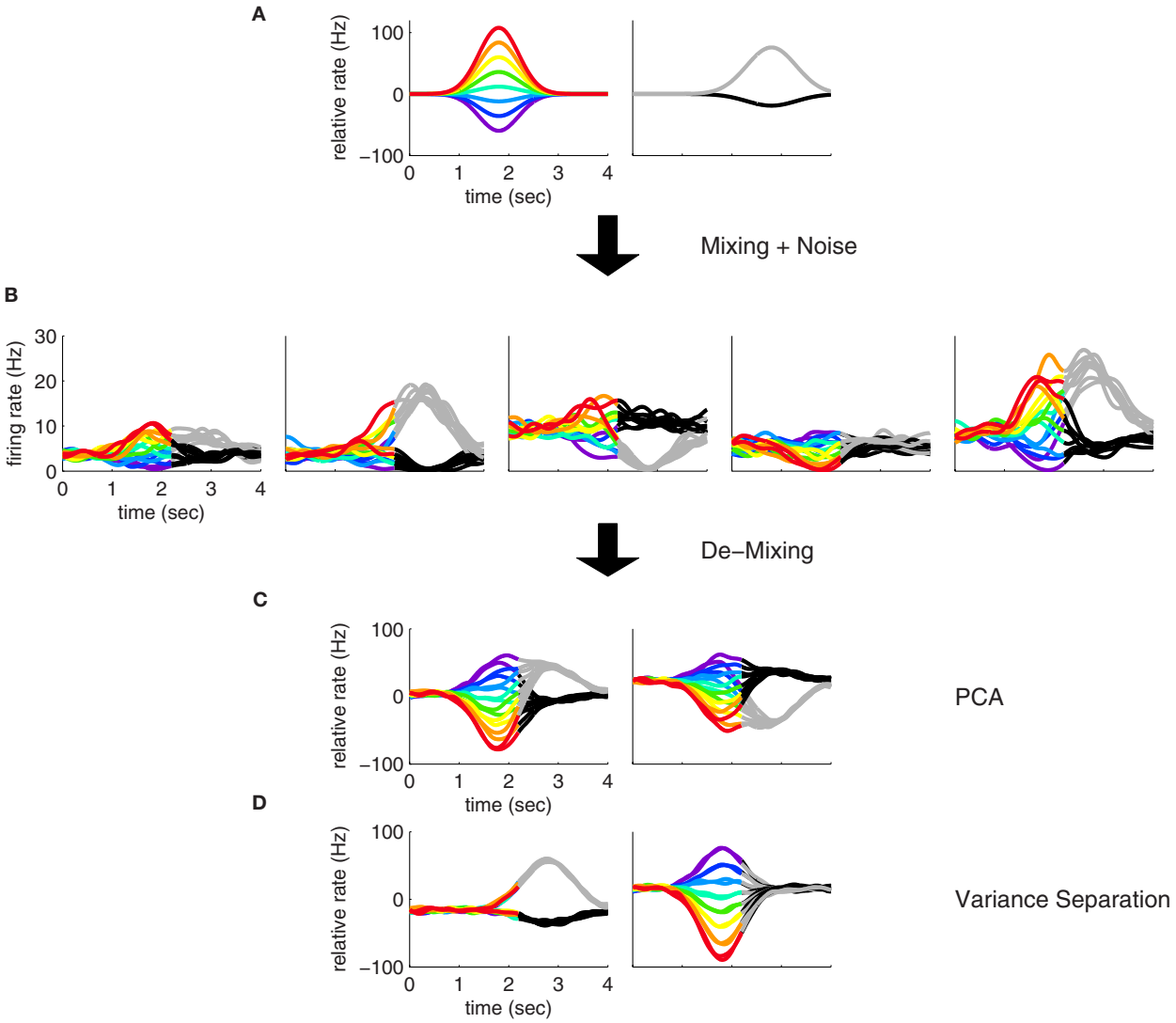


FIGURE 1 | Mixing and demixing of neural responses in a simulated two-alternative forced choice task. (A) We assume that neural responses are linear mixtures of two underlying components, one of which encodes the stimulus (left, colors representing different stimuli), and one of which encodes the binary decision (right) of a two-alternative-forced choice task. For concreteness, we assume that the task comprised $M_s = 8$ stimuli and $M_d = 2$ decisions. **(B)** Single cell responses are random combinations of these two components. We assume that $N = 50$ neurons have been recorded, five of which are shown here. The noisy variability of the responses was obtained by transforming the deterministic, linear mixture of each neuron into 10 inhomogeneous Poisson spike trains, and then re-estimating the firing rates by low-pass filtering and averaging the spike trains. This type of noise may be considered more realistic, even if it deviates from the assumptions in the main text. In our numerical example, this did not prove to be a problem. To systematically address such problems, however, one may apply a variance-stabilizing transformation to the data, such as taking the square-root of the firing rates before computing the covariance matrix (see e.g., Efron, 1982). **(C)** PCA uncovers the underlying two-dimensionality of the data, but the resulting coordinates do not demix the separate sources of firing rate variance. **(D)** By explicitly contrasting these separate sources, we can retrieve the original components up to a sign.

that captures most of the variance of the data by maximizing the function L with respect to $\mathbf{u}$ subject to the normalization constraint $\mathbf{u}^T\mathbf{u} = 1$. The solution corresponds to the first axis of the coordinate system that PCA constructs. If we are looking for several mutually orthogonal axes, these can be conveniently summarized into an $N \times n$ orthogonal matrix, $U = [\mathbf{u}_1,\ldots,\mathbf{u}_n]$. To find the maximum amount of variance that falls into the subspace spanned by these axes, we need to maximize

$$L = \sum_{i=1}^{n} \mathbf{u}_i^T C \mathbf{u}_i = \mathrm{tr}(U^T C U) \quad \text{subject to} \quad U^T U = I_n, \tag{7}$$

where the trace-operation, $\mathrm{tr}(\cdot)$, sums over all the diagonal entries of a matrix, and I_n denotes the $n \times n$ identity matrix.

Mathematically, the *principal* axes $\mathbf{u}_i$ correspond to the eigenvectors of the covariance matrix, C, which can nowadays be computed quite easily using numerical methods. Subsequently, the data can be plotted in the new coordinate system. The new coordinates of the data are given by

$$\mathbf{y}(t,s,d) = U^T\left(\mathbf{r}(t,s,d) - \mathbf{r}\right). \tag{8}$$

These new coordinates are called the *principal components*. Note that the new coordinate system has a different origin from the old one, since we subtracted the vector of mean firing rates, **r**. Consequently, the principal components can take both negative and positive values. Note also that the principal components are only defined up to a minus sign since every coordinate axis can be reflected along the origin. For our artificial data set, only two eigenvalues are non-zero, so that two principal components suffice to capture the complete variance of the data. The data in these two new coordinates, $y_1(t,s,d)$ and $y_2(t,s,d)$, are shown in **Figure 1C**.

Our toy model shows how PCA can succeed in summarizing the population response, yet it also illustrates the key problem of PCA: just as the individual neurons, the components mix information about the different task parameters (**Figure 1C**), even though the original components do not (**Figure 1A**). The underlying problem is that PCA ignores the causes of firing rate variability. Whether firing rates have changed due to the external stimulus s, due to the internally generated decision d, or due to some other cause, they will enter equally into the computation of the covariance matrix and therefore not influence the choice of the coordinate system constructed by PCA.

To make these notions more precise, we compute the covariance matrix of the simulated data. Inserting Eq. 3 into Eq. 6, we obtain

$$C = \mathbf{a}_1\mathbf{a}_1^T M_{11} + \mathbf{a}_2\mathbf{a}_2^T M_{22} + \left[\mathbf{a}_1\mathbf{a}_2^T + \mathbf{a}_2\mathbf{a}_1^T\right] M_{12} + H, \tag{9}$$

where M_{11} and M_{22} denote firing rate variance due to the first and second component, respectively, M_{12} denotes firing rate variance due to a mix of the two components, and H is the covariance matrix of the noise. Using the short-hand notations $z_1(t) = \langle z_1(t,s)\rangle_s$, $z_2(t) = \langle z_2(t,d)\rangle_d$, and $z_i = \langle z_i(t)\rangle_t$ for $i \in [1,2]$, the different variances are given by

$$M_{11} = \left\langle (z_1(t,s) - z_1)^2 \right\rangle_{t,s}, \tag{10}$$

$$M_{22} = \left\langle (z_2(t,d) - z_2)^2 \right\rangle_{t,d}, \tag{11}$$

$$M_{12} = \left\langle (z_1(t) - z_1)(z_2(t) - z_2) \right\rangle_t. \tag{12}$$

Principal component analysis will only be able to segregate the stimulus- and decision-dependent variance if the mixture term M_{12} vanishes and if the variances of the individual components, M_{11} and M_{22}, are sufficiently different from each other. However, if the two underlying components $z_1(t,s)$ and $z_2(t,d)$ are temporally correlated, then the mixture term M_{12} will be non-zero. Its presence will then force the eigenvectors of C away from $\mathbf{a}_1$ and $\mathbf{a}_2$. Moreover, even if the mixture term vanishes, PCA may still not be able to retrieve the original mixture coefficients, if the variances of the individual components, M_{11} and M_{22} are too close to each other when compared to the magnitude of the noise: in this case the eigenvalue problem becomes degenerate. In general, the covariance matrix therefore mixes different origins of firing rate variance rather than separating them. While PCA allows us to reduce the dimensionality of the data, the coordinate system found may therefore provide only limited insight into how the different task parameters are represented in the neural activities.

DEMIXING RESPONSES USING COVARIANCES OVER MARGINALIZED DATA

To solve these problems, we need to separate the different causes of firing rate variability. In the context of our example, we can attribute changes in the firing rates to two separate sources, both of which contribute to the covariance in Eq. 6. First, firing rates may change due to the externally applied stimulus s. Second, firing rates may change due to the internally generated decision d.

To account for these separate sources of variance in the population response, we suggest to estimate one covariance matrix for every source of interest. Such a covariance matrix needs to be specifically targeted toward extracting the relevant source of firing rate variance without contamination by other sources. Naturally, this step is somewhat problem-specific. For our example, we will first focus on the problem of estimating firing rate variance caused by the stimulus separately from firing rate variance caused by the decision. When averaging over all stimuli, we obtain the marginalized firing rates $\mathbf{r}(t,d) = \langle \mathbf{r}(t,s,d)\rangle_s$. The covariance caused by the stimulus is then given by the $N \times N$ matrix

$$C_s = \left\langle \left(\mathbf{r}(t,s,d) - \mathbf{r}(t,d)\right)\left(\mathbf{r}(t,s,d) - \mathbf{r}(t,d)\right)^T \right\rangle_{t,s,d}. \tag{13}$$

We will refer to C_s as the marginalized covariance matrix for the stimulus. We can repeat the procedure for the decision-part of the task. Marginalizing over decisions, we obtain $\mathbf{r}(t,s) = \langle \mathbf{r}(t,s,d)\rangle_d$ and

$$C_d = \left\langle \left(\mathbf{r}(t,s,d) - \mathbf{r}(t,s)\right)\left(\mathbf{r}(t,s,d) - \mathbf{r}(t,s)\right)^T \right\rangle_{t,s,d}. \tag{14}$$

Having two different covariance matrices, one may now perform two separate PCAs, one for each covariance matrix. In turn, one obtains two separate coordinate systems, one in which the principal axes point into the directions of state space along which firing rates vary if the stimulus is changed, the other in which they point into the directions along which firing rates vary if the decision changes.

For the toy model, it is readily seen that the marginalized covariance matrices are given by $C_s = \mathbf{a}_1\mathbf{a}_1^T M_{s,11} + H$ and $C_d = \mathbf{a}_2\mathbf{a}_2^T M_{d,22} + H$ with $M_{s,11} = \langle (z_1(t,s) - z_1(t))^2\rangle$ and $M_{d,22} = \langle (z_2(t,d) - z_2(t))^2\rangle$. Consequently, the principal eigenvectors of C_s and C_d will be equivalent to the mixing coefficients $\mathbf{a}_1$ and $\mathbf{a}_2$, at least as long as the variances $M_{s,11}$ and $M_{d,22}$ are much larger than the size of the noise, which is given by $\mathrm{tr}(H)$.

If the noise term is not negligible, it will force the eigenvectors away from the actual mixing coefficients. This problem can be alleviated by using the orthogonality condition, $\mathbf{a}_1^T\mathbf{a}_2 = 0$, which implies that there are separate sources of variance for the stimulus- and decision-components. To this end, we can seek to divide the full space into two subspaces, one that captures as much as possible about the stimulus-dependent covariance C_s, and another, that captures as much as possible about the decision-dependent covariance C_d. Our goal will then be to maximize the function

$$L = \mathrm{tr}\left(U_1^T C_s U_1\right) + \mathrm{tr}\left(U_2^T C_d U_2\right) \tag{15}$$

with respect to the two orthogonal matrices U_1 and U_2 whose columns contain the basis vectors of the respective subspaces. The first term in Eq. 15 captures the total variance falling into

the subspace spanned by the columns of U_1, and the second term the total variance falling into the subspace given by U_2. Writing $U = [U_1, U_2]$, we obtain an orthogonal matrix for the full space, and the orthogonality conditions are neatly summarized by $UU^T = I$. As shown in the Appendix, the maximization of Eq. 15 under these orthogonality constraints can be solved by computing the eigenvectors and eigenvalues of the difference of covariance matrices,

$$D = C_s - C_d. \tag{16}$$

In this case, the eigenvectors belonging to the positive eigenvalues of D form the columns of U_1 and the eigenvectors belonging to the negative eigenvalues of D form the columns of U_2. As with PCA, the positive or negative eigenvalues can be sorted according to the amount of variance they capture about C_s and C_d.

For the simulated example, we obtain

$$D = \mathbf{a}_1 \mathbf{a}_1^T M_{s,11} - \mathbf{a}_2 \mathbf{a}_2^T M_{d,22}, \tag{17}$$

where the noise term H has now dropped out. Diagonalization of D results in two clearly separated eigenvalues, $M_{s,11}$ and $-M_{d,11}$, and in two eigenvectors, $\mathbf{a}_1$ and $\mathbf{a}_2$, that correspond to the original mixing coefficients.

LINKING THE POPULATION LEVEL AND THE SINGLE CELL LEVEL

As a result of the above method, we obtain a new coordinate system, whose basis vectors are given by the columns of the matrix U. This coordinate system provides simply a different, and hopefully useful, way of representing the population response. One major advantage of orthogonality is that one can easily move back and forth between the single cell and population level description of the neural activities. Just as in PCA, we can project the original firing rates of the neurons onto the new coordinates,

$$\mathbf{y}(t,s,d) = U^T \left(\mathbf{r}(t,s,d) - \mathbf{r} \right), \tag{18}$$

and the two leading coordinates for the toy model are shown in **Figure 1D**. These components correspond approximately to the original components, $z_1(t,s)$ and $z_2(t,d)$. In turn, we can reconstruct the activity of each neuron by inverting the coordinate transform,

$$\mathbf{r}(t,s,d) = U\mathbf{y}(t,s,d) + \mathbf{r}. \tag{19}$$

For every neuron this yields a set of N reconstruction coefficients which correspond to the rows of U.

Since two coordinates were sufficient to capture most of the variance in the toy example, the firing rate of every neuron can be reconstructed by a linear combination of these two components, $y_1(t,s,d)$ and $y_2(t,s,d)$. For each neuron, we thereby obtain two reconstruction coefficients, u_{i1} and u_{i2}. The set of all reconstruction coefficients constitutes a cloud of points in a two-dimensional space. The distribution of this cloud, together with the activities of several example neurons are shown in **Figure 2**. This plot allows us to link the single cell with the population level by visualizing how the activity of each neuron is composed out of the two underlying components.

GENERALIZATIONS TO MORE THAN TWO PARAMETERS

In our toy example, we have assumed that each task parameter is represented by a single component. We note that this is a feature of our specific example. In more realistic scenarios, a single

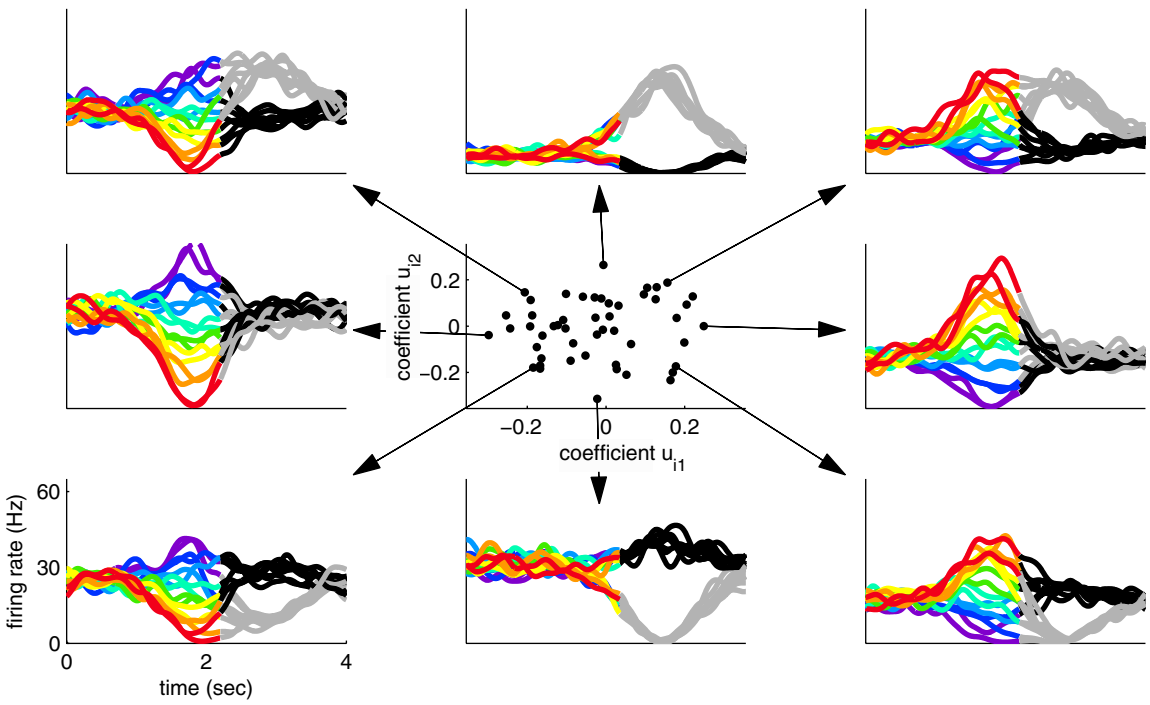


FIGURE 2 | Linking the single cell and population level. Using the coordinate system retrieved by separating variances, we can illuminate the contributions of each component to the individual neural responses. The center shows the contribution of the stimulus- and decision-related components to each individual neuron. The surrounding plots show the activity of eight example neurons, corresponding to the respective dots in the center.

task parameter could potentially be represented by more than one component. For instance, if one set of neurons fires transiently with respect to a stimulus *s*, but another set of neurons fires tonically, then the firing rate dynamics of the stimulus representation are already two-dimensional, even without taking the decision into account. In such a case, we can still use the method described above to retrieve the two subspaces in which the respective components lie.

However, the number of task parameters will often be larger than two. In the two-alternative-forced choice task, there are at least four parameters that could lead to changes in firing rates: the timing of the task, *t*, potentially related to anticipation or rhythmic aspects of a task, the stimulus, *s*, the decision, *d*, and the reward, *r*. Even more task parameters could be of interest, such as those extracted from previous trials etc.

These observations raise the question of how the method can be generalized if there are more than two task parameters to account for. To do so, we write the relevant parameters into one long vector $\boldsymbol{\theta} = (\theta_1,\theta_2,\ldots,\theta_M)$, and assume that the firing rates of the neurons are linear mixtures of the form

$$\mathbf{r}(t,\boldsymbol{\theta}) = \mathbf{a}_{11} z_{11}(t,\theta_1) + \mathbf{a}_{12} z_{12}(t,\theta_1) + \ldots \tag{20}$$

$$+\mathbf{a}_{21} z_{21}(t,\theta_2) + \mathbf{a}_{22} z_{22}(t,\theta_2) + \ldots \tag{21}$$

$$+\mathbf{a}_{M1} z_{M1}(t,\theta_M) + \ldots, \tag{22}$$

where each task parameter is now represented by more than one component. For each parameter, θ_i, we can compute the marginalized covariance matrix,

$$C_i = \left\langle \left(\mathbf{r}(t,\boldsymbol{\theta}) - \langle \mathbf{r}(t,\boldsymbol{\theta})\rangle_{\theta_i}\right)\left(\mathbf{r}(t,\boldsymbol{\theta}) - \langle \mathbf{r}(t,\boldsymbol{\theta})\rangle_{\theta_i}\right)^T \right\rangle_{t,\boldsymbol{\theta}}, \tag{23}$$

which measures the covariance in the firing rates due to changes in the parameter θ_i. Diagonalizing each of these covariance matrices will retrieve the various subspaces corresponding to the different mixture coefficients. For instance, when diagonalizing C_1, we obtain the subspace for the components that depend on the parameter θ_1. The relevant eigenvectors of C_1 will therefore span the same subspace as the mixture coefficients $\mathbf{a}_{11}$, $\mathbf{a}_{12}$, etc., in Eq. 22.

As before, the method's performance under additive noise can be enhanced by maximizing a single function (see Appendix)

$$L = \sum_{i=1}^{M} \mathrm{tr}\left(U_i^T C_i U_i\right) \tag{24}$$

subject to the orthogonality constraint $U^T U = I$ for $U = [U_1,U_2,\ldots,U_M]$. Maximization of this function will force the firing rate variance due to different parameters θ_i into orthogonal subspaces (as required by the model). If $M = 1$, then maximization results in a standard PCA. In the case $M = 2$, maximization requires the diagonalization of the difference of covariance matrices $C_1 - C_2$, as in Eq. 16. In the case $M > 2$, various algorithms can be constructed to find local maxima of L (see e.g., Bolla et al., 1998). To our knowledge, a full understanding of the global solution structure of the maximization problem does not exist for $M > 2$. In the Appendix, we show how to maximize Eq. 24 with standard gradient ascent methods. In any case, it may often be a good idea to use PCA on the full covariance matrix of the data, Eq. 6, to reduce the dimensionality of the data set prior to the demixing procedure. Indeed, this preprocessing step was applied in Machens et al. (2010).

FURTHER GENERALIZATIONS AND LIMITATIONS OF THE METHOD

The above formulation of the problem may be further generalized by allowing individual components to mix parameters in nontrivial ways. To study this scenario in a simple example, imagine that in the above two-alternative-forced choice task, in addition to the stimulus- and decision-dependent component, there were a purely time-dependent component, $z_3(t)$, locked to the time structure of the task, so that

$$\mathbf{r}(t,s,d) = \mathbf{a}_1 z_1(t,s) + \mathbf{a}_2 z_2(t,d) + \mathbf{a}_3 z_3(t) + \mathbf{c} + \mathbf{n}(t). \tag{25}$$

This scenario is illustrated in **Figures 3A,B**. As before, we can compute marginalized covariance matrices, that capture the covariance due to the stimuli *s*, the decisions *d*, or the time points *t*. While the marginalized covariance matrices for the stimuli and decisions, C_s and C_d, have one significant eigenvalue each, and thereby capture the relevant component (**Figure 3C**), the marginalized covariance matrix for time, C_t, now has three significant eigenvalues, and therefore does not allow us to retrieve the purely time-dependent component $z_3(t)$. The reason for this failure is that all three components in Eq. 25 have a time-dependence that cannot be averaged out. By design, the stimulus-averaged first component, $z_1(t) = \langle z_1(t,s)\rangle_s$, and the decision-averaged second component, $z_2(t) = \langle z_2(t,d)\rangle_d$ do not vanish. In other words, the stimulus- and decision-components have intrinsic time-dependent variance that cannot be separated from the stimulus- or decision-induced variance.

Consequently, the subspace spanned by the first three eigenvectors of C_t overlaps with the respective subspaces spanned by the first eigenvectors of C_s and C_d. One way to visualize this overlap is to take the five relevant eigenvectors (three for C_t, one for C_s, and one for C_d) and compute how much of the variance of each marginalized covariance matrix they capture. To do so, we compute the "confusion matrix"

$$S_{ij} = \frac{\mathbf{u}_i^T C_j \mathbf{u}_i}{\mathrm{tr}(C_j)}. \tag{26}$$

This confusion matrix measures what percentage of the variance attributed to the *j*-th cause is captured by the *i*-th coordinate. For the above example, it is illustrated in **Figure 3D**. If in one row of this matrix, more than one entry is significantly above 0, then more than one covariance matrix has significant variance along that direction of state space. Whereas the eigenvectors of the C_s and C_d matrix do not interfere with each other, i.e., they are approximately orthogonal, the eigenvectors of the C_t matrix interfere with both the C_s and C_d eigenvectors, i.e., the respective subspaces overlap. The method introduced above will still yield a result in this case, however, the new coordinate system will generally not retrieve the original components.

An *ad hoc* solution to this problem may be to section the three-dimensional eigenvector subspace of C_t, and identify a direction that is orthogonal to the first eigenvectors of C_s and C_d, which will

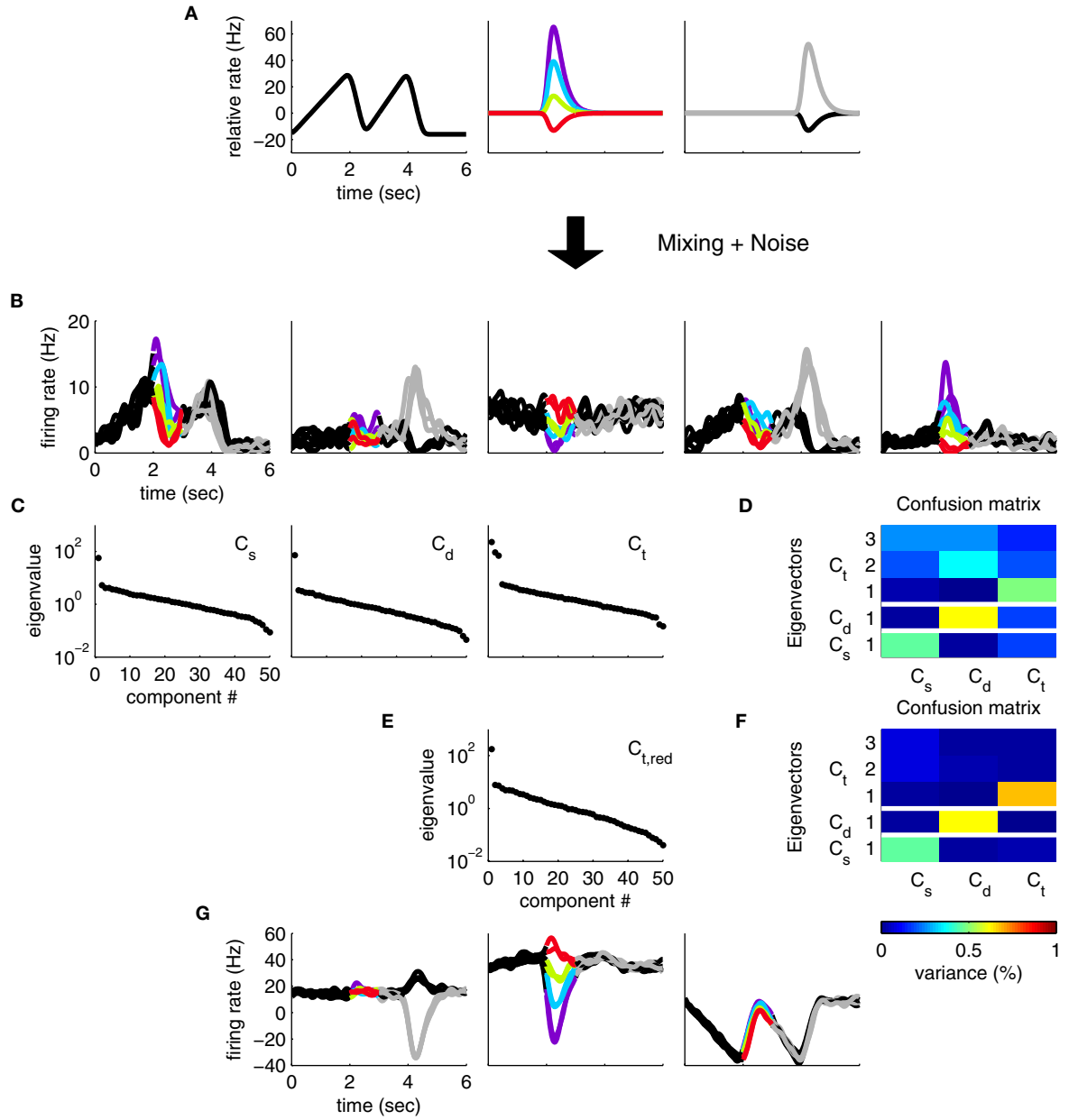


FIGURE 3 | Heterogeneous responses as a result of linear mixture of three components. (A) We now assume that neural responses are constructed from three underlying components, one of which encodes only the task rhythm (left), while the two others encode stimuli and responses, as in **Figure 1**. **(B)** Single cell responses are random combinations of these three components. We again assume that $N = 50$ neurons have been recorded, five of which are shown here, and response noise was generated as in **Figure 1**. **(C)** Whereas the marginalized covariance matrices for stimulus and decision, C_s and C_d, have only one significant eigenvalue, the one for time, C_t, features three eigenvalues above the noise floor. **(D)** The confusion matrix displays the percentage of variance captured by the different components (*y*-axis) with respect to the different marginalized covariance matrices (*x*-axis). The C_s row shows, for instance, that the first eigenvector of C_s captures a significant amount of variance from the C_s matrix (which is good), but it also captures a significant amount of variance from the C_t matrix (which is bad). **(E)** By estimating the time-dependent covariance over a time window limited to $t \in [0,2]$, we can reduce the number of significant eigenvalues to one. **(F)** In turn, every row of the confusion matrix has only one entry, showing that every component captures the variance of one marginalized covariance matrix only. **(G)** The firing rates projected onto the components from **(F)**.

then correspond to the purely time-dependent component $z_3(t)$. Alternatively, we could restrict the estimation of C_t to the time before stimulus onset, so that the covariance matrix is no longer contaminated by time-dependent variance from the stimulus- or decision-components. The rank of C_t then reduces to one, and the different components separate nicely (**Figures 3E,F,G**). While feasible in our toy scenario, these *ad hoc* procedures are not guaranteed to work for real data, when more dimensions are involved, and

more complex confusion matrices may result. However, the latter solution demonstrates that by a judicious choice of marginalized covariance matrices, one may sometimes be able to avoid such problems of non-separability.

CONNECTION TO BLIND SOURCE SEPARATION METHODS

In all of these scenarios, we assumed that the firing rates $\mathbf{r}$ are linear mixtures of a set of underlying sources $\mathbf{z}$, each with mean 0, so that

$$\mathbf{r} = A\mathbf{z} + \mathbf{c}. \tag{27}$$

The problem that we have been describing then consists in estimating the unknown sources, $\mathbf{z}$, the unknown mixture coefficients, A, and the unknown bias parameters $\mathbf{c}$ from the observed data, $\mathbf{r}$. Without loss of generality, we can assume that the sources are centered so that $\langle z \rangle = 0$. Ours is therefore a specific version of the much-studied blind source separation problem (see e.g., Molgedey and Schuster, 1994; Bell and Sejnowski, 1995). In many standard formulations of this problem, one assumes that the sources are uncorrelated, or even statistically independent, which implies that the covariance matrix of the sources, $M = \langle \mathbf{z}\mathbf{z}^T \rangle_t$, is diagonal.

In our case, we do not want to make this assumption, which rules out the use of many blind source separation methods, such as independent component analysis (Hyvärinen et al., 2001). On the upside, we do have additional information, in the form of n task parameters, that provide indirect clues toward the underlying sources. More specifically, we assume that the sources are of the form $z_k(t,\theta_k)$ where θ_k denotes a single task parameter, or a specific combination of task parameters. For each task parameter, we can estimate the marginalized covariance matrix C_i, which in turn is given by $C_i = AM_iA^T$ with

$$M_i = \left\langle \left(\mathbf{z}(t,\boldsymbol{\theta}) - \langle \mathbf{z}(t,\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}_i}\right)\left(\mathbf{z}(t,\boldsymbol{\theta}) - \langle \mathbf{z}(t,\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}_i}\right)^T \right\rangle_{t,\boldsymbol{\theta}} \tag{28}$$

As long as different task parameters are distributed over different components, the matrix M_i will be block-diagonal. In the most general case, however, as discussed above, this will not be true. If one parameter is shared among several components, then the respective marginalized covariance matrix will capture variance from all of these components, and maximization of Eq. 24 will not necessarily retrieve the original components. Future work may show how this general, semi-blind source separation problem can be solved by using knowledge about the structure of the marginalized M-matrices. For now, we suggest that in many practical scenarios, a judicious choice of covariance measurements, for instance, by focusing on particular time intervals of a task etc., may help to partly reduce the problem to those that are completely separable, as in Eq. 22.

DISCUSSION

In this article, we addressed the problem of analyzing neural recordings with strong response heterogeneity. A key problem for these data sets is first and foremost the difficulty of visualizing the neural activities at the population level. Simply parsing through individual neural responses is often not sufficient, hence the quest for methods that provide a useful and interpretable summary of the population response.

To provide such a summary, we made one crucial assumption. We assumed that the heterogeneity of neural responses is caused by a simple mixing procedure in which the firing rates of individual neurons are random, linear combinations of a few fundamental components. We believe that such a scenario is likely to be responsible for at least part of the observed response diversity. Higher-level areas of the brain are known to integrate and process information from many other areas in the brain. The presumed fundamental components could be given by the inputs and outputs of these areas. If such components are mixed at random at the level of single cells, then upstream or downstream areas can access the relevant information with simple linear and orthogonal read-outs. Such linear population read-outs have long been known to work quite well in various neural systems (Seung and Sompolinsky, 1993; Salinas and Abbott, 1994).

To retrieve the components from recorded neural activity, and thereby at least partly reduce the response heterogeneity, we suggest to estimate the covariances in the firing rates that can be attributed to the experimentally controlled, external task parameters. Using these marginalized covariance matrices, we showed how to construct an orthogonal coordinate system such that individual coordinates capture the main aspects of the task-related neural activities and the coordinate system as a whole captures all aspects of the neural activities. In the new coordinate system, firing rate variance due to different task parameters is projected onto orthogonal coordinates, making visualization and interpretation of the data particularly easy. We note, though, that the existence of a useful, orthogonal coordinate system is not guaranteed by the method, but can only be a feature of the data. Our method will generally not return useful results if mixing is linear, but not orthogonal, or if mixing is non-linear. Nonetheless, the case of non-orthogonal, linear mixing, may still be investigated through separate PCAs on the different marginalized covariance matrices.

Other methods exist that address similar goals. Most prominently, application of canonical correlation analysis (CCA) to the type of data discussed here would also construct a coordinate system whose choice is influenced by knowledge about the task structure. In our context, CCA would seek a coordinate axis in the state space of neural responses and a coordinate axis in the space of task parameters, such that the correlation between the two is maximized. Whether this method would yield a useful, i.e., interpretable, coordinate system for real data sets remains open to investigation. CCA has recently been proposed as a method to construct population responses in sensory systems (Macke et al., 2008) and as a way to correlate electrophysiological with fMRI data (Biessmann et al., 2009).

Further extensions and generalizations of PCA exist, some of which are specifically targeted to the type of data we have discussed here. The work of Yu et al. (2009), for instance, explicitly addresses the problems that are incurred by estimating firing rates

prior to the dimensionality reduction. They show how to combine these two separate steps into a single one using the theory of Gaussian processes. Their work is therefore complementary to ours, and could potentially be incorporated into the methodology introduced here.

Methods to summarize population activity have been employed in many different neurophysiological settings (Friedrich and Laurent, 2001; Stopfer et al., 2003; Paz et al., 2005; Narayanan and Laubach, 2009; Yu et al., 2009). Our main aim here was to modify these methods such that experimentally controlled parameters are taken into account and influence the construction of a new coordinate system. A first application of this method to neural responses from the prefrontal cortex revealed new aspects of a previously studied data set (Machens et al., 2010). Many other data sets with strong response heterogeneity may be amenable to a similar analysis.

ACKNOWLEDGMENTS

I thank Claudia Feierstein, Naoshige Uchida, and Ranulfo Romo for access to their multi-electrode data which have been the main source of inspiration for the present work. I furthermore thank Carlos Brody, Matthias Bethge, Claudia Feierstein, and Thomas Schatz for helpful discussions along various stages of the project. My work is supported by an Emmy-Noether grant from the Deutsche Forschungsgemeinschaft and a Chair d'excellence grant from the Agence Nationale de la Recherche.

REFERENCES

Averbeck, B. B., Sohn, J.-W., and Lee, D. (2006). Activity in prefrontal cortex during dynamic selection of action sequences. *Nat. Neurosci.* 9, 276–282.

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159.

Biessmann, F., Meinecke, F. C., Gretton, A., Rauch, A., Rainer, G., Logothetis, N., and Müller, K. R. (2009). Temporal kernel canonical correlation analysis and its application in multimodal neuronal data analysis. *Mach. Learn.* 79, 5–27.

Bolla, M., Michaletzky, G., Tusnády, G., and Ziermann, M. (1998). Extrema of sums of heterogeneous quadratic forms. *Linear Algebra Appl.* 269, 331–365.

Brody, C. D., Hernandez, A., Zainos, A., and Romo, R. (2003). Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb. Cortex* 13, 1196–1207.

Churchland, M. M., and Shenoy, K. V. (2007). Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. *J. Neurophysiol.* 97, 4235–4257.

Efron, B. (1982). Transformation theory: how normal is a family of distributions? *Ann. Stat.* 10, 328–339.

Eliasmith, C., and Anderson, C. H. (2003). *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*. Cambridge, MA: MIT Press.

Feierstein, C. E., Quirk, M. C., Uchida, N., Sosulski, D. L., and Mainen, Z. F. (2006). Representation of spatial goals in rat orbitofrontal cortex. *Neuron* 51, 495–507.

Friedrich, R. W., and Laurent, G. (2001). Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. *Science* 291, 889–894.

Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning Theory*. New York: Springer.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. New York: Wiley InterScience.

Jun, J. K., Miller, P., Hernández, A., Zainos, A., Lemus, L., Brody, C., and Romo, R. (2010). Heterogeneous population coding of a short-term memory and decision-task. *J. Neurosci.* 30, 916–929.

Kepecs, A., Uchida, N., Zariwala, H. A., and Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455, 227–231.

Kiani, R., and Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324, 759–764.

Machens, C. K., Romo, R., and Brody, C. D. (2010). Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *J. Neurosci.* 30, 350–360.

Macke, J. H., Zeck, G., and Bethge, M. (2008). "Receptive fields without spike-triggering," in *Advances in Neural Information Processing Systems 20*, eds J. C. Platt, D. Koller, Y. Singer, and S. Roweis (Red Hook, NY: Curran), 969–976.

Miller, E. K. (1999). The prefrontal cortex: complex neural properties for complex behavior. *Neuron* 22, 15–17.

Molgedey, L., and Schuster, H. G. (1994). Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* 72, 3634–3637.

Narayanan, N. S., and Laubach, M. (2009). Delay activity in rodent frontal cortex during a simple reaction time task. *J. Neurophysiol.* 101, 2859–2871.

Newsome, W. T., Britten, K. H., and Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature* 341, 52–54.

Nicolelis, M. A. L., Baccala, L. A., Lin, R. C. S., and Chapin, J. K. (1995). Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science* 268, 1353–1358.

Paz, R., Natan, C., Boraud, T., Berman, H., and Vaadia, E. (2005). Emerging patterns of neuronal responses in supplementary and primary motor areas during sensorimotor adaptation. *J. Neurosci.* 25, 10941–10951.

Rao, S. C., Rainer, G., and Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science* 276, 821–824.

Romo, R., Brody, C. D., Hernandez, A., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399, 470–473.

Romo, R., Hernández, A., Zainos, A., Lemus, L., and Brody, C. D. (2002). Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat. Neurosci.* 5, 1217–1225.

Salinas, E., and Abbott, L. F. (1994). Vector reconstruction from firing rates. *J. Comput. Neurosci.* 1, 89–107.

Seo, H., Barraclough, D. J., and Lee, D. (2009). Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. *J. Neurosci.* 29, 7278–7289.

Seung, H. S., and Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proc. Natl. Acad. Sci. U.S.A.* 90, 10749–10753.

Stopfer, M., Jayaraman, V., and Laurent, G. (2003). Intensity versus identity coding in an olfactory system. *Neuron* 39, 991–1004.

Sugrue, L. P., Corrado, G. S., and Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science* 304, 1782–1787.

Uchida, N., and Mainen, Z. F. (2003). Speed and accuracy of olfactory discrimination in the rat. *Nat. Neurosci.* 6, 1224–1229.

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* 102, 614–635.

Zacksenhouse, M., and Nemets, S. (2008). "Strategies for neural ensemble data analysis for brain–machine interface (BMI) applications," in *Methods for Neural Ensemble Recordings*, ed. M. A. L. Nicolelis (Boca Raton, FL: CRC Press), 57–82.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 November 2009; paper pending published: 19 November 2009; accepted: 27 July 2010; published online: 06 October 2010.

*Citation: Machens CK (2010). Demixing population activity in higher cortical areas. Front. Comput. Neurosci. **4**:126. doi:10.3389/fncom.2010.00126*

APPENDIX

MAXIMIZATION FOR TWO COVARIANCE MEASUREMENTS

Assume that our goal is to separate the state space into two mutually orthogonal subspaces, such that most of the variance measured by C_1 falls into one subspace, and most of the variance measured by C_2 into the orthogonal subspace. To do so, we define a matrix U_1 whose columns contain a set of vectors $\mathbf{u}_i$ with $i = 1,\ldots,M$, and a matrix U_2 whose columns contain a set of vectors $\mathbf{u}_i$ with $i = M+1,\ldots,N$. All vectors are mutually orthonormal, so that $\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$. Our goal will then be to maximize

$$L = \mathrm{tr}\left(U_1^T C_1 U_1\right) + \mathrm{tr}\left(U_2^T C_2 U_2\right). \tag{29}$$

The orthogonality constraint is given by the condition $U_1 U_1^T + U_2 U_2^T = I$. By the rules of traces, and using this constraint, we obtain

$$\begin{aligned} L &= \mathrm{tr}\left(U_1 U_1^T C_1\right) + \mathrm{tr}\left(U_2 U_2^T C_2\right) \\ &= \mathrm{tr}\left(U_1 U_1^T C_1 + \left(I - U_1 U_1^T\right) C_2\right) \\ &= \mathrm{tr}\left(U_1 U_1^T (C_1 - C_2)\right) + tr(C_2). \end{aligned}$$

The last line is maximized if the matrix U_1 contains all the eigenvectors that correspond to the positive eigenvalues of $C_1 - C_2$. Consequently, the matrix U_2 will contain all the eigenvectors corresponding to the negative eigenvalues of $C_1 - C_2$. The extremal eigenvalues of the difference matrix, i.e., the largest and the smallest, correspond to the two eigenvectors that capture most of the variance in C_1 and C_2 under the given trade-off.

ADDITIVE NOISE DOES NOT AFFECT THE MAXIMUM

To study the maximization problem under condition of additive noise, we assume n covariance measurements so that

$$C_i = S_i + H, \tag{30}$$

where S_i is the signal-part and H the noise part of the covariance matrix. Since the noise acts additively on the firing rates, every covariance measurement is polluted with the same amount of noise, H, compare Eq. 23. When maximizing Eq. 24 with respect to an orthogonal transform, $U = [U_1,\ldots,U_n]$, we will then target only the signal part of the covariance matrices, but not the noise part. To see that, we note that

$$L = \sum_{i=1}^{n} \mathrm{tr}\left(U_i^T C_i U_i\right) \tag{31}$$

$$= \mathrm{tr}\left(\sum_{i=1}^{n} U_i U_i^T C_i\right) \tag{32}$$

$$= \mathrm{tr}\left(\sum_{i=1}^{n-1} U_i U_i^T C_i + \left(I - \sum_{i=1}^{n-1} U_i U_i^T\right) C_n\right) \tag{33}$$

$$= \mathrm{tr}\left(\sum_{i=1}^{n-1} U_i U_i^T (C_i - C_n) + C_n\right) \tag{34}$$

$$= \mathrm{tr}\left(\sum_{i=1}^{n-1} U_i U_i^T (S_i - S_n)\right) + \mathrm{tr}(S_n + H). \tag{35}$$

Accordingly, the projection operators, $U_i U_i^T$, which project the variance into the relevant subspaces, target the difference of covariance matrices, $C_i - C_n$, so that the noise drops out, since $C_i - C_n = S_i - S_n$.

MAXIMIZATION FOR *N* COVARIANCE MEASUREMENTS

Maximization of Eq. 24,

$$L = \sum_{i=1}^{n} \mathrm{tr}\left(U_i^T C_i U_i\right) \quad \text{subject to} \quad UU^T = 1 \tag{36}$$

is a quadratic optimization problem under quadratic constraints which can be solved numerically by any of a standard set of methods. A specific method to solve a related problem has been proposed in Bolla et al. (1998). Here, we present an algorithm based on a simple gradient ascent.

First, we need an initial guess for the U_i. We suggest to use the first principal axes (eigenvector with largest eigenvalue) of the marginalized covariance matrix C_i. This procedure, however, will generally yield a set of matrices U_i which are not mutually orthogonal. To orthogonalize these vectors, one can use the method of symmetric orthogonalization. Given the initial guess for the matrix, $U = [U_1,\ldots,U_n]$, the transform

$$U \to U(U^T U)^{-1/2} \tag{37}$$

will yield a matrix with mutually orthogonal columns so that $U^T U = I$. We will use this matrix U as our initial guess for the gradient ascent.

Next, let us define the matrix Q_i as an $n \times n$ matrix of zeros in which only the entry in the i-th column and i-th row is 1. The maximization over the captured variances, Eq. 36, can then be rewritten as

$$L = \sum_{i=1}^{n} \mathrm{tr}\left(U^T C_i U Q_i\right) \quad \text{subject to} \quad U^T U = I, \tag{38}$$

which allows us to compactly write the matrix derivative of L as

$$\frac{\partial L}{\partial U} = \sum_{i=1}^{n} C_i U Q_i. \tag{39}$$

Hence, to maximize L on the manifold of orthogonal matrices, U, we need to iterate the equations,

$$U \to U + \alpha \frac{\partial L}{\partial U} \tag{40}$$

$$U \to U(U^T U)^{-1/2}, \tag{41}$$

where the first equation performs a step toward the maximum, whose length is determined by the learning rate α, and the second step projects U back onto the manifold of orthogonal matrices.

frontiers in
COMPUTATIONAL NEUROSCIENCE

METHODS ARTICLE
published: 02 July 2010
doi: 10.3389/fncom.2010.00016

Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference

Benjamin Staude[1], Sonja Grün[2,3] and Stefan Rotter[1]*

[1] *Bernstein Center Freiburg and Faculty of Biology, Albert-Ludwig University, Freiburg, Germany*
[2] *Unit of Statistical Neuroscience, RIKEN Brain Science Institute, Wako-Shi, Japan*
[3] *Bernstein Center for Computational Neuroscience, Humboldt Unverstität zu Berlin, Berlin, Germany*

Edited by:
Jakob H. Macke, Max Planck Institute for Biological Cybernetics, Germany

Reviewed by:
Yasser Roudi, NORDITA, Sweden
Don H. Johnson, Rice University, USA
Jonathan D. Victor, Weill Cornell Medical College, USA

***Correspondence:**
Stefan Rotter, Bernstein Center Freiburg and Faculty of Biology, Albert-Ludwig University, Hansastrasse 9a, 79104 Freiburg, Germany.
e-mail: stefan.rotter@biologie.uni-freiburg.de

The extent to which groups of neurons exhibit higher-order correlations in their spiking activity is a controversial issue in current brain research. A major difficulty is that currently available tools for the analysis of massively parallel spike trains (N >10) for higher-order correlations typically require vast sample sizes. While multiple single-cell recordings become increasingly available, experimental approaches to investigate the role of higher-order correlations suffer from the limitations of available analysis techniques. We have recently presented a novel method for cumulant-based inference of higher-order correlations (CuBIC) that detects correlations of higher order even from relatively short data stretches of length T = 10–100 s. CuBIC employs the compound Poisson process (CPP) as a statistical model for the population spike counts, and assumes spike trains to be stationary in the analyzed data stretch. In the present study, we describe a non-stationary version of the CPP by decoupling the correlation structure from the spiking intensity of the population. This allows us to adapt CuBIC to time-varying firing rates. Numerical simulations reveal that the adaptation corrects for false positive inference of correlations in data with pure rate co-variation, while allowing for temporal variations of the firing rates has a surprisingly small effect on CuBICs sensitivity for correlations.

Keywords: multiple unit activity, higher-order correlations, non-stationarity, statistical population model

INTRODUCTION

It has long been suggested that fundamental insight into the nature of neuronal computation requires the understanding of the cooperative dynamics of populations of neurons (Hebb, 1949). A controversial issue in this debate is the role of correlations among nerve cells. On the one hand, an increasing body of both experimental (e.g., Gray and Singer, 1989; Vaadia et al., 1995; Riehle et al., 1997; Bair et al., 2001; Kohn and Smith, 2005; Shlens et al., 2006; Fujisawa et al., 2008; Pillow et al., 2008) and theoretical (Abeles, 1991; Diesmann et al., 1999; Kuhn et al., 2003) literature supports the concept of cooperative computation on various temporal and spatial scales. On the other hand, the mostly detrimental effect of correlations on rate-based information transmission and processing (Abbott and Dayan, 1999; Averbeck and Lee, 2006; Josić et al., 2009) has generated a strong opposition toward correlation-based concepts of cortical coding (Shadlen and Newsome, 1998; Averbeck et al., 2006; Schneidman et al., 2006; Ecker et al., 2010). Evidently, a thorough description of the correlation structure of neuronal populations is an indispensable prerequisite to resolve these opposing theoretical viewpoints (Brown et al., 2004).

Experimental reports on coordinated activity at the level of spike trains resort almost exclusively to correlations between pairs of nerve cells (e.g., Eggermont, 1990; Vaadia et al., 1995; Kreiter and Singer, 1996; Riehle et al., 1997; Kohn and Smith, 2005; Sakurai and Takahashi, 2006; Fujisawa et al., 2008; Ecker et al., 2010). Such pairwise correlations cannot, as a matter of principle, resolve the cooperative activity of neuronal populations to the extent required for rigorous hypothesis testing (Gerstein et al., 1989; Martignon et al., 2000; Brown et al., 2004). In particular, whether or not coincident spikes of pairs of neurons participate in synchronized "cluster-events" cannot be decided on measurements of pairwise correlation alone; this can only be achieved by the systematic assessment of higher-order correlations, i.e., statistical couplings among triplets, quadruplets, and larger groups (Martignon et al., 1995; Staude et al., 2010). Importantly, the nonlinear dynamics of spike generation makes neurons extremely sensitive for synchrony in their input pools (Softky, 1995; König et al., 1996). Ignoring these higher-order correlations in the statistical description of spiking populations is therefore hardly advisable (Bohte et al., 2000; Kuhn et al., 2003).

Initially, the main obstacle for assessing the higher-order structure of neuronal populations were limitations in experimental methodology, as until recently state-of-the-art electrophysiological setups allowed to record only few neurons simultaneously. The advent of multi-electrode arrays and optical imaging techniques, however, now reveals fundamental shortcomings of available analysis tools (Brown et al., 2004). Mathematical frameworks to model and estimate higher-order correlations typically assign one "interaction parameter" for every subgroup of the population, leading to a $2^N - 1$ dimensional model for a population comprising N neurons (Martignon et al., 1995, 2000). The associated estimation problem greatly suffers from this combinatorial explosion: the number of parameters to be estimated from the available sample size (a population of $N = 100$ neurons implies $\sim 10^{30}$ parameters while 100 s of data provide only $\sim 10^6$ samples) illustrates the principal infeasibility of this approach. In fact, the estimation of such higher-order

correlations runs into severe practical problems even for populations of $N \sim 10$ neurons (Martignon et al., 1995, 2000; Del Prete et al., 2004; Shlens et al., 2006; Montani et al., 2009). The severeness of this limitation is further underscored by the fact that the significance of higher-order correlations computed from small populations ($N \sim 10$; Schneidman et al., 2006; Shlens et al., 2006) can generally not be extrapolated to large populations (Roudi et al., 2009). Taken together, while recent progress in experimental technique allows for the simultaneous recording of the spiking activity of tens to hundreds nerve cells, a faithful statistical description of the resulting activity that includes correlations of higher order is greatly hampered by the limitations of available data analysis techniques.

We have recently presented a novel method for a cumulant-based inference for the presence of higher-order correlations (CuBIC) that avoids the need for extensive sample sizes (Staude et al., 2007, 2009). Instead of directly estimating correlation parameters from all subgroups, CuBIC aims only at population-average correlations, estimated via the cumulants of the pooled and discretely sampled spiking activity of all recorded neurons (population spike counts). The presence of higher-order correlations is then inferred from measured cumulants of low order by exploiting certain constraining relations among correlations of different orders in a statistical model of correlated spiking. CuBIC avoids the direct estimation of higher-order correlations, but decides whether or not lower order cumulants require the presence of higher-order correlations. Focusing on such less specific questions drastically reduces the requirements with respect to the sample size: when applied to artificial data, CuBIC reliably infers higher-order correlations from large ($N \gtrsim 100$), even weakly correlated populations (pairwise correlation coefficient $c \sim 0.01$) that were generated with reasonable sample sizes ($T < 100$ s, Staude et al., 2009).

As a statistical model, CuBIC employs the compound Poisson process (CPP), where correlations are induced by the insertion of coincident events in continuous time, i.e., before binning is applied (Ehm et al., 2007; Johnson and Goodman, 2007; Brette, 2009; Staude et al., 2010). Interestingly, this model of correlation fits perfectly to measuring (higher-order) correlations via connected cumulants of the binned spike trains (Staude et al., 2010), a common framework for (higher-order) correlation measures. The simple relationship of the unknown model parameters, i.e., the orders of correlation present in the data, and the observable cumulants of the population spike count allows to devise null-hypotheses concerning the orders of correlation in the data (the details of the CPP and CuBIC are explained in Section "The Stationary Case"). Combining tests against different null-hypothesis yields a lower bound $\hat{\xi}$ for the maximal order of correlation in the data.

A central assumption in the original presentation of CuBIC (Staude et al., 2009) was that the statistics of spiking in the population does not change over time (stationarity). As both experimental cues and/or internal processes often induce transients or fluctuations of firing rates, this central assumption is frequently violated in electrophysiological data.

In the present study, we describe a non-stationary version of the CPP by decoupling the correlation structure from the spike intensity of the population (see Section "The Non-stationary Case"). Using the "law of total cumulance" we are able to incorporate non-stationarities in firing rate into the computation of the cumulants of the population spike counts. These rate-adjusted cumulants are then used to adapt CuBIC to infer higher-order correlations also from non-stationary data. This adaptation requires a specification of the kind of non-stationarity in terms of a parametric family of distributions for the bin-wise mean firing rates (the "carrier distribution"). Allowing for uniform rate fluctuations, for instance, yields as a result that the data must have correlations of some minimal order $\hat{\xi}$ even if firing rates fluctuated uniformly from bin to bin. In this sense, the choice of a family for the carrier distribution implies a demarcation line between "genuine" correlation and "artifacts" due to rate (co-) variation (Staude et al., 2008). Numerical simulations reveal that the adaptation corrects for false positive inference of correlations in data with pure rate co-variation, while allowing for potential variations in firing rates has a surprisingly small effect on CuBIC's sensitivity for correlations (see Case Studies). Furthermore, we find that a perfect match between the true carrier family and the family allowed in the adapted CuBIC does not seem to be fundamentally important to guarantee reliable test performance.

THE STATIONARY CASE

CUMULANTS AND THE COMPOUND POISSON PROCESS

Population spike count

The basic observable of this study is the pattern vector $\mathbf{X}(s) := (X_1(s),\ldots,X_N(s))$, where $X_i(s)$ is the discretized spike count of the ith neuron in the bin $[sh,(s+1)h)$ of width h (a complete list of symbols is provided in Section "List of Symbols" in Appendix). Given $\mathbf{X}(s)$, we define the population spike count $Z(s)$ as the total number of spikes in the population in the sth bin (**Figure 1**)

$$Z(s) = \sum_{i=1}^{N} X_i(s).$$

In the case where the X_i are binary ("1" for one or more spikes in the bin, "0" for no spike), $Z(s)$ is simply the number of neurons that spike in the sth bin. As opposed to other frameworks for correlation analysis (e.g., Aertsen et al., 1989; Martignon et al., 1995; Grün et al., 2002a; Nakahara and Amari, 2002; Shlens et al., 2006), however, the method presented in this study does not assume binary variables.

We here assume that $Z(s)$ and $Z(s + k)$ are independent for $k \neq 0$ (zero memory). Furthermore, let us for now assume that the distribution of $Z(s)$ does not depend on the time bin s (stationarity). This critical assumption will be relaxed in Section "The Non-stationary Case".

Correlations and cumulants

In the present framework, correlations among the variables X_i are measured by mixed or "connected" cumulants. Like the more familiar (raw) moments $\mathrm{E}[Z^m]$ of a random variable Z, the univariate cumulants $\kappa_m[Z]$ characterize the shape of its distribution (see, e.g., Stratonovich, 1967; Gardiner, 2003). For the first two cumulants, the expectation value and the variance, the latter can be expressed in terms of the former by the well-known expressions $\kappa_1[Z] = \mathrm{E}[Z]$ and $\kappa_2[Z] = \mathrm{E}[Z^2] - \mathrm{E}[Z]^2 = \mathrm{Var}[Z]$. Similar equalities for higher cumulants are exceedingly complicated, but algorithms for their computations are available (see Stuart and Ord, 1987 for explicit expressions for $m \leq 10$, Section "Cumulants of the Non-stationary CPP" for a straightforward, and Di Nardo et al., 2008 for a more advanced algorithm). For notational consistency, we will from now on use the cumulant notation, e.g., use the terms "first/second cumulant" instead of the more familiar "mean/variance".

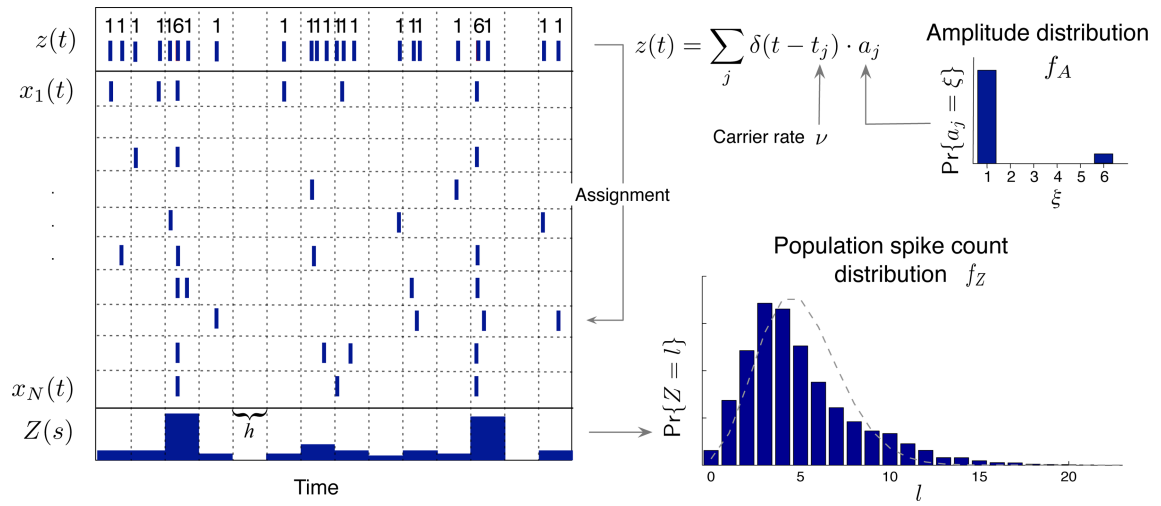


FIGURE 1 | Schema of the compound Poisson process and its measurement. Left: spike event times (horizontal bars) of individual neurons $x_1(t),\ldots,x_N(t)$ and tick marks of the carrier process $z(t)$ (top) with the associated amplitudes (numbers above the ticks). The population spike count $Z(s)$ (below the spike trains) counts the number of spikes across all neurons in bins of width h (dotted lines). Right: distributions of the amplitudes a_j, f_A (top) and of the population spike counts, f_Z (bottom; blue bars: f_Z from 100 s of data with the given amplitude distribution, estimated using a bin size of $h = 5$ ms; dashed line: Poisson fit, corresponding to an independent population with the same firing rates). To construct a population of correlated spike trains, amplitudes a_j are drawn for all events t_j in the carrier process i.i.d from f_A. Individual processes $x_i(t)$ are constructed by assigning subsequent events of the carrier process $z(t)$ into a_j "child" processes $x_i(t)$ (here, events are assigned randomly to the specific process IDs). Correlations of order ξ are induced, whenever events in the carrier process are copied into more than ξ processes, i.e., if the amplitude distribution assigns non-zero probabilities for amplitudes ≥ ξ.

Multivariate, or "connected", cumulants arise when the variable under consideration is a sum of correlated variables. For $m = 2$ and $Z = \Sigma_{i=1}^{N} X_i$, for instance, we have the well-known formula

$$\kappa_2[Z] = \mathrm{Var}[Z] = \mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{Var}[X_i] + \sum_{i\neq j} \mathrm{Cov}[X_i, X_j]$$
$$=: \sum_{i=1}^{N} \kappa_2[X_i] + \sum_{i\neq j} \kappa_{1,1}[X_i, X_j] \quad (1)$$

Hence, the second-order cumulant correlations $\mathrm{Cov}[X_i,X_j] = \kappa_{1,1}[X_i,X_j]$ measure the degree of additive linearity of $\kappa_2[Z]$. Higher-order cumulant correlations are generalizations of the covariance in exactly this sense, and mth order correlations arise when $\kappa_m[Z]$ is decomposed into expressions involving the individual X_i. The following definition fixes the notation used in the remainder of this study, for a precise definition we refer to the literature (e.g., Stuart and Ord, 1987; Gardiner, 2003; Staude et al., 2009; for details on cumulant correlations see also Streitberg, 1990; Staude et al., 2010).

Definition 1 *Let* $\mathbf{X} = (X_1,\ldots,X_N)$ *be an N-dimensional random variable, e.g., the spike counts of N parallel spike trains, let* $M = \{m_1,\ldots,m_\kappa\}$ *be a subset of* $\{1,\ldots,N\}$ *of size k, and denote by* $\sigma(M) \in \{0,1\}^N$ *the binary indicator vector of the set M, whose ith component is 1 if* $i \in M$ *and 0 otherwise. Then we measure kth order correlations among* $(X_{m_1},\ldots,X_{m_k})$ *by the connected cumulant* $\kappa_{\sigma(M)}[\mathbf{X}]$. *We say that* $\mathbf{X}$ *has correlations of order k if and only if at least one kth order connected cumulant of* $\mathbf{X}$ *is non-zero.*

The following generalization of Eq. 1 is a straightforward consequence of the construction of connected cumulants (Staude et al., 2010).

Theorem 1 *The mth cumulant* $\kappa_m[Z]$ *of* $Z = \Sigma_{i=1}^{N} X_i$ *depends on the summed correlations among the* X_i *of all orders* $\leq m$, *but is independent of correlations of orders* $> m$.

By the above theorem, $\kappa_m[Z]$ is a measure for the total correlation in the population of all orders $\leq m$. While a correction of the second cumulant for the influence of the single process statistics would be straightforward (subtracting $\Sigma_{i=1}^{N}\mathrm{Var}[X_i]$ in Eq. 1, see Staude et al., 2009), correcting higher cumulants for the influence of correlations of lower order is exceedingly complicated. We therefore employ a parametric model for Z, the CPP (see next section), the parameters of which can be interpreted straightforwardly in terms of higher-order correlations among the X_i.

Before a discussion of our model, we wish to stress that the cumulant correlations presented here do not comply with the interaction parameters of the more familiar log-linear model. In particular, data sets can have higher-order log-linear interactions without having higher-order cumulant correlations and vice versa (see Staude et al., 2010 for concrete examples, and e.g., Darroch and Speed, 1983; Streitberg, 1990; Staude et al., 2009 for more general discussions).

The compound poisson process

As opposed to the *discretized*, binned population spike count $Z(s)$ of the previous section, the proposed model operates in *continuous*, i.e., unbinned time. That is, we model the process $z(t) = \Sigma_{i=1}^{N} x_i(t)$, where $x_i(t) = \Sigma_j \delta(t - t_j^i)$ denotes the ith unbinned, continuous-time spike train with spike event times t_j^i ($i = 1,\ldots,N, j \in \mathbb{N}$). The model we propose for $z(t)$ is that of a CPP

$$z(t) = \sum_j \delta\left(t - t_j\right) a_j, \quad (2)$$

where the event times t_j constitute a Poisson process, and the marks a_j are i.i.d. integer-valued random variables, drawn independently for all t_j (**Figure 1**, left). The marks a_j determine the number of neurons that fire at time t_j, and will be referred to as the "amplitude" of the event at time t_j. The probability that an event has a specific amplitude is determined by the amplitude distribution f_A, i.e., $f_A(\xi) = \Pr\{a_j = \xi\}$ for all $j \in \mathbb{N}$ (**Figure 1**, top right). The Poisson process that generates the events t_j is called the "carrier process" of the model and its rate ν is the "carrier rate". Processes of this type are also referred to as generalized, or marked, Poisson processes (see e.g., Snyder and Miller, 1991 for a general definition, and Ehm et al., 2007 for an alternative application to spike train analysis).

With the above model, the generation of a population of spike trains proceeds in two steps. First, realize a Poissonian carrier processes $m(t) = \Sigma_j \delta(t = t_j)$ and draw for each of its events t_j an i.i.d. amplitude a_j from the amplitude distribution f_A. In the second step, assign the spike at t_j to a_j individual processes, where the process IDs are determined from a separate "assignment distribution". The simplest scenario assumes uniform assignment, where the a_j neuron IDs that receive the spike at t_j are drawn randomly from $\{1,\ldots,N\}$, resulting in a homogeneous population. As CuBIC only aims for a lower bound on the order of correlation, irrespective of the neuron IDs that realize these correlations, we here ignore the assignment distribution, and focus on the amplitude distribution only. The following theorem clarifies the relationship between cumulant correlations and the amplitude distribution in the framework of the CPP (see Staude et al., 2009 for a proof).

Theorem 2 *Let $z(t) = \Sigma_{i=1}^{N} x_i(t)$ be a CPP with amplitude distribution f_A and carrier rate ν, and let $\mathbf{X} = (X_1,\ldots,X_N)$ be the vector of counting variables obtained from the $x_i(t)$ with bin width h. Then*:

1. *The components of* **X** *have correlations of order m (in the sense of Definition 1) if and only if f_A assigns non-zero probabilities to amplitudes $\geq m$.*
2. *With $\mu_m := \mathrm{E}[A^m] = \Sigma_{k=1}^{N} k^m f_A(\xi)$, the cumulants of Z are given as*

$$\kappa_m[Z] = \mu_m \nu h. \tag{3}$$

Note that correlations in the above theorem are measured strictly on the basis of the discretized counting variables X_i. As a consequence, they do not resolve (and do not depend on) the perfect temporal precision of the coincident events in the CPP. That is, if the events of $z(t)$ were assigned to the individual processes with a temporal jitter that is small with respect to the bin size h, the effect of the jitter on the correlations is negligible.

Now let $\xi \leq N$ be the maximal order of correlation in the model, i.e., $f_A(k) = 0$ for $k > \xi$, and denote by ν_k $(k = 1,\ldots\xi)$ the compound rates of events of amplitude k, i.e., $\nu_k = \nu \cdot f_A(k)$. Then Eq. 3 can be written as

$$\kappa_m[Z] = \vec{\xi}_m \cdot \vec{\nu}_\xi h, \tag{4}$$

where $\vec{\xi}_m = (1^m,\ldots,\xi^m)$, $\vec{\nu}_\xi := (\nu_1,\ldots,\nu_\xi)$, and $\vec{\xi}_m \cdot \vec{\nu}_\xi := \Sigma_{i=1}^{\xi} i^m \nu_i$ is the vector dot product.

CuBIC

This section summarizes the stationary version of CuBIC to the extent that is needed to understand its adaptation to non-stationary populations (see Staude et al., 2009 for details). In brief, CuBIC quantifies the following thought experiment. Consider the situation of four simultaneously recorded neurons, where all neuron pairs have a correlation coefficient of $c = 1$. As $c = 1$ implies identity for all pairs of spike trains, all four spike trains must in fact be identical. In the framework of the CPP, this translates to the existence of events of amplitude $a_j = 4$, and hence correlation of order 4 (Theorem 1). This illustrates that it is possible, in principle, to infer the existence of fourth order correlations from estimated pairwise correlations. In Staude et al. (2009), this inference was generalized by a hierarchy of statistical hypothesis tests $H_0^{m,\xi}$, labeled by the order m of the correlation estimated from the population spike count, and the test parameter ξ, which indicates the maximal order of correlation allowed in the null hypothesis. For given m and ξ, the rejection of a hypothesis $H_0^{m,\xi}$ means that estimated correlations of order m in the data imply the presence of correlations of at least order $\xi + 1$. Combining tests for different values ξ then provides $\hat{\xi}_m = \max\{\xi \mid H_0^{m,\xi} \text{ is rejected}\} + 1$ as a lower bound on the order of correlation in the data. In the thought experiment above, we estimated pairwise correlation, hence $m = 2$, and rejected tests with $\xi = 1,2,3$, such that $\hat{\xi}_2 = 4$. In principle, the order of the estimated correlation m is a free parameter. However, as shown in Staude et al. (2009), tests with $m = 3$ are already extremely sensitive, such that we will present both the stationary CuBIC and the non-stationary adaptation only for the case $m = 3$.

Assume one is given the first three cumulants of a population spike count variable Z'. Then, for a fixed value of the test parameter ξ, consider the following constrained maximization problem

$$\kappa_{3,\xi}^{*} := \max_{\nu, f_A} \{\kappa_3[Z]\} \tag{5}$$

subject to $\kappa_2[Z'] = \kappa_2[Z]$

$$\kappa_1[Z'] = \kappa_1[Z]$$
$$f_A(k) = 0 \text{ for } k > \xi,$$

where Z is the population spike count of a model with parameters ν and f_A. The model that solves Eq. 5 has the maximal third cumulant, i.e., triplet correlations, among all models that do not have any correlations beyond order ξ, and have the same population-averaged first- and second-order properties, i.e., firing rates and pairwise correlations, as the given spike count variable Z'. As a consequence, $\kappa_3 > \kappa_3[Z']$ implies that the third order correlations in Z' cannot be realized with correlations of orders $\leq \xi$. Thus Z' must have correlations of order $\geq \xi + 1$.

To solve Eq. 5, we use Eq. 4 and obtain the equivalent problem

$$\kappa_{3,\xi}^{*} = \max_{\vec{\nu}_\xi} \left\{ \vec{\xi}_3 \cdot \vec{\nu}_\xi h \right\} \tag{6}$$

subject to $\kappa_2[Z'] = \vec{\xi}_2 \cdot \vec{\nu}_\xi h$

$$\kappa_1[Z'] = \vec{\xi}_1 \cdot \vec{\nu}_\xi h.$$

In Eq. 6, the objective function and the constraints depend linearly on the model parameters $\vec{\nu}_\xi$. Problems of this type, so-called Linear Programming Problems, are uniquely solvable, e.g., by the Simplex

Method or any of its variants (Press et al., 1992, Chapter 10.8). The solution yields an upper bound for the third cumulant $\kappa^*_{3,\xi}$ and the corresponding parameter vector $\vec{\nu}^*_\xi$. As it turns out, the only non-zero components of the solver $\vec{\nu}^*_\xi = (\nu^*_1, \ldots, \nu^*_\xi)$ are ν^*_1 and ν^*_ξ, i.e., $\nu^*_k = 0$ for all $k \notin \{1, \xi\}$ (see The Solution of the Maximization in Appendix). The carrier rate and amplitude distribution of the CPP that maximizes Eq. 6 are then given by $\nu^* = \nu^*_1 + \nu^*_\xi$ and $f_{A^*}(l) = \nu^*_l / \nu^*$.

With the solution of Eq. 5, the null hypothesis $H_0^{3,\xi}$ is

$$H_0^{3,\xi} : \kappa_3[Z'] \leq \kappa^*_{3,\xi}$$

To test a sample $\{Z'\} = \{Z'_1, \ldots, Z'_L\}$ against $H_0^{3,\xi}$, we estimate its cumulants by the so-called k-statistics k_m (Stuart and Ord, 1987; the well-known sample mean and unbiased sample variance are the first two k-statistics). To derive the required distribution of the test statistic k_3 under $H_0^{3,\xi}$, we assume that Z' is the population spike count of the model with parameters ν^* and f^*_A, i.e., the solution of Eq. 6 after the unknown cumulants $\kappa_1[Z']$ and $\kappa_2[Z']$ have been replaced by their estimates k_1 and k_2. Thus, under $H_0^{3,\xi}$ the distribution of k_3 has expectation value $\kappa^*_{3,\xi}$, and its variance is given as (e.g., Stuart and Ord, 1987)

$$\mathrm{Var}[k_3] = \frac{\kappa_6[Z']}{L} + 9\frac{\kappa_2[Z']\kappa_4[Z']}{L-1} + 6\frac{\kappa_2[Z']^3}{(L-1)(L-2)}, \tag{7}$$

where $\kappa_i[Z']$ are the cumulants of Z', obtained by inserting ν^* and f^*_A into Eq. 3. Finally, with sample sizes of $L > 10000$ (corresponding to a data set of 10 s duration, sampled with a bin width of $h = 1$ ms), the distribution of k_3 is well approximated by a normal distribution, such that the p-value of $H_0^{3,\xi}$ is given by

$$p_{3,\xi} = \int_{k_m}^{\infty} \frac{1}{\sqrt{2\pi \mathrm{Var}[k_3]}} \exp\left(-\frac{\left(t - \kappa^*_{3,\xi}\right)^2}{2\mathrm{Var}[k_3]} \right) dt. \tag{8}$$

As mentioned above, the rejection of a specific hypothesis $H_0^{3,\xi}$ implies that the data have correlations beyond order ξ or, in other words, that $\xi + 1$ is a lower bound for the order of correlation. The final result of CuBIC is the maximum of these lower bounds

$$\hat{\xi} := \max\{\xi \mid p_{3,\xi} < \alpha\} + 1, \tag{9}$$

where α is a predefined test level (see Staude et al., 2009 for details).

THE NON-STATIONARY CASE

THE MODEL

In the previous section, the CPP was presented as a model for populations with constant firing rates. However, (co-)variations in firing rate are a common feature of neuronal populations. To incorporate potential non-stationarities into the CPP, recall its central ingredients: the intensity of the population is described by the carrier rate ν, the population-averaged correlation structure is determined by the amplitude distribution f_A, and the precise composition of spikes and correlations within the population is determined by the assignment distribution. Given this parametrization, non-stationarities can in principle be included into the CPP by allowing a time-dependent carrier rate, amplitude distribution, assignment distribution, or combinations thereof.

Rather than presenting a general model for non-stationary populations, the focus of this study is to adapt the analysis technique CuBIC for potential variations in the firing rates. CuBIC, however, aims only at the population-average correlation structure f_A, and inference is based on the population spike count Z. Time dependencies in the assignment distribution only change the neuron IDs that realize correlations over time, but do no alter the order of correlations. As a consequence, the population spike count Z is not influenced by potential temporal variations of the assignment distribution (see **Figure 2A**). Furthermore, CuBIC aims only at a lower bound for the maximal order of correlation, not on the precise values of correlations of different orders. As such, it aims for the largest entry with non-zero probability in the amplitude distribution, irrespective of whether this entry was present in the whole data stretch or if it occurred only within a short period. Taken together, CuBIC is blind for non-stationarities in the amplitude distribution and the assignment distribution. We therefore assume both these objects to be constant over time and consider only time-varying carrier rates $\nu(t)$ (see top panels in **Figures 2B,C**).

CUMULANTS OF THE NON-STATIONARY CPP

To relate the model parameters to the cumulants of Z for time varying $\nu(t)$, observe that the value of $Z(s)$ in the window $[sh, (s+1)h)$ does not depend on the precise time course of the carrier rate $\nu(t)$ in this window, but only on the integral $R_s := 1/h \int_{sh}^{(s+1)h} \nu(t)dt$. Substituting the carrier rate $\nu(t)$ with a piecewise constant function whose value in the interval $[sh, (s+1)h)$ is R_s thus results in an identical population spike count Z. Furthermore, CuBIC ignores the temporal order of $Z(1), \ldots, Z(L)$ and assumes subsequent values of Z to be i.i.d. variables. As a consequence, CuBIC is also blind for the temporal order of the rate values R_s, and we therefore assume them to be i.i.d. with a common "carrier distribution" f_R (compare panels B and C of **Figure 2**).

The above setting characterizes the population spike count Z as a parameter-dependent random variable, where the outcome of Z in the sth bin depends not only on the outcome of the CPP realization, but also on the (random) value of the rate variable R_s. For such "doubly stochastic" variables, the raw moments are given as

$$\mu_m[Z] := \mathrm{E}[Z^m] = \mathrm{E}^R\left[\mathrm{E}[Z^m \mid R]\right] \quad (m \in \mathbb{N}),$$

where the inner expectation is the expectation value of Z^m for a given value of the rate R, and the outer expectation is with respect to the distribution f_R. Now recall the definition of the moments as the coefficients of the Taylor series expansion of the characteristic function

$$\phi_Z(s) := \mathrm{E}\left[e^{isZ}\right] \tag{10}$$

$$= \sum_m i^m \frac{s^m}{m!} \mu_m[Z], \tag{11}$$

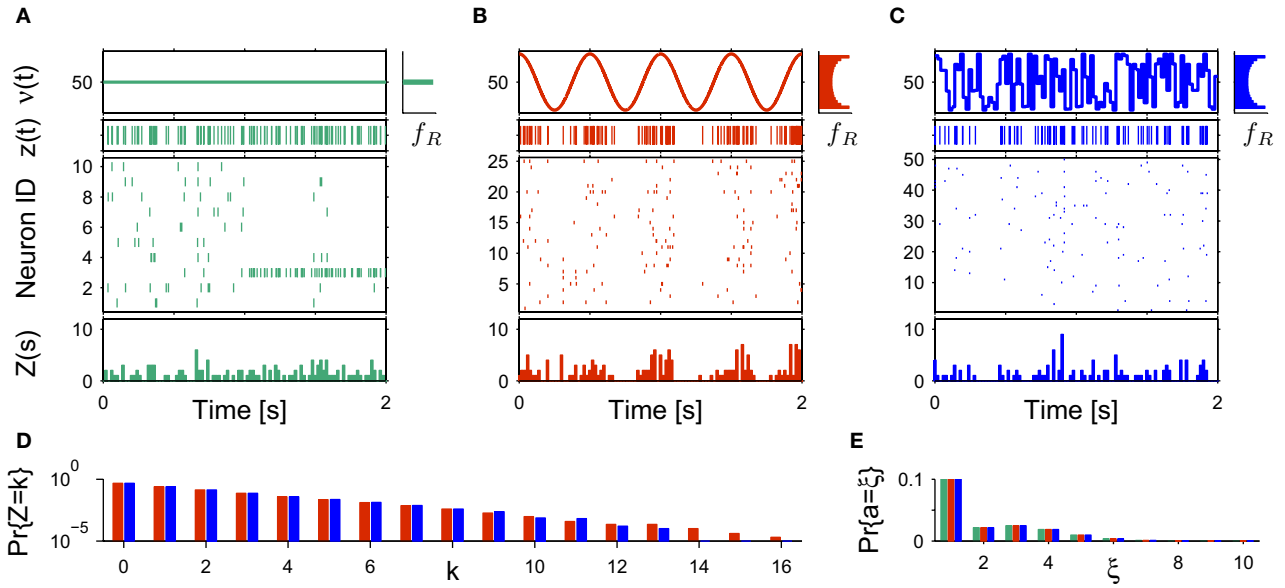


FIGURE 2 | Non-stationarities in the CPP. Panels **(A–C)** show carrier rates $\nu(t)$ in Hz (top panels), the distributions of the bin-wise mean carrier rates ("carrier distributions" f_R, small panels on the right; bin size $h = 20$ ms), the event-times of the carrier process $z(t)$ (second panels from top), raster plots (third panels from top) and population spike counts $Z(s)$ (bottom panels) for three different data sets. **(A)** Time varying assignment distribution produces non-stationary single processes, although the carrier process $z(t)$ is stationary (constant carrier rate $\nu(t) = 50$ Hz; all carrier events with $t_j < 1$ s are assigned randomly, all carrier events with $t_j \geq 1$ and amplitude $a_j = 1$ are assigned to neuron 3). **(B)** Cosine carrier rate results in non-stationary carrier process $z(t)$, the subsequent uniform assignment to $N = 25$ neurons generates a homogeneous, non-stationary population. **(C)** The carrier rate is constant in bins of length $h = 20$ ms and subsequent values of $\nu(t)$ are i.i.d. realizations with the carrier distribution of the cosine carrier rate in **(B)**. Carrier events in **(C)** are assigned uniformly to $N = 50$ neurons. As illustrated by the virtually identical population spike count distributions shown in **(D)** (estimated from 1000 s of artificial data, color code as in **B,C**; note logarithmic y-scale), the differences in both the carrier rate, in particular the temporal order of the bins, and the assignment, in particular the number of neurons of the generated population, do not influence the statistics of the population spike count Z. **(E)** Amplitude distributions of all three data sets are superpositions of a "background-peak" at $\xi = 1$ and a binomial distribution $B(10,0.3)$ (color code as in **A–C**).

such that the moments can be obtained from ϕ_Z via

$$\mu_m[Z] = \frac{1}{i^m}\frac{\partial^m \phi_Z(s)}{\partial s^m}\bigg|_{s=0}.$$

Analogously, cumulants are the coefficients of the logarithm of the characteristic function

$$\psi_Z(s) = \log \mathrm{E}[e^{isZ}] \tag{12}$$

$$= \sum_m i^m \frac{s^m}{m!}\kappa_m[Z], \tag{13}$$

such that

$$\kappa_m[Z] = \frac{1}{i^m}\frac{\partial^m \psi_Z(s)}{\partial s^m}\bigg|_{s=0} \tag{14}$$

A straightforward strategy to relate cumulants to moments is to insert Eq. 11 into the right hand side of Eq. 12, writing the logarithm as a power series, and collecting coefficients for identical powers of s. This procedure illustrates in particular that the mth cumulant is a function of the first m raw moments only, and, reversely, that the mth moment can be expressed as a function of the first m cumulants. We denote the maps relating cumulants to moments and moments to cumulants by F_m and G_m, respectively, such that

$$\mu_m[Z] = F_m\left(\kappa_1[Z],\ldots,\kappa_m[Z]\right) \tag{15}$$

$$\kappa_m[Z] = G_m\left(\mu_1[Z],\ldots,\mu_m[Z]\right). \tag{16}$$

With these maps at hand, the mth cumulant of the non-stationary CPP can be computed by the following procedure.

1. For $i = 1,\ldots,m$ express the conditional moment $\mu_i[Z \mid R]$ in terms of the first i cumulants, i.e., write $\mu_i[Z \mid R] = F_i(\kappa_1[Z \mid R],\ldots,\kappa_i[Z \mid R])$.
2. Apply Eq. 3 to the individual cumulants, such that $\mu_i[Z \mid R] = F_i(\mu_1 Rh,\ldots,\mu_i Rh)$, where, as before, $\mu_i = \mu_i[A]$ are the moments of the amplitude distribution f_A.
3. Compute the mth cumulant of Z by applying G_m

$$\kappa_m[Z] = G_m\left(F_1(\mu_1 Rh),\ldots,F_m(\mu_1 Rh,\ldots,\mu_m Rh)\right). \tag{17}$$

The results are summarized as the "law of total cumulance". The first three orders read

$$\kappa_1[Z] = \mu_1\kappa_1[R]h \tag{18}$$

$$\kappa_2[Z] = \mu_2\kappa_1[R]h + \mu_1^2\kappa_2[R]h^2 \tag{19}$$

$$\kappa_3[Z] = \mu_3\kappa_1[R]h + \mu_1^3\kappa_3[R]h^3 + 3\mu_1\mu_2\kappa_2[R]h^2. \tag{20}$$

A similar parameter transformation that leads from Eq. 3 to Eq. 4 simplifies Eqs. 18–20. In slight abuse of notation, we use the same symbol for the *expected* compound rates as for the constant compound rates of Eq. (4), i.e., write $\nu_k := f_A(k)\cdot\kappa_1[R]$. Using the

standardized cumulants of the rate variable $\beta_k := \kappa_k[R]/\kappa_1[R]^k$ and the vector notation $\vec{\nu}_\xi := (\nu_1,\ldots,\nu_\xi)$ and $\vec{\xi}_m = (1^m,\ldots,\xi^m)$, Eqs. 18–20 can be written as

$$\kappa_1[Z] = \vec{\xi}_1 \cdot \vec{\nu}_\xi h \tag{21}$$

$$\kappa_2[Z] = \vec{\xi}_2 \cdot \vec{\nu}_\xi h + \left(\vec{\xi}_1 \cdot \vec{\nu}_\xi\right)^2 h^2 \beta_2 \tag{22}$$

$$\kappa_3[Z] = \vec{\xi}_3 \cdot \vec{\nu}_\xi h + \left(\vec{\xi}_1 \cdot \vec{\nu}_\xi\right)^3 h^3 \beta_3 + 3\left(\vec{\xi}_1 \cdot \vec{\nu}_\xi\right)\left(\vec{\xi}_2 \cdot \vec{\nu}_\xi\right) h^2 \beta_2 \tag{23}$$

CuBIC FOR NON-STATIONARY DATA

To adapt CuBIC to non-stationary populations, we need to formulate the general maximization problem (Eq. 5) for the case of time-dependent carrier rates. Using Eqs. 21–23 instead of Eq. 4, we obtain

$$\kappa_{3,\xi}^{*,ns} = \max_{\vec{\nu}_\xi,\beta_2,\beta_3} \left\{ \vec{\xi}_3 \cdot \vec{\nu}_\xi h + \left(\vec{\xi}_1 \cdot \vec{\nu}_\xi\right)^3 h^3 \beta_3 + 3\left(\vec{\xi}_1 \cdot \vec{\nu}_\xi\right)\left(\vec{\xi}_{12} \cdot \vec{\nu}_\xi\right) h^2 \beta_2 \right\} \tag{24}$$

$$\text{subject to } \kappa_2[Z'] = \vec{\xi}_2 \cdot \vec{\nu}_\xi h + \left(\vec{\xi}_1 \cdot \vec{\nu}_\xi\right)^2 h^2 \beta_2$$

$$\kappa_1[Z'] = \vec{\xi}_1 \cdot \vec{\nu}_\xi h,$$

with $\vec{\nu}_\xi \in [0,\infty)^\xi$, $\beta_2 \in [0,\infty)$ and $\beta_3 \in (-\infty,\infty)$. After some algebra and substitution of the unknown cumulants $\kappa_i[Z']$ by their estimates k_i, we arrive at the equivalent problem

$$\kappa_{3,\xi}^{*,ns} = \max_{\vec{\nu}_\xi,\beta_2,\beta_3} \left\{ \vec{\xi}_3 \cdot \vec{\nu}_\xi h + k_1^3 \beta_3 - 3k_1^3 \beta_2^2 + 3k_1 k_2 \beta_2 \right\} \tag{25}$$

$$\text{subject to } k_2 = \vec{\xi}_2 \cdot \vec{\nu}_\xi h + k_1^2 \beta_2$$

$$k_1 = \vec{\xi}_1 \cdot \vec{\nu}_\xi h.$$

As opposed to the Linear Programming Problem of the stationary case (Eq. 6), the constraints in Eq. 25 do not apply to all free variables: the third standardized cumulant of the rate variable β_3 can, in principle, be arbitrarily large. As a consequence, also the objective function in Eq. 25 is unbounded. We therefore have to impose additional constraints on the carrier distribution f_R in order to ensure convergence of the maximization. The approach taken here is to prescribe a two-dimensional parametric family for the distribution of the rate variable, such that its third (standardized) cumulant can be expressed in terms of its first two cumulants. The choice for the family of carrier distributions determines the form of the objective function. Let us present two specific cases (see Section "Discussion" for more details on the role of this choice).

Symmetric carrier distributions

If the carrier distribution is symmetric about its mean, like e.g., the uniform distribution, we can exploit the fact that the third cumulant $\kappa_3[R]$ of symmetric distributions vanishes[1]. As a consequence, also $\beta_3 = \kappa_3[R]/\kappa_1[R]^3 = 0$. The objective function of the maximization problem (Eq. 25) for symmetric carrier distributions thus becomes

$$F\left(\nu_1,\ldots,\nu_\xi,\beta_2\right) = \vec{\xi}_3 \cdot \vec{\nu}_\xi h + 3\beta_2 k_1 k_2 - 3k_1^3 \beta_2^2. \tag{26}$$

This objective function is linear in the (expected) compound rates ν_k ($k = 1,\ldots,\xi$) and quadratic in β_2. As the constraints are also linear, using Eq. 26 as the objective function in Eq. 25 yields a convex quadratic maximization problem. Problems of this type have a unique solution, and effective implementations of numerical solvers are available.

Gamma-distributed carrier rate

If the carrier variable R follows a Gamma-distribution, the third cumulant can be expressed in terms of the first two as $\kappa_3[R] = 2\kappa_2[R]^2/\kappa_1[R]$. For the normalized third cumulant β_3 we thus have $\beta_3 = 2\kappa_2[R]^2/\kappa_1[R]^4 = 2\beta_2^2$. The objective function then reads

$$F\left(\nu_1,\ldots,\nu_\xi,\beta_2\right) = \vec{\xi}_3 \cdot \vec{\nu}_\xi h + 3\beta_2 k_1 k_2 - k_1^3 \beta_2^2. \tag{27}$$

As for symmetric carrier distributions, the objective function in Eq. 25 is linear in the expected component rates ν_k ($k = 1,\ldots,\xi$) and quadratic in β_2. Thus, the resulting problem is also a convex quadratic programming problem and can be solved with the same solvers.

We wish to stress once again that the choice made here concerns only the family of the carrier distribution, not its particular shape. That is, if we chose e.g., a uniform distribution, we only determine that $\kappa_3[R] = 0$ but do not fix the support of the distribution. Finally, we note that the only non-zero components of the solution $\vec{\nu}_\xi^* = \left(\nu_1^*,\ldots,\nu_\xi^*\right)$ to the maximization problem are ν_1^* and ν_ξ^*, i.e., $\nu_k^* = 0$ for all $k \notin \{1,\xi\}$, as in the stationary case (see The Solution of the Maximization in Appendix).

Variance of test statistic

As a final ingredient, CuBIC requires the variance of the test statistic k_3 in order to be able to compute the p-values. Assuming that the solution of Eq. 25 is available, we compute the second, fourth and sixth cumulant of the solution by inserting the solving $\vec{\nu}_\xi^*$ and β_2^* into the algorithm leading to Eq. 17. Plugging these cumulants into Eq. 7 yields $\mathrm{Var}[k_3]$, the variance of the test statistic k_3 under the non-stationary null-hypothesis. Insertion into Eq. 8 yields the corresponding p-value.

CASE STUDIES

To illustrate the application of the adapted CuBIC, we consider here two typical experimental situations. In the first situation, data are recorded in early sensory systems, where characteristic firing rate profiles closely follow the stimulus. In such a scenario, information about the rate distribution can be obtained from the type of the stimulus. In the second situation, there is no direct relation between the experimental paradigm and non-stationarities in firing rates. In this case, ad hoc assumptions of the family of carrier rate are the only option. We illustrate both scenarios and the steps required when applying the non-stationary version of CuBIC by analyzing simulated spike trains.

STIMULUS-DRIVEN NON-STATIONARITY

Pure non-stationarity

Our first example mimics a recording in visual cortex with a oriented sinusoidal grating as stimulus. We model the population response in such an experiment by a CPP model with sinusoidal carrier rate $\nu(t)$, i.e.,

$$\nu(t) = B + C\cos(2\pi\omega t - d), \tag{28}$$

where B is the offset, $C \leq B$ is the amplitude of the modulation, ω is the temporal frequency of the stimulus and d is the phase, i.e., the sum of the stimulus phase and the delay it takes for the stimulus-driven activity to propagate to the recording electrodes

[1]Generally, $\kappa_3[R]$ is a measure for the skewness of f_R, such that $\kappa_3[R] < 0$ for left-skewed, $\kappa_3[R] = 0$ for symmetric, and $\kappa_3[R] > 0$ for right-skewed distributions.

(**Figure 3**, top panels of left and right columns). The first data set of this example models the case of pure rate non-stationarity (**Figure 3**, left column). The amplitude distribution is a delta-peak at $\xi = 1$, such that all events of the marked process $z(t)$ have amplitude $a_j = 1$. As a consequence, $z(t)$ is a simple Poisson process with rate $\nu(t)$. Setting $B = C = 500$ and $\omega = 500$ ms, $d = 0$, and assigning the carrier spikes uniformly to a population of $N = 50$ spike trains, we obtain a homogeneous population where each neuron oscillates with a temporal frequency of 2 Hz, assuming firing rates between 0 and 20 Hz (second panel from top of **Figure 3**, left). Note that assigning the events of $z(t)$ to individual spike trains was done only for illustrative purposes, as it does not influence the results of CuBIC.

To analyze the data, we chose a bin size of $h = 5$ ms, computed the population spike count $Z(s)$, and estimated its distribution $f_Z(k) = \Pr\{Z = k\}$ from a total simulation time of $T = 100$ s (**Figure 3**, third and fourth panels from top, respectively). Note that the population spike count Z does not provide unambiguous information about the carrier rate. Next, we applied the stationary version of CuBIC with $m = 3$. That is, we ignored the apparent non-stationarities that are present in the data and computed p-values $p_{3,\xi}$ for $\xi = 1,\ldots,7$ as described in Section "CuBIC" (**Figure 3**, green bars in bottom panel). We found that only the p-value with $\xi = 1$ fell below the chosen significance level of $\alpha = 0.05$ (**Figure 3**, green arrowhead in bottom left panel). Hence, $H_0^{3,1}$ was rejected while all hypotheses with $\xi > 1$ were retained. As a consequence, the stationary CuBIC yields $\hat{\xi} = \max\{\xi \mid p_{3,\xi} < \alpha\} + 1 = 2$ (Eq. 9) as a lower bound for the maximal order of correlation. As the data do not contain correlations beyond those induced by the population-wide non-stationarity, the stationary CuBIC thus leads to false positive inference.

Application of the adapted CuBIC requires the specification of a parametric family of carrier distributions f_R ("carrier family"). We put ourselves in the position of an experimenter, to whom the carrier rate is unknown and cannot be estimated directly; the only observable quantity is the population activity, which is a combination of the carrier rate and potential events of high amplitude. However, as the stimulus was a cosine, we assume also the carrier rate to be a cosine as in Eq. 28, only that the parameters B, C and d are unknown (ω can be estimated from the stimulus frequency). Now recall that the adaptation of CuBIC does not require knowledge of the exact time-course of the carrier rate: what matters is a model for the distribution of the average rate values in the bins $R_s = 1/h \int_{sh}^{(s+1)h} \nu(t)dt$. Given $h << 1/\omega$, we can assume that the sequence of rate values R_s

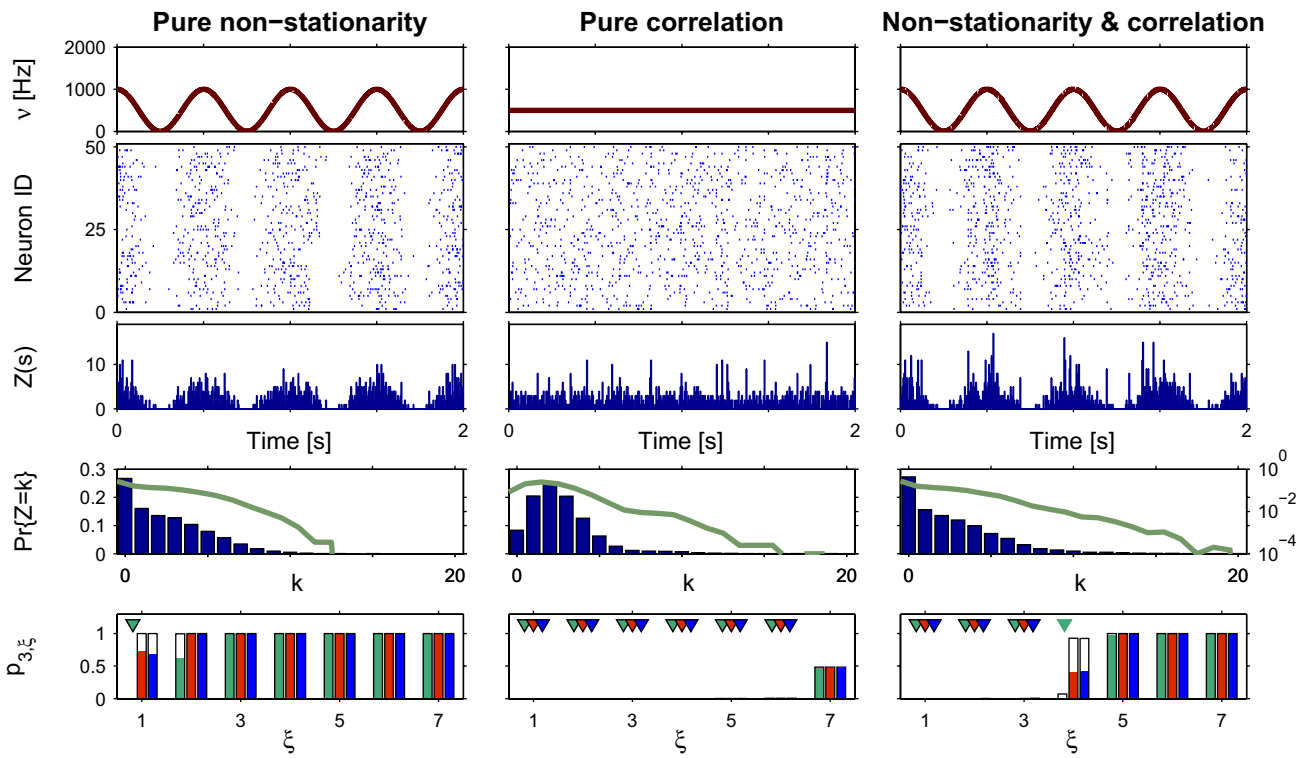


FIGURE 3 | Population statistics and CuBIC results for cosine-like non-stationarity for three data sets. Shown are the carrier rate $\nu(t)$ (top panels), raster plots of $N = 50$ spike trains (second panels from top) and population spike counts $Z(s)$ obtained with a bin width of length $h = 5$ ms (third panels from top) for the first 2 s of a data stretch of length $T = 100$ s. Below are the empirical distributions $f_Z(k) = \Pr\{Z = k\}$ estimated from the entire data set (second panel from bottom; blue bars: linear y-scale with axes on the left, green solid line: logarithmic y-scale with axes on the right). Bottom panels show p-values of the stationary CuBIC (green), the adapted CuBIC with cosine-like rate variations (red) and with bimodal rate variations (blue), where rejected null-hypotheses, i.e., p-values below a significance level of $\alpha = 0.05$, are marked by arrowheads. Outlined bars and arrowheads in bottom panel show results of data where interspike intervals below 2 ms were removed before the analysis. Data with pure rate non-stationarity (left column) have a sinusoidal carrier rate $\nu(t)$ (top panel) and an amplitude distribution with mass concentrated at $\xi = 1$ ($f_A(k) = 0$ for $k > 1$; see text for details). Pure correlation (middle column) is modeled with a stationary carrier rate and correlation up to order 7 ($\nu(t) = const.$, $f_A(k) = 0$ for $k \notin \{1,7\}$). The probability for the high-amplitude events results in a pairwise correlation coefficient of $c = 0.01$ if the events of the carrier process are distributed uniformly among the processes $N = 50$. The combined data set with non-stationarity and correlation (right column) has the same correlation structure as the data in the middle column, and the same carrier rate as in the first column.

samples the cosine faithfully, such that both the phase offset d and the oscillation frequency ω do not influence f_R. In this case, the moments (and cumulants) of all orders can be computed from the parameters B and C in a straightforward manner (see Cumulants of Carrier Distributions in Appendix). In our example, f_R is symmetric, which implies $\beta_3 = 0$ (see Symmetric Carrier Distributions). Furthermore, as the carrier rate $\nu(t)$ has to be non-negative, the model parameters must satisfy $B \geq C$, which implies $\beta_2 \leq 1/2$. The red bars of the lower left panel in **Figure 3** show the p-values computed from Eq. 8, after (a) the solution of stationary problem, $\kappa^*_{3,\xi}$, has been substituted with the solution of Eq. 25 with objective function Eq. 26 and additional constraint $\beta_2 \leq 1/2$, and (b) $\mathrm{Var}[k_3]$ was computed with the algorithm explained in Section "Variance of Test Statistic" with the cumulants of the rate variable R given in Section "Cumulants of Carrier Distributions " in Appendix. We find that $p_{3,\xi} > 0.05$ for all $\xi = 1,\ldots,7$, hence no hypothesis is rejected and the lower bound is $\hat{\xi} = 1$. Thus, the adapted CuBIC does not infer correlation in this data set, and the adaptation successfully corrects for the faulty inference of correlation of the stationary version.

Reduced sensitivity in stationary data?

In the first data set, the stationary CuBIC assigned the rate-generated correlation among the counting variables X_i to events of amplitude ≥2. Allowing for potential (co-)variations of firing rates in the adapted CuBIC corrected for this faulty inference of correlations. Mathematically, allowing for rate (co-)variations allows non-zero values of the parameter β_2 in the maximization of the third cumulant (Eq. 25). Maximizing over a larger parameter space may then increase $\kappa^{*,ns}_{3,\xi}$ as compared to the stationary maximum $\kappa^*_{3,\xi}$. Thus, hypotheses that are rejected by the stationary CuBIC may be retained by the adapted version, as the latter can produce larger maximal third cumulants for a given value of ξ. Consequently, we expected the adapted version to be generally less sensitive for existing events of high amplitude. To investigate this issue, we generated a data set of pure correlation by choosing a constant carrier rate, but allowing for events of amplitude $\xi_{\mathrm{syn}} = 7$ on top of the "background spikes" with $\xi = 1$ (**Figure 3**, middle column). The probability of these events was set to $f_A(7) = 0.0125$, which resulted in a average pairwise correlation coefficient of $c = 0.01$ if the events of $z(t)$ were assigned homogeneously to a population of $N = 50$ spike trains. The value for $\nu = 500$ Hz was chosen to match the average carrier rate of the first example. Note that the additional events of amplitude $\xi = 7$ resulted in a slight increase of the population spike count from $\kappa_1[Z] = \kappa_1[R]h\mu_1 = 500{\cdot}0.005{\cdot}1 = 2.5$ in the first example to $\kappa_1[Z] \approx 500{\cdot}0.005{\cdot}1.0753 \approx 2.7$ in this example, and thereby also increased the firing rates of the $N = 50$ spike trains. The remaining parameters in this and all following examples were: bin width $h = 5$ ms, simulation time $T = 100$ s.

Compared to the size of the population ($N = 50$), the rate of the high-amplitude events ($\nu_7 = \kappa_1[I]{\cdot}f_A(7) \approx 6$ Hz) is relatively small. As a consequence, these are hardly visible in the raster displays and population spike counts (**Figure 3**, second and third panels from top in the middle column). In the distributions of the population spike count, they lead to a slight increase for the frequency of patterns of size $k \approx 10$, visible only on a logarithmic scale (**Figure 3**, middle column, green solid line in fourth panel from top).

In this data set, the stationary CuBIC rejects all hypotheses up to a value of $\xi = 6$. Hence, CuBIC performs optimally in this data set, as the resulting lower bound $\hat{\xi} = 7$ corresponds to the maximal order of correlation $\xi_{\mathrm{syn}} = 7$ in this data set.

To our big surprise and satisfaction, p-values for the adapted version of CuBIC were very similar to those of the stationary CuBIC (**Figure 3**, bottom panel in middle column, red and green bars respectively). In particular, the adapted CuBIC rejected the same hypotheses, and, as a consequence, also yielded the same lower bound $\hat{\xi} = 7$. Contrary to our assumption, the proposed adaptation thus did not compromise CuBIC's sensitivity in this stationary data set.

Combined effects

Finally, we generated a third data set that combined the properties of the first two examples (**Figure 3**, right column). The amplitude distribution was the same as in the example of pure correlation, i.e., with an additional entry at $\xi_{\mathrm{syn}} = 7$, while the carrier rate was the same cosine as in the data with pure non-stationarity (**Figure 3**, top panels of right and left columns, respectively).

For this data set, we expected the stationary CuBIC to overestimate the order of correlation, i.e., to yield $\hat{\xi} > \xi_{\mathrm{syn}} = 7$, as the considerable rate co-variation produces correlation among the X_i in addition to the events of high amplitude (Staude et al., 2008). To the contrary, however, p-values for the stationary CuBIC fell below the significance level only up to $\xi = 4$ (**Figure 3**, green arrowheads in the bottom panel of the left column), yielding $\hat{\xi} = 5$. Allowing for cosine-like non-stationarities in the null-hypotheses reduces the lower bound to $\hat{\xi} = 3$ (**Figure 3**, red arrowheads in the bottom panel of the left column). Thus, rather than a false positive inference of correlation, the additional non-stationarity resulted in a reduction of the inferred order of correlation as compared to the stationary scenario.

Different carrier family

To investigate the robustness of the proposed adaptation with respect to faulty choices for the carrier family, we also analyzed the data under the assumption that the carrier rate switches between a state of low rate, ν_{min}, and of high rate, ν_{max}. This can be described by a bimodal carrier distribution

$$f_R(\nu;\nu_{\mathrm{min}},\nu_{\mathrm{max}},\eta) = (1-\eta)\delta(\nu-\nu_{\mathrm{min}}) + \eta\delta(\nu-\nu_{\mathrm{max}}),$$

where $\delta(\nu)$ is the Dirac-delta function and $\eta \in [0,1]$ parametrizes the relative proportion of bins where $\nu(t)$ is in the low (high) rate state. Inspecting the time course of $Z(s)$, we chose the mass of the two peaks identical ($\eta = 0.5$), which leaves a two-parameter family for f_R (its cumulants are derived in Section "Cumulants of Carrier Distributions" in Appendix). For the data shown in **Figure 3**, allowing for such drastic rate dynamics hardly changed p-values as compared to the cosine carrier rate (compare red with blue bars in bottom panel of **Figure 3**). Most importantly, the bimodal adaptation resulted in the same lower bounds $\hat{\xi}$ as the cosine adaptation.

INTERNALLY GENERATED NON-STATIONARITIES

Gamma-distributed carrier rate

In the examples of the previous section, we inferred the type of non-stationarity from the experimental paradigm, i.e., from the properties of the stimulus. In many experimental situations, such

a priori assumptions on the rate variable cannot be justified, and changes in firing rates can have diverse statistical properties. General excitability of the tissue can change firing rates either in the form of slow drifts or abrupt transitions between so called up- and down-states. Furthermore, both local computations and additional cortical or sub-cortical inputs may change firing rates of the recorded population.

A further feature of the previous examples was that the carrier rate $\nu(t)$ changed rather slowly. As mentioned above, however, the temporal dynamics of the carrier rate does not influence the statistics required for CuBIC, i.e., the population spike count Z, as long as the distribution of the rate values per bin, f_R, is not altered. The carrier rate $\nu(t)$ of the second class of examples is a piecewise constant function that changes its value in subsequent bins, i.e.,

$$\nu(t) = \sum_{s=1}^{L} R_s \Pi_s(t),$$

where $L = T/h$ is the sample size (number of bins) $\Pi_s(t) = 1$ for $t \in [sh,(s+1)h)$ and $\Pi_s(t) = 0$ otherwise, and the R_s are the rate values drawn i.i.d. from the carrier distribution f_R (compare the model of rate-covariations in Staude et al., 2008 and the carrier rate in **Figure 2C**). Here, f_R is a gamma distribution, i.e., $f_R(r;k,\theta) = \frac{r^{k-1}}{\theta^k \Gamma(k)} e^{-r/\theta}$ where $\Gamma(k)$ is the gamma function. As in the previous section, the parameters of the carrier rate (k and θ) are such that distributing the events of the carrier process $z(t)$ uniformly among $N = 50$ spike trains leads to time varying firing rates $\lambda(t)$ with mean value 10 Hz, and we set its variance to 40 Hz^2. The maximal value of $\lambda(t)$ in the entire simulation was ≈79 Hz. The remaining parameters in the three examples shown in **Figure 4** are as in those of **Figure 3**. In data with pure rate non-stationarity (**Figure 4**, left column) the amplitude distribution has an isolated peak at $\xi = 1$. For pure correlation (**Figure 4**, middle column), the carrier rate is constant [$\nu(t) = 500$ Hz] and the amplitude distribution has an additional peak at $\xi_{syn} = 7$ that resulted in a pairwise correlation coefficient of $c = 0.01$ among the $N = 50$ spike trains. Finally, the combined data set has the same time-varying carrier rate as the purely non-stationary data, and the same amplitude distribution as the data with pure correlation (**Figure 4**, right column).

The gamma-distributed rate variable generates strongly fluctuating rate profiles, with peak values of the carrier rate above 3000 Hz. This leads to strong fluctuations in the population spike count even in the case of pure rate non-stationarity (**Figure 4**, third panel from top in left column), such that its distribution has a fairly elongated tail (**Figure 4**, fourth panel from top in left column).

Results of the stationary CuBIC

Despite quantitative differences, stationary CuBIC performs qualitatively similarly for gamma- and cosine-like carrier rate. For pure rate non-stationarity, it wrongly interprets the rate-induced correlations among the X_i as events of high amplitudes. The null-hypotheses are rejected up to $\xi = 3$, yielding a lower bound of $\hat{\xi} = 4$ (**Figure 4**, green arrowheads in the bottom panel of the left column). Similarly, adding gamma-like non-stationarities to a data set with correlation decreases the inferred lower bound, here from $\hat{\xi} = 7$ in the stationary data (**Figure 4**, green arrowheads

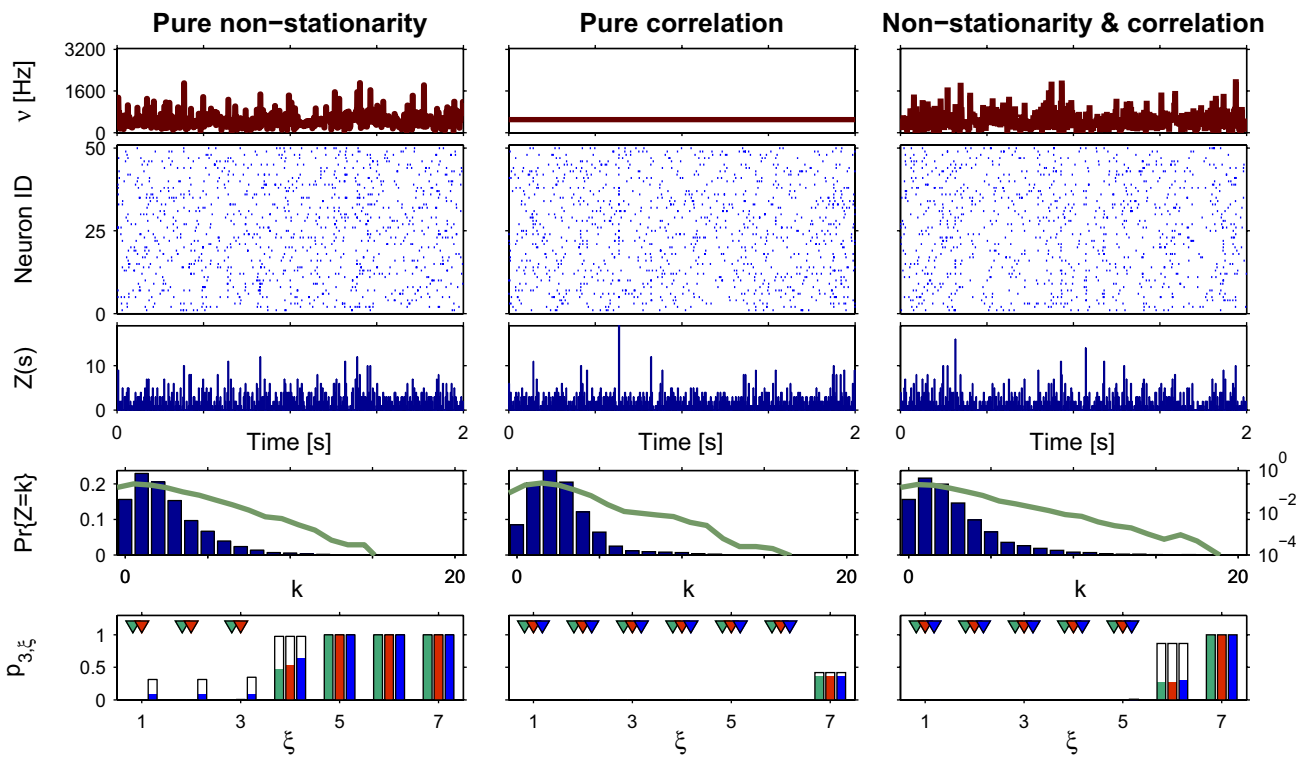


FIGURE 4 | Population statistics and CuBIC results for non-stationarities with gamma-distributed carrier rate. The figure has the same setup as **Figure 3**, only that the bottom panels show results for stationary CuBIC (green bars and arrowheads), allowed uniform carrier distribution (red bars and arrowheads) and gamma-distributed carrier rate (blue bars and arrowheads).

in the bottom panel of the middle column) to $\hat{\xi} = 6$ in the non-stationary data (**Figure 4**, green arrowheads in the bottom panel of the right column).

The adapted CuBIC

As opposed to Section "Stimulus-driven Non-stationarity" where properties of the carrier rate could be inferred from the stimulus, we here cannot make qualitative guesses about the type of non-stationarity. The fact that firing rates fluctuate on a bin-to-bin basis makes it very difficult to infer the type of non-stationarity from the raster plots, the population spike counts Z or its distribution f_Z. As a consequence, we can only make ad hoc assumptions on f_R. We consider two cases. First, we allow f_R to be a uniform distribution (**Figure 4**, red bars and arrowheads in bottom panels). As the uniform distribution is symmetric, we use Eq. 26 as the objective function with the additional constraint $\beta_2 \leq 1/3$ imposed by the non-negativity of the carrier rate. The cumulants of the uniform distribution (see Cumulants of carrier distributions in Appendix) are then used for the computation of $\mathrm{Var}[k_3]$. Second, we allow for gamma-distributed non-stationarities, where we use Eq. 27 as the objective function and the cumulants of the gamma-distribution for the computation of $\mathrm{Var}[k_3]$ (**Figure 4**, blue bars and arrowheads in bottom panels).

For the data with pure rate non-stationarity (**Figure 4**, left column), allowing a uniform rate variable rejects hypotheses up to $\xi = 3$ and thus yields a lower bound of $\hat{\xi} = 4$. Allowing f_R to be gamma-distributed, on the other hand, produces p-values above 0.05 for all $\xi = 1,\ldots,7$, thereby retaining all hypotheses. Thus, while the uniform null-hypotheses fail to reduce the lower bound as compared to the stationary version, allowing for the true, gamma-distributed non-stationarities corrects for false positive inference of correlation.

For data with correlation, the choice of the carrier distribution has only little influence on the resulting p-values (**Figure 4**, bottom panels in middle and left column). For pure correlation, the lower bounds are identical for all three rate models ($\hat{\xi} = 7 = \xi_{\mathrm{syn}}$). In the combined data set (**Figure 4**, right column), the additional non-stationarity reduces the lower bound as compared to the stationary data with correlation to $\hat{\xi} = 6$, irrespective of the non-stationarity allowed in the null-hypotheses.

REFRACTORY PERIODS

Besides the stationarity, CuBIC's second central assumption is that spike trains of single neurons can be described as Poisson processes, i.e., have exponential interspike interval (ISI) distributions. While tails of ISI distributions can often be relatively well described by exponentials, the high probability for short intervals of the Poisson process contradicts the absolute refractory period of a few milliseconds found in most nerve cells. We investigated the extent to which refractoriness influences test results of CuBIC by re-analyzing the data of the previous sections after short ISIs were removed. Specifically, for each data set we assigned the events of the carrier process randomly to the $N = 50$ spike trains, removed spikes of all spike trains that occurred with a temporal difference of $\tau \leq 2$ ms, and constructed the population spike counts of these thinned spike trains. The analysis of the refractory data showed that introducing refractoriness has a very limited effect on test results (outlined bars and arrowheads in lower panels of **Figures 3 and 4**). If p-values changed, they increased, which makes CuBIC more conservative but does not generate false positives. From all simulation, the increased p-value reduced the lower bound $\hat{\xi}$ only in a single instance (cosine rate modulation with correlation, analyzed with stationary CuBIC at $\xi = 4$; green arrowhead in bottom right panel of **Figure 3**); in all remaining cases the refractory period changed p-values where they were above the significance level of $\alpha = 0.05$. We thus conclude that CuBIC is reasonably robust if spike deviate from Poisson processes in terms of short refractory periods (here 2 ms).

DISCUSSION

The analysis of massively parallel spike trains for higher-order correlations is a prerequisite for understanding the mechanisms of cooperative neuronal computation in the brain. However, analysis techniques to estimate higher-order correlations typically require enormous sample sizes, rendering the analysis of large ($N > 10$) populations for higher-order effects extremely difficult. In Staude et al. (2009), we have presented a novel method (CuBIC) that avoids the need for extensive sample sizes. Numerical simulations illustrated that CuBIC reliably infers correlations of very high order (>10) from recordings of large ($N \sim 100$), even weakly correlated neuronal populations (average pairwise correlation coefficient $c < 0.01$) with sample sizes corresponding to 10–100 s recording time.

As a statistical model, CuBIC assumes the CPP, where correlation among the spike trains is modeled by the insertion of additional coincident events (Ehm et al., 2007; Johnson and Goodman, 2007; Brette, 2009; Staude et al., 2010). The linear relation between the model parameters, i.e., the orders of correlation present in the data, and the cumulants $\kappa_m[Z]$ of the population spike count Z allows the computation of the maximal value of $\kappa_m[Z]$ under the assumption that the orders of correlation in the data do not exceed a predefined value ξ. Comparing the estimated cumulants to these upper bounds then yields a collection of statistical hypothesis tests $H_0^{m,\xi}$, labeled by m, the order of the *estimated cumulant*, and ξ, the *maximal order of correlation* allowed in the null-hypothesis. In this paper, we chose a fixed value of $m = 3$, and for given ξ, the rejection of $H_0^{3,\xi}$ implies that the data must have correlations of order at least $\xi + 1$ (for a discussion on the choice of m see below). A combination of tests with different values for ξ finally yields a lower bound for the maximal order of correlation, denoted by $\hat{\xi}$. For a discussion of the properties and limitations of the CPP (e.g., positivity of correlations), general issues concerning CuBIC, and the relationship between cumulant-correlations and the higher-order parameters of the log-linear model used by, e.g., Martignon et al. (1995), Schneidman et al. (2006), Shlens et al. (2006), we refer to the extensive discussion of Staude et al. (2009). The latter point is discussed in more detail also in Staude et al. (2010). In this section, we focus on issues that relate directly to the novelties of the present study.

Before going into detail, we need to make a general remark. CuBIC is a parametric procedure, and assumes that the data, i.e., the population spike count, can be described sufficiently well by a discretized, potentially non-stationary CPP. If this model does not fit, results of CuBIC may be unreliable. The extent to which results are reliable then depends on the mismatch between the CPP and the data. In practice, where single spike trains deviate from Poisson

processes, for example due to refractory properties, this mismatch may be evaluated by analyzing surrogate data (e.g., Grün, 2009). If, for instance, CuBIC returns $\hat{\xi} = 10$ in a data set of non-Poissonian spike trains, and the analysis of surrogate data with the same single process properties but without correlation yields $\hat{\xi} = 1$, we can conclude that the value of $\hat{\xi} = 10$ is really due to existing correlations. This kind of analysis, however, has to be performed specifically for a given data set. As a first step, we have here investigated the effect of a spike train's most obvious deviation from the Poisson process: absolute refractory periods (here 2 ms, **Figures 3 and 4**). Its relatively small effect on *p*-values makes us confident that CuBIC is a promising analysis method even if single spike trains deviate from the Poisson assumption. The detailed analysis of CuBIC's robustness with respect to these violation is a central aspect of our current research (e.g., Staude et al., 2007; Reimer et al., 2009). We wish to stress that in the case of small bin sizes and hard refractory periods, assuming single processes to have Poisson statistics is essentially identical to the popular assumption of independence of subsequent bins in Bernoulli models (as in e.g., Martignon et al., 1995; Shlens et al., 2006). A more detailed discussion of this issue, together with an analysis of the parameter dependence of CuBIC can be found in Staude et al. (2009).

A central assumption in the original presentation of CuBIC (Staude et al., 2009) is that the statistics of the population does not change over time (stationarity). In the present study, we have adapted CuBIC to be able to analyze also non-stationary data. The basis of this adaptation is a non-stationary version of the CPP, where the intensity of the population, parametrized by the ν, is decoupled from the correlation structure, parametrized by the amplitude distribution f_A. Describing the population spike count as a doubly stochastic CPP, potential non-stationarities in the data can by integrated into the quantification of the null-hypotheses of CuBIC. We wish to stress once again, however, that non-stationarities in single neurons do not necessarily imply time varying carrier rates (see, e.g., **Figure 2A**), such that not every non-stationary data set requires the application of the adapted CuBIC.

In this study, we presented the adaptation only for the third cumulant, i.e., $m = 3$. Although the derivation of the mathematical equations is straightforward also for higher values of m, the resulting expressions become increasingly complicated. This may result in particular in strong non-linearities in the maximization problem, such that additional care is necessary to ensure convergence of numerical solvers at the global maximum. Furthermore, the estimation of cumulants of order >3 becomes unreliable and their estimators have large variance, which may compromise test power. As CuBIC proved to be very sensitive even for $m = 3$ (Staude et al., 2009), we currently have not developed the theory for higher m. Nevertheless, this might be a fruitful direction of future research.

The main difference between the stationary CuBIC and its non-stationary adaptation lies in the maximization of the third cumulant of Z. Here, the adapted version requires that the third standardized cumulant $\beta_3 = \kappa_3[R]/\kappa_1[R]^3$ of the binned carrier rate R does not assume arbitrary large values. In this study, this is achieved by prescribing a two-parameter family for the carrier distribution f_R (the "carrier family"). The remainder of this section is therefore primarily concerned with elucidating the role of this particular choice.

CORRECTING FOR RATE EFFECTS

Classically, correlations are measured on a small time scale, and subsequently corrected for effects from the firing rates. The estimation of the firing rate, in turn, proceeds along one of the following two lines. Either, rate variations are identified with artifacts that are locked to the stimulus or some other external cue. Then, firing rates are estimated by averaging over repeated presentation of the stimulus, or trials (e.g., Perkel et al., 1967a,b; Aertsen et al., 1989). Alternatively, changes in firing rates are assumed to fluctuate only on slow time scales; they are then estimated by averaging over time. Although combinations, refinements and optimizations of both strategies have been developed (e.g., Grün et al., 2002b; Ventura et al., 2005; Shimazaki et al., 2009), we wish to stress that all such corrections make strong a priori assumptions on the differences between "genuine" correlations and "artifacts" induced by non-stationary firing rates (see also Staude et al., 2008).

The correction of CuBIC, i.e. the choice of a carrier family, follows a fundamentally different strategy. Rather than assuming that firing rate profiles are identical over trials (first strategy) or that rate-fluctuations are band-limited (second strategy), the carrier family limits the extent to which correlations among the counting variables X_i are assigned to rate-variations. As discussed above, this choice ensures boundedness of the third standardized cumulant $\beta_3 = \kappa_3[R]/\kappa_1[R]^3$ of the rate variable R. As $\kappa_3[R]$ measures the skewness of the carrier distribution, large values of β_3 imply a long right tail. As a consequence, a population with large β_3 has bins with very large values of the carrier rate R. If binned firing rates can assume arbitrary high values, however, the difference between global rate variations and precise spike coordination vanishes. Thus, the choice of the carrier family determines the extent to which one assigns patterns with high spike counts to (co-)variations of firing rates. In other words: the carrier family sets CuBICs demarcation line between rate co-variation and genuine correlation.

In Section "Case Studies", we illustrated how CuBIC operates in two alternative scenarios. In Section "Stimulus-driven Non-stationarity", we assumed that properties of the firing rates can be inferred from the stimulus. Here, the stimulus was a cosine, but the reasoning there can be generalized to a broad class of stimuli. The effect of oriented bars, for instance, might be modeled by a bimodal carrier distribution as in Section "Different Carrier Family",

$$f_R(\nu;\nu_{\min},\nu_{\max},\eta) = (1-\eta)\delta(\nu-\nu_{\min}) + \eta\delta(\nu-\nu_{\max}), \tag{29}$$

where the mixing parameter η determines the relative duration of the respective stimulus phases (light/dark). The carrier family of Eq. 29 also describes experiments with a well-defined stimulus onset, as e.g., odor presentation or whisker stimulation. In such a scenario, η has to be chosen as the relative duration of the stimulus-off/stimulus-on epoch. Furthermore, properties of the carrier family in "free viewing" conditions might be estimated from the statistical properties of the visual scene.

In the second scenario (see Internally Generated Non-stationarities), the experimental paradigm did not provide information about the carrier family. Here, we argued that ad hoc assumptions are the only option. However, if the activity of individual neurons is available, their statistics can provide additional information that can be exploited. If, for instance, individual spike

trains do not show evidence of time varying firing rates, or provide upper bounds for the cumulants of their firing rates, this information may be extrapolated to the level of the carrier rate. Together with general moment inequalities (e.g., Kumar, 2002), such information may help to dispose of the explicit choice of the carrier family by providing an explicit upper bound for β_3. Although the absence of a precise parametric model for the carrier distribution impedes the faithful computation of $\mathrm{Var}[k_3]$ under $H_0^{3,\xi}$, the close similarity of the error-bars in **Figure 5A** for the different methods makes us confident that $\mathrm{Var}[k_3]$ can be reasonably approximated by any carrier distribution as long as the upper bound $\kappa^{*,ns}_{3,\xi}$ is faithful (see also Section 'What is the "true" Carrier Family?' for conceptual issues concerning the choice for f_R).

DISCUSSION OF CASE STUDIES

The analysis of artificial data (**Figures 3 and 4**) can be summarized by four major points:

1. For data with pure firing rate non-stationarity but no higher-order correlations, the stationary CuBIC misinterprets the common rate variations as events of amplitude ≥2, i.e., as correlations. The order of the inferred correlation depends on the kind of non-stationarity in the data ($\hat{\xi}=2$ for cosine carrier rate, $\hat{\xi}=4$ for gamma-distributed rate variable).
2. Allowing for potential non-stationarities in the null-hypotheses can reduce this lower bound. Using the correct family for the carrier distribution, i.e., a cosine-like model for the data in **Figure 3** and a gamma-distribution for the data in **Figure 4**, corrects entirely for the false positive inference of the stationary CuBIC, such that the lower bound becomes $\hat{\xi}=1$. Choosing the wrong family, however, may not account properly for rate variations, especially if the family assigns smaller values of β_3 (e.g., allowing a uniform carrier distribution if the data had a gamma-distributed carrier rate as in **Figure 4**).
3. For stationary data with correlation, allowing for non-stationarities in the null-hypotheses has no effect on the inferred lower bound. This result holds irrespective of the type of non-stationarity allowed in the null-hypotheses.
4. Non-stationarities in data with correlation reduce the inferred lower bound as compared to data with the same correlation structure but constant firing rates. The degree of reduction depends mostly on the kind of non-stationarity in the data. The family allowed for the carrier distribution did not affect the lower bound.

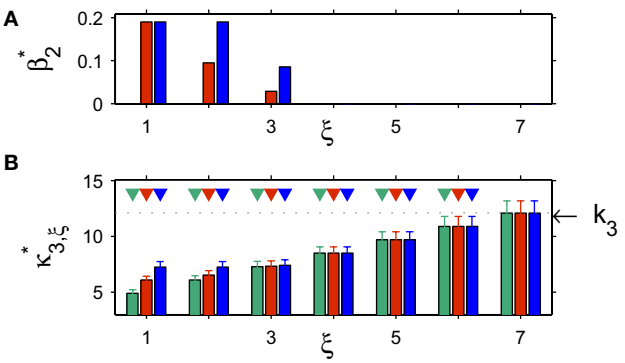


FIGURE 5 | Results of the maximization (Eq. 25 with appropriate objective functions) for data with pure correlation (see legend of Figure 3 for parameters). (A) Normalized variance of the rate variable β_2^* that solves Eq. 25 for $\xi = 1,\ldots,7$ for different types of allowed non-stationarities (green: stationary CuBIC; red: non-stationary CuBIC allowing uniform carrier distribution; blue: non-stationary CuBIC allowing gamma carrier distribution). **(B)** The maximal third cumulant $\kappa^*_{3,\xi}$ as a function of the test parameter ξ for different types of allowed non-stationarities (same color code as in **A**). Error bars denote $2\sqrt{\mathrm{Var}[k_3]}$, corresponding to a significance level of $\alpha \approx 0.05$, the dashed line is the value of the test statistic k_3 and arrowheads denote rejected null-hypothesis where $\kappa^*_{3,\xi} + 2\sqrt{\mathrm{Var}[k_3]} < k_3$ (compare bottom panel of middle column in **Figure 4**).

It is well-known that co-variations in firing rates induce correlations between spike counts. Thus, it comes as no surprise that analyzing non-stationary data with the stationary CuBIC generates false positives (point 1). In the adaptation, parts of the correlation can be assigned to (co)-variations in firing rates. Allowing for non-stationarities therefore corrects for this faulty inference (point 2).

To understand why the adaptation did not generally reduce the lower bound (point 3), we investigated the interplay between the constraints and the objective function in the maximization of the third cumulant in further detail. Due to the increased number of degrees of freedom, our intuition was that allowing for non-stationarities additional to the correlations of order ξ would simply increase the maximal value of the third cumulant $\kappa^*_{3,\xi}$ as compared to the stationary version. This, however, seems not to be the case. In the data sets with pure correlation (**Figures 3 and 4**, middle panels), for example, the values of $\kappa^*_{3,\xi}$ and $\kappa^{*,ns}_{3,\xi}$ are identical for $\xi \geq 4$ (**Figure 5B** shows results for stationary, uniform and gamma-distributed rate variables, results of combined data sets are very similar, data not shown). The reason is that for a given value of the first estimated cumulant k_1, the constraint $k_2 = \vec{\xi}_2 \cdot \vec{\nu}_\xi h + k_1^2 \beta_2$ penalizes stronger non-stationarity, i.e., larger values of the standardized variance β_2, with a reduction of the (expected) component rates $\nu_1,\ldots,\nu_\xi$, and vice versa. In the objective function (Eq. 25)

$$F\left(\nu_1,\ldots,\nu_\xi,\beta_2\right) = \vec{\xi}_3 \cdot \vec{\nu}_\xi h + k_1^3\beta_3 - 3k_1^3\beta_2^2 + 3k_1k_2\beta_2, \tag{30}$$

the component rates enter via $\vec{\xi}_3 \cdot \vec{\nu}_\xi$, while the β_2-dependence is a parabola with negative curvature. As

$$\vec{\xi}_3 \cdot \vec{\nu}_\xi = \sum_{k=1}^{\xi} k^3 \nu_k \gg \sum_{k=1}^{\xi} k^2 \nu_k = \vec{\xi}_2 \cdot \vec{\nu}_\xi, \tag{31}$$

especially for large values of ξ, the objective function profits more from high component rates than the constraint penalizes these. As a consequence, the maximization favors high component rates over strong rate fluctuations, especially for large ξ. The results of the maximization procedure for the data with pure correlation supports this interpretation, as the standardized variance of the model that maximizes $\kappa_3[Z]$, β_2^*, decreases with ξ (**Figure 5A**). For $\xi \geq 4$, we have $\beta_2^* = 0$, and, as a consequence, the maximizing model of the adapted maximization problem is the same as that of the stationary maximization. Consequently, also the solutions $\kappa^*_{3,\xi}$ and the *p*-values

are identical. Contrary to our initial intuition, the maximization thus does not generally favor strong rate variations. Evidently, however, the extent to which the inclusion of non-stationarities in the null-hypothesis alter test results depends crucially on the parameters of the data.

PARAMETER-DEPENDENCE OF TEST RESULTS

The results of the case studies summarized above suggest that including potential non-stationarities in the null-hypothesis is always a safe bet: it corrects for false positive inference if correlation originates from rate effects, but does not alter *p*-values if the stationary CuBIC did not overestimate the order of correlation. To sketch the parameter range where including non-stationarities reduces *p*-values only if necessary, recall that we have identified the reason for the unchanged maximal third cumulant in the interplay between the constraint and the objective function. Considering the general objective function (Eq. 25) however, we find that the influence of the non-stationarity (via β_2 and β_3) depends on the value of the first sample cumulant k_1. For $k_1 < 1$, non-stationarities (positive β_2 and β_3) hardly influence the objective function, hence the maximization can be assumed to favor high component rates over non-stationarities, yielding identical test results for the stationary and the adapted CuBIC. For $k_1 \gg 1$ a small increase in β_3 has a strong effect on the objective function, which may in turn favor strong non-stationarities, thereby producing different test results for the stationary and the adapted method. Thus, a crucial parameter for the performance of the adaptation is the first sample cumulant k_1. Now k_1 is the estimator of $\kappa_1[Z] = \kappa_1\left[\Sigma_{i=1}^N X_i\right] = h\Sigma_{i=1}^N \lambda_i$, where λ_i is the (average) firing rate of the *i*th neuron. Thus, given the summed firing rate $\Sigma_{i=1}^N \lambda_i$ of the recorded population, we may chose the bin size h in order to keep k_1 in a range where the adaptation can be expected to have reasonable test power. As all simulations of Section "Case Studies" had $k_1 \sim 2.5$, achieving a value of $k_1 < 1$ is not always necessary.

WHAT IS THE "TRUE" CARRIER FAMILY?

In Section "Correcting for Rate Effects" we have presented a few guidelines for the choice of the family of carrier rate distributions. There are cases, however, where this choice is not easily justifiable by resorting to observable quantities. To discuss the status of this problem in more depth, we wish to stress that measures of correlation and their various corrections are purely statistical in nature. To be of scientific value, statistical results have to be put in context and must be interpreted, e.g., in terms of biophysical mechanisms. The "firing rate" of a neuron, for example, is not a biophysical entity as such. The term arises only if one describes the variable behavior of a neuron using point processes. As a consequence, whether or not the choice of a chosen carrier family f_R is "valid" depends entirely on the intended biological interpretation. If, for example, the choice of f_R is guided by the properties of the stimulus, the natural interpretation of rejected null-hypothesis is that the dynamic properties of the neuronal network under investigation generates correlations beyond direct stimulus effects. In an alternative situation, f_R may be chosen to reflect the slow ongoing dynamics observed in a simultaneously recorded mass signal (as e.g., described in Tsodyks et al., 1999). Significant higher-order correlations are then interpreted as coordinated activity that is not covered by such large-scale phenomena. In either case, the term "correlations beyond the rate" is biologically meaningful only after an explicit interpretation of the term "rate" by the experimenter. Finally, we stress that the different choices for the carrier family affected the test results only weakly, especially if the data had genuine spike correlations (**Figures 3 and 4**, left and right columns). Thus, a perfect match between the carrier distribution underlying a Monte Carlo simulation and the family assigned in CuBIC does not seem to be of great importance for a reliable interpretation of test results.

APPENDIX

LIST OF SYMBOLS

Symbol	Meaning
$x_i(t)$	*i*th spike train in continuous time
X_i	Counting variable of *i*th spike train
$z(t)$	Carrier process, i.e., summed activity in continuous time
Z	population spike count
$\kappa_i[Z]$	*i*th cumulant of Z
k_i	*i*th sample cumulant of Z, i.e., *k*-statistic
h	Bin width
λ	Average firing rate of individual spike trains
f_A	Amplitude distribution, i.e., population-average correlation structure
μ_i	*i*th raw moment of amplitude distribution
ν	Carrier rate
ν_k	(Expected) compound rate of all events with amplitude *k*
$\vec{\nu}_\xi$	Vector of (expected) compound rates $\nu_1,\ldots,\nu_\xi$
$\vec{\xi}_i$	Vector of *i*th powers of $1,\ldots,\xi$
$H_0^{3,\xi}$	Null-hypothesis stating that the data has correlation of maximal order ξ
$\kappa^*_{3,\xi}$	Maximal value of $\kappa_3[Z]$ given correlations of orders $\leq\xi$
R_s	Carrier variable, mean value of the carrier rate in the sth bin
f_R	Carrier distribution: distribution of the $\{R_s\}_s$
β_k	*k*th standardized cumulant of the carrier distribution
$\kappa^{*,ns}_{3,\xi}$	Maximal value of $\kappa_3[Z]$ given correlations of orders $\leq\xi$ and non-stationarity

THE SOLUTION OF THE MAXIMIZATION

This Appendix shows that the solution $(\nu_1^*,\ldots,\nu_\xi^*)$ of the stationary (Eq. 6) and the non-stationary maximization problem (Eq. 25) fulfills $\nu_k^* = 0$ for $k = 2,\ldots,\xi - 1$.

The non-stationary problem (Eq. 25) has the objective function

$$F = \sum_{k=1}^{\xi} k^3\nu_k + k_1^3\beta_3 - 3k_1^3\beta_2^2 + 3k_1k_2\beta_2 \tag{32}$$

with constraints

$$k_1 = \sum_{k=1}^{\xi} k\nu_k \tag{33}$$

$$k_2 = \sum_{k=1}^{\xi} k^2\nu_k + k_1^2\beta_2 \tag{34}$$

Simple computations starting from Eqs. 33 and 34 yield

$$\nu_1 = \frac{1}{\xi - 1}\left(\xi k_1 - k_2 - k_1^2\beta_2 - \sum_{k=2}^{\xi-1}\left(\xi k - k^2\right)\nu_k\right)$$

$$\nu_\xi = \frac{1}{\xi(\xi-1)}\left(k_2 - k_1 + k_1^2\beta_2 - \sum_{k=2}^{\xi-1}\left(k^2 - k\right)\nu_k\right),$$

which, after insertion into Eq. 32 yield

$$F = \frac{1}{\xi - 1}\sum_{k=2}^{\xi-1}\left(k^3 + k^2\left(1-\xi^2\right) + k\left(\xi^2 - \xi\right)\right)\nu_k + H, \tag{35}$$

with

$$H = k_1\left(k_1\beta_2(\xi+1) - \xi\right) + k_2(\xi+1) + k_1^3\beta_3 - 3k_1^3\beta_2^2 + 3k_1k_2\beta_2.$$

Now for $k = 2,\ldots,\xi - 1$ we have

$$\frac{\partial F}{\partial \nu_k} = \frac{k}{\xi - 1}\left(k^2 + k\left(1-\xi^2\right) + \xi^2 - \xi\right) \tag{36}$$

$$k \le \xi - 1\; k^2 + \xi^2 - \xi^2 k + k - \xi \tag{37}$$

$$= k^2 + \xi^2 - \xi^2 k + k - \xi \tag{38}$$

$$k < \xi\; \xi^2 + \xi^2 - \xi^2 k + \xi - \xi \tag{39}$$

$$= 2\xi^2 - k\xi^2 \tag{40}$$

$$2 \le k\, 0 \tag{41}$$

The gradient of the objective function therefore points to negative values of ν_k for $k = 2,\ldots,\xi - 1$. Because $\nu_k \ge 0$, the maximum is achieved for $\nu_k = 0$. In the stationary maximization problem, insertion of the constraints into the objective function yields the same F as Eq. 35, only with $H = k_2(\xi + 1)$. Consequently, Eqs. 36–41 hold also in the stationary scenario.

CUMULANTS OF CARRIER DISTRIBUTIONS

For the computation of $\mathrm{Var}[k_3]$ (Eq. 7) from the solutions of the maximization procedure, i.e., the parameters $\vec{\nu}_\xi^*$ and β_2^*, we require explicit expressions for the cumulants of the carrier distribution f_R up to order $m = 6$. Recall that we here considered only two-dimensional parametric families for f_R, which implies that these parameters can be computed from the known mean value $\mathrm{E}[R] = \Sigma_{k=1}^{\xi}\nu_k^*$ and the normalized variance β_2^*. In the following, we provide explicit expressions that relate the raw moments to these parameters. Expressions for the cumulants are then computed by applying the conversion map G constructed in Section "Cumulants of the Non-stationary CPP".

Cosine carrier rates

For $\nu(t) = B + C\cos(2\pi\omega t + d)$, we derive the distribution of the rate values $R_s = 1/h\int_{sh}^{(s+1)h}\nu(t)dt$ under the assumption that subsequent values of R_s sample the cosine faithfully. In this case, we can express R as a function of the $[0,1]$-uniform variable T as $R = g(T)$, with $g(t) = B + C\cos(T)$. Denoting the uniform distribution by f_T, the distribution function of R is thus given by

$$f_R(r) = f_T\left(g^{-1}(r)\right)\left|\left(\frac{dg^{-1}(r)}{dr}\right)\right| = \frac{1}{C\sqrt{1-\left(\frac{R-B}{C}\right)^2}}$$

The first six moments can be computed by solving the integrals

$$\mathrm{E}\left[R^m\right] = \int_{B-C}^{B+C} r^m \frac{1}{C\sqrt{1-\left(\frac{r-B}{C}\right)^2}}dr,$$

which yields

$$\mathrm{E}[R] = B$$

$$\mathrm{E}\left[R^2\right] = B^2 + \frac{C^2}{2}$$

$$\mathrm{E}\left[R^3\right] = B^3 + \frac{3BC^2}{2}$$

$$\mathrm{E}\left[R^4\right] = B^4 + 3B^2C^2 + \frac{3C^4}{8}$$

$$\mathrm{E}\left[R^5\right] = B^5 + 5B^3C^2 + \frac{15BC^4}{8}$$

$$\mathrm{E}\left[R^6\right] = B^6 + \frac{15B^4C^2}{2} + \frac{45B^2C^4}{8} + \frac{5C^6}{16}$$

Bimodal carrier rates

Let $f_R(\nu;\nu_{\min},\nu_{\max},\eta) = (1-\eta)\delta(\nu - \nu_{\min}) + \eta\delta(\nu - \nu_{\max})$ be a bimodal rate distribution, where $\eta \in [0,1]$ and $\nu_{\min},\nu_{\max} \in [0,\infty]$. The raw moments of f_R are

$$\mathrm{E}\left[R^m\right] = (1-\eta)\nu_{\min}^m + \eta\nu_{\max}^m.$$

Uniform carrier rates

If R is uniformly distributed between a and b, the raw moments of R are given as

$$E\left[R^m\right] = \frac{b^{m+1} - a^{m+1}}{(m+1)(b-a)}.$$

Gamma carrier rates

The moments of the gamma distribution $f(R;k,\theta) = R^{k-1}\frac{e^{-R/\theta}}{\theta^k\Gamma(k)}$ with parameters k and θ are given as

$$E\left[R^m\right] = \frac{\theta^m\Gamma(k+m)}{\Gamma(k)},$$

where Γ denotes the gamma-function.

ACKNOWLEDGMENTS

We are grateful to Stuart Baker for a fruitful discussion, Leona Schild for help in the initial phase of this project, and Imke Reimer for comments on an earlier version of the manuscript. Supported by the German Federal Ministry of Education and Research (BMBF grants 01GQ0420 and 01GQ01413 to the BCCN Freiburg and Berlin), the Helmholz Alliance on Systems Biology (Germany), and SFB 780.

REFERENCES

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

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*, 1st Edn. Cambridge: Cambridge University Press.

Aertsen, A., Gerstein, G., Habib, M., and Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of "effective connectivity". *J. Neurophysiol.* 61, 900–917.

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366.

Averbeck, B. B., and Lee, D. (2006). Effects of noise correlations on information encoding and decoding. *J. Neurophysiol.* 95, 3633–3644.

Bair, W., Zohary, E., and Newsome, W. (2001). Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J. Neurosci.* 21, 1676–1697.

Bohte, S. M., Spekreijse, H., and Roelfsema, P. R. (2000). The effects of pair-wise and higher-order correlations on the firing rate of a postsynaptic neuron. *Neural Comput.* 12, 153–179.

Brette, R. (2009). Generation of correlated spike trains. *Neural Comput.* 21, 188–215.

Brown, E. N., Kaas, R. E., and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 7, 456–461.

Darroch, J., and Speed, T. (1983). Additive and multiplicative models and interactions. *Ann. Stat.* 11, 724–738.

Del Prete, V., Martignon, L., and Villa, A. E. P. (2004). Detection of syntonies between multiple spike trains using a coarse-grain binarization of spike count distributions. *Network* 15, 13–28.

Diesmann, M., Gewaltig, M.-O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402, 529–533.

Di Nardo, E., Guarino, G., and Senato, D. (2008). A unifying framework for k-statistics, polykays and their multivariate generalizations. *Bernoulli* 14, 440–468.

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science* 327, 584–587.

Eggermont, J. J. (1990). *The Correlative Brain*, Vol. 16, *Studies of Brain Function*. Berlin: Springer-Verlag.

Ehm, W., Staude, B., and Rotter, S. (2007). Decomposition of neuronal assembly activity via empirical de-poissonization. *Electron. J. Stat.* 1, 473–495.

Fujisawa, S., Amarasingham, A., Harrison, M., and Buzsaki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* 11, 823–833.

Gardiner, C. W. (2003). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, 3 Edn. *Springer Series in Synergetics*, Vol. 13. Berlin: Springer.

Gerstein, G. L., Bedenbaugh, P., and Aertsen, A. (1989). Neuronal assemblies. *IEEE Trans. Biomed. Eng.* 36, 4–14.

Gray, C. M., and Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 86, 1698–1702.

Grün, S. (2009). Data-driven significance estimation of precise spike correlation. *J. Neurophysiol.* 101, 1126–1140 (invited review).

Grün, S., Diesmann, M., and Aertsen, A. (2002a). 'Unitary Events' in multiple single-neuron spiking activity. I. Detection and significance. *Neural Comput.* 14, 43–80.

Grün, S., Diesmann, M., and Aertsen, A. (2002b). 'Unitary Events' in multiple single-neuron spiking activity. II. Non-Stationary data. *Neural Comput.* 14, 81–119.

Hebb, D. O. (1949). The organization of behavior: a neuropsychological theory. New York: John Wiley & Sons.

Johnson, D., and Goodman, I. (2007). Jointly Poisson processes. arXiv:0911.2524.

Josić, K., Shea-Brown, E., Doiron, B., and de la Rocha, J. (2009). Stimulus-dependent correlations and population codes. *Neural Comput.* 21, 2774–2804.

Kohn, A., and Smith, M. A. (2005). Stimulus dependence of neuronal correlations in primary visual cortex of the Macaque. *J. Neurosci.* 25, 3661–3673.

König, P., Engel, A. K., and Singer, W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci.* 19, 130–137.

Kreiter, A. K., and Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of awake macaque monkey. *J. Neurosci.* 16, 2381–2396.

Kuhn, A., Aertsen, A., and Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Comput.* 1, 67–101.

Kumar, P. (2002). Moments inequalities of a random variable defined over a finite interval. *J. Ineq. Pure Appl. Math.* 3, 41.

Martignon, L., Deco, G., Laskey, K., Diamond, M., Freiwald, W., and Vaadia, E. (2000). Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies. *Neural Comput.* 12, 2621–2653.

Martignon, L., von Hasseln, H., Grün, S., Aertsen, A., and Palm, G. (1995). Detecting higher-order interactions among the spiking events in a group of neurons. *Biol. Cybern.* 73, 69–81.

Montani, F., Ince, R. A. A., Senatore, R., Arabzadeh, E., Diamond, M. E., and Panzeri, S. (2009). The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. *Philos. Trans. R. Soc. Lond. A* 367, 3297–3310.

Nakahara, H., and Amari, S. (2002). Information-geometric measure for neural spikes. *Neural Comput.* 14, 2269–2316.

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967a). Neuronal spike trains and stochastic point processes. I. The single spike train. *Biophys. J.* 7, 391–418.

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967b). Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J.* 7, 419–440.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*, 2 Edn. New York, NY: Cambridge University Press.

Reimer, I. C. G., Staude, B., and Rotter, S. (2009). "Detecting assembly activity in massively parallel spike trains," in *Proceedings of the 8th Meeting of the German Neuroscience Society/30th Göttingen Neurobiology Conference*, Vol. 1, *Neuroforum, Supplement*, eds H. Bähr and I. Zerr. Heidelberg: Spektrum Akademischer Verlag.

Riehle, A., Grün, S., Diesmann, M., and Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278, 1950–1953.

Roudi, Y., Nirenberg, S., and Latham, P. E. (2009). Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS Comput. Biol.* 5, e1000380. doi: 10.1371/journal.pcbi.1000380.

Sakurai, Y., and Takahashi, S. (2006). Dynamic synchrony of firing in the monkey prefrontal cortex during working-memory tasks. *J. Neurosci.* 6, 10141–10153.

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012.

Shadlen, M. N., and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* 18, 3870–3896.

Shimazaki, H., Amari, S., Brown, E. N., and Grün, S. (2009). "State-space analysis on time-varying correlations in parallel spike sequences," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Washington, DC: IEEE Computer Society), 3501–3504.

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M., and Chichilnisky, E. J. (2006). The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* 26, 8254–8266.

Snyder, D. L., and Miller, M. I. (1991). Random Point Processes in Time and Space. New York: Springer.

Softky, W. R. (1995). Simple codes versus efficient codes. *Curr. Opin. Neurobiol.* 5, 239–247 (commentary).

Staude, B., Grün, S., and Rotter, S. (2010). "Higher order correlations and cumulants," in *Analysis of Parallel Spike Trains*, eds S. Grün and S. Rotter. Springer Series in Computational

Neuroscience. Berlin: Springer-Verlag.
Staude, B., Rotter, S., and Grün, S. (2007). "Detecting the existence of higher-order correlations in multiple single-unit spike trains," in *Abstract Viewer/Itinerary Planner*, Vol. 103.9/AAA18 (Washington, DC: Society for Neuroscience).
Staude, B., Rotter, S., and Grün, S. (2008). Can spike coordination be differentiated from rate covariation? *Neural Comput.* 20, 1973–1999.
Staude, B., Rotter, S., and Grün, S. (2009). CuBIC: cumulant based inference of higher-order correlations. *J. Comput. Neurosci.* doi: 10.1007/s10827-009-0195-x.
Stratonovich, R. L. (1967). *Topics in the Theory of Random Noise*. New York, Gordon & Breach Science.
Streitberg, B. (1990). Lancaster interactions revisited. *Ann. Stat.* 18, 1878–1885.
Stuart, A., and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*, 5 Edn. London: Griffin and Co.
Tsodyks, M., Kenet, T., Grinvald, A., and Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* 286, 1943–1946.
Vaadia, E., Aertsen, A., and Nelken, I. (1995). 'Dynamics of neuronal interactions' cannot be explained by 'neuronal transients'. *Proc. Biol. Sci.* 261, 407–410.
Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., and Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* 373, 515–518.
Ventura, V., Cai, C., and Kass, R. E. (2005). Trial-to-trial variability and its effect on time-varying dependency between two neurons. *J. Neurophysiol.* 94, 2928–2939.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 December 2009; paper pending published: 25 December 2009; accepted: 11 May 2010; published online: 02 July 2010.
Citation: Staude B, Grün S and Rotter S (2010) Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference. Front. Comput. Neurosci. ***4****:16. doi: 10.3389/fncom.2010.00016*

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 26 April 2010
doi: 10.3389/fncom.2010.00010

Estimating the amount of information carried by a neuronal population

Yunguo Yu[1], Marshall Crumiller[1], Bruce Knight[2] and Ehud Kaplan[1]*

[1] *Neuroscience Department, The Mount Sinai School of Medicine, New York, NY, USA*
[2] *Laboratory of Biophysics, The Rockefeller University, New York, NY, USA*

Edited by:
Jakob H. Macke, MPI for Biological Cybernetics, Germany

Reviewed by:
Simon R. Schultz, Imperial College London, UK
Stefano Panzeri, Italian Institute of Technology, Italy

Correspondence:
Ehud Kaplan, Department of Neuroscience, The Mount Sinai School of Medicine, One Gustave Levy Place, Box 1065, New York, NY 10029, USA.
e-mail: ehud.kaplan@mssm.edu

Although all brain functions require coordinated activity of many neurons, it has been difficult to estimate the amount of information carried by a population of spiking neurons. We present here a Fourier-based method for estimating the information delivery rate from a population of neurons, which allows us to measure the redundancy of information within and between functional neuronal classes. We illustrate the use of the method on some artificial spike trains and on simultaneous recordings from a small population of neurons from the lateral geniculate nucleus of an anesthetized macaque monkey.

Keywords: information, neural population, thalamus, redundancy

INTRODUCTION

The brain processes information, and it is therefore natural to estimate the amount of information that a neuron transmits to its targets. In the past, several methods that derive such estimates from the firing pattern (Optican and Richmond, 1987; Richmond and Optican, 1987; Richmond et al., 1987; Bialek et al., 1991; Rieke et al., 1997; Strong et al., 1998; Brenner et al., 2000) or membrane potential (Borst and Theunissen, 1999; DiCaprio, 2004) of individual neurons have been used. The information from spike trains was estimated by calculating the entropy associated with the various temporal patterns of spike discharge, using Shannon's formula (Shannon and Weaver, 1949).

Since all brain functions involve many neurons, it is desirable to provide similar information estimates for a neuronal population (Knight, 1972). To simply add up the information amounts from individual neurons in the population would be valid only if the neurons were all independent of one another, an assumption that usually is incorrect (see, for example, Zohary et al., 1994; Bair et al., 2001; Pillow et al., 2008). Approaches like the *Direct Method* (Strong et al., 1998) are impractical for a population, because the multi-dimensional space occupied by many spike trains can be sampled only sparsely by most neurophysiological experiments. Calculating the information carried by a population of many neurons thus has remained a challenge (Brown et al., 2004; Quiroga and Panzeri, 2009). At the same time, the need for such estimates has become increasingly urgent, since the technology of recording simultaneously from many neurons has become much more affordable and wide-spread, and data from such recordings are becoming common.

We describe here a method that estimates the amount of information carried by a population of spiking neurons, and demonstrate its use, first with simulated data and then with data recorded from the *lateral geniculate nucleus* (LGN) of an anesthetized macaque monkey.

MATERIALS AND METHODS

SURGICAL PREPARATION

The experimental methods were similar to those used in our lab in the past (Uglesich et al., 2009). Housing, surgical and recording procedures were in accordance with the National Institutes of Health guidelines and the Mount Sinai School of Medicine Institutional Animal Care and Use Committee. Adult macaque monkeys were anesthetized initially with an intramuscular injection of xylazine (Rompun, 2 mg/kg) followed by ketamine hydrochloride (Ketaset, 10 mg/kg), and then given propofol (Diprivan) as needed during surgery. Local anesthetic (xylocaine) was used profusely during surgery, and was used to infiltrate the areas around the ears. Anesthesia was maintained with a mixture of propofol (4 mg/kg-hr) and sufentanil (0.05 μg/kg-hr), which was given intravenously (IV) during the experiment. Propofol anesthesia has been shown to cause no changes in blood flow in the occipital cortex (Fiset et al., 1999), and appears to be optimal for brain studies. Cannulae were inserted into the femoral veins, the right femoral artery, the bladder, and the trachea. The animal was mounted in a stereotaxic apparatus. Phenylephrine hydrochloride (10%) and atropine sulfate (1%) were applied to the eyes. The corneas were protected with plastic gas-permeable contact lenses, and a 3-mm diameter artificial pupil was placed in front of each eye. The blood pressure, electrocardiogram, and body temperature were measured and kept within the physiological range. Paralysis was produced by an infusion of pancuronium bromide (Norcuron, 0.25 mg/kg-hr), and the animal was artificially respired. The respiration rate and stroke volume were adjusted to produce an end-expiratory value of 3.5–4% CO_2 at the exit of the tracheal cannula. Penicillin (750,000 units) and gentamicin sulfate (4 mg) were administered IM to provide antibacterial coverage, and dexamethasone was injected IV to prevent cerebral edema. A continuous IV flow (3–5 ml/kg-hr) of lactated Ringer's solution with 5% dextrose was maintained throughout the experiment to keep the animal properly hydrated,

and the urinary catheter monitored the overall fluid balance. Such preparations are usually stable in our laboratory for more than 96 h. The animal's heart rate and blood pressure monitored the depth of anesthesia, and signs of distress, such as salivation or increased heart rate, were watched for. If such signs appeared, additional anesthetics were administered immediately.

VISUAL STIMULATION

The eyes were refracted, and correcting lenses focused the eyes for the usual viewing distance of 57 cm. Stimuli were presented monocularly on a video monitor (luminance: 10–50 cd/m^2) driven by a VSG 2/5 stimulator (CRS, Cambridge, UK). The monitor was calibrated according to Brainard (1989) and Wandell (1995). Gamma corrections were made with the VSG software and photometer (OptiCal). Visual stimuli consisted of homogeneous field modulated in luminance according to a pseudo-random naturalistic sequence (van Hateren, 1997). Eight second segments of the luminance sequences were presented repeatedly 128 times ('repeats'), alternating with 8 s non-repeating ('uniques') segments of the sequence (Reinagel and Reid, 2000). In addition, we used steady (unmodulated) light screens and dark screens, during which spontaneous activity was recorded.

ELECTROPHYSIOLOGICAL RECORDING

A bundle of 16 stainless steel microwires (25 μ) was inserted into a 22 gauge guard tube, which was inserted into the brain to a depth of 5 mm above the LGN. The microwire electrodes were then advanced slowly (in 1 μ steps) into the LGN, until visual responses to a flashing full field screen were detected. The brain over the LGN was then covered with silicone gel, to stabilize the electrode bundle. Based on the electrode depth, dominant eye sequence and cell properties (Kaplan, 2007), we are confident that all the electrodes were within the parvocellular layers of the LGN. The receptive fields of the recorded cells covered a relatively small area (~4 in diameter), which suggests that the electrodes bundle remained relatively compact inside the LGN.

The output of each electrode was amplified, band-pass filtered (0.75–10 kHz), sampled at 40 kHz and stored in a *Plexon* MAP computer for further analysis.

DATA ANALYSIS

Spike sorting

Sorting procedures. The spike trains were first thresholded (SNR ≥5) and sorted using a template-matching algorithm under visual inspection (*Offline Sorter,* Plexon Inc., Dallas, TX, USA). In most cases, spikes from several neurons recorded by a given electrode could be well-separated by this simple procedure. In more difficult cases, additional procedures (peak- or valley-seeking, or multivariate t-distributions) (Shoham et al., 2003) were employed. Once the spikes were sorted, a firing times list was generated for each neuron and used for further data analysis.

Quality assurance. To ensure that all the spikes in a given train were fired by the same neuron, we calculated for each train the interspike interval (ISI) histogram. If we found intervals that were shorter than the refractory period of 2 ms, the spike sorting was repeated to eliminate the misclassified spikes. We ascertained that all the analyzed data came from responsive cells by calculating the coefficients of variation of the peristimulus time histogram bin counts for the responses to the repeated and unique stimuli, and taking the ratio of these two coefficients. Only cells for which that ratio exceeded 1.5 were included in our analysis.

Generation of surrogate data

To test our method we generated synthetic spike trains from a Poisson renewal process, in which the irregularities of neuronal spike times are modeled by a stochastic process whose mathematical properties are well defined. Recent interest and success in modeling a neuron spike-train as an inhomogeneous Poisson process (Pillow et al., 2005, 2008; Pillow and Simoncelli, 2006) led us to that choice.

Firing rates and input. Our modeling necessarily addressed two major features of the laboratory data. The nine real neurons show a range of mean firing rates, from 3.04 impulses per second (ips) to 28.72 ips, which span an order of magnitude. To mimic this, we gave our 12 model cells 12 inputs which consecutively incremented by a factor of $10^{(1/11)}$, to give firing rates spanning an order of magnitude. The second major feature was that our laboratory neurons evidently received inputs processed in several ways following the original retinal stimulus. To make a simple caricature of this, we drove each of our Poisson model neurons with a separate input that was a weighted mean admixture of two van Hateren-type stimuli. The first was that which we used in the laboratory and the second was the time-reversal of that stimulus. Calling these A and B, the stimuli were of the form $S = (1 - x)\cdot A + x\cdot B$, where the admixture variable x took on 12 equally spaced values starting with 0 and ending with 1. As shown in **Table 1**, the pairs (admixture, mean rate) were chosen in a manner that allowed the whole grouping of model cells to be divided into smoothly changing subsets in different ways, and evenly distributed the range of properties across all cells.

Estimation of the information delivered by a subset of neurons

If we have data from numerous parallel spike trains, the familiar *Direct method* (Strong et al., 1998) for computing signal information delivered requires an impractical time span of data. As a

Table 1 | Parameters for stimulating the surrogate neurons. Each surrogate neuron was driven by a mixture of two *van Hateren* inputs, chosen to cover uniformly the range of firing rates and mixture ratios.

Cell #	1	2	3	4	5	6	7	8	9	10	11	12
Firing rate	4.98	6.18	7.58	9.38	11.42	14.13	17.47	21.64	26.79	32.74	40.60	50.09
Admixture	0	0.27	0.55	0.82	0.09	0.36	0.64	0.91	0.18	0.45	0.73	1

practical alternative we advance a straightforward multi-cell generalization of a method of information computation from basis-function coefficients.

Shannon has observed (Shannon and Weaver, 1949, Chapter 4; see also Shannon, 1949) that the probability structure of a stochastic signal over time may be well approximated in many different ways by various equivalent multivariate distribution density functions of high but finite dimension. He further observed that when some specific scheme is used to characterize both the distribution of signal-plus-noise and the distribution of noise alone, the information quantity one obtains for the signal alone, by taking the difference of the information quantities (commonly called 'entropies') evaluated from the two distributions, has a striking invariance property: the value of the signal information is universal, and does not depend on which of numerous possible coordinate systems one has chosen in which to express the multivariate probability density (see extensive bibliography, and discussion, in Rieke et al., 1997, chapter 3). We will follow Shannon (1949), whose choice of orthonormal functions was Fourier normalized sines and cosines, over a fixed, but long, time span T. This choice has the added virtue of lending insight into the frequency structure of the information transfer under study.

Here we outline our approach for obtaining the signal-information rate, or 'mutual information rate', transmitted by the simultaneously recorded spikes of a collection of neurons. The mathematical particulars are further elaborated in the Appendix. Following Shannon (1949), if one has a data record that spans a time T, it is natural to use the classical method of Fourier analysis to resolve that signal into frequency components, each of which speaks of the information carried by frequencies within a frequency bandwidth of $1/T$. If this is repeated for many samples of output, one obtains a distribution of amplitudes within that frequency band. In principle, that probability distribution can be exploited to calculate how many bits would have to be generated per second (the information rate) to describe the information that is being transmitted within that frequency band.

However, part of that information rate represents not useful information but the intrusion of noise. To quantify our overestimate we may repeat the experiment many times without variation of input stimulus, and in principle may employ the same hypothetical means as before to extract the 'information', which now more properly may be called 'noise entropy'. When this number is subtracted from the previous, we obtain the mutual information rate, in bits per second, carried by the spikes recorded from that collection of neurons.

In order to reduce the above idea to practice, we have exploited the following fact (which apparently is not well known nor easily found in the literature): if our response forgets its past history over a correlation time span that is brief compared to the experiment time span, T, then the central limit theorem applies, and our distribution of signal measurements within that narrow bandwidth will follow a Gaussian distribution. If we are making simultaneous recordings from a collection of neurons, their joint probability distribution within that bandwidth will be multivariate Gaussian. A Gaussian with known center of gravity is fully characterized by its variance, and similarly a multivariate Gaussian by its covariance matrix. Such a covariance matrix, which can be estimated directly from the data, carries with it a certain entropy. By calculating the covariance matrices for responses to both unique and repeated stimuli, one can determine the total signal information flowing through each frequency channel for a population of neurons.

To verify that our Gaussian assumption is valid, we have applied to our Fourier-coefficient sample sets two standard statistical tests that correctly identify a sample as Gaussian with 95% accuracy. For our 12 surrogate cells and 9 laboratory LGN cells, the degree of verification across the frequency range for 2560 distribution samples (160 Hz × 8 bins/Hz × 2, with each sine and cosine term sampled 128 times) is shown in **Table 2**. Because of its importance, we return to this issue in the Discussion, where further evidence is provided for the Gaussian nature of the underlying distributions.

RESULTS

ANALYSIS OF SIMULATED SPIKE TRAINS

Entropy vs temporal frequency

In anticipation of analyzing simultaneous laboratory records of actual neurons, we have created 12 Poisson model neurons with firing rates that overlap those of our laboratory neurons and with inputs as discussed above in Section 'Materials and Methods', presented at the same rate (160 Hz) used in the laboratory experiments. **Figure 1** shows, for a single simulated cell, the entropy rate per frequency, for responses to unique and repeat stimuli. The entropy from the responses to the unique stimulus (signal plus noise) exceeds that of the responses to the repeated stimulus (noise alone) at low frequencies, and the two curves converge near the monitor's frame-rate of 160 Hz, beyond which signal-plus-noise is entirely noise. Hence we will terminate the sum in (Eq. A26) at that frequency. The difference between the two curves at any temporal frequency is the mutual information rate at that frequency.

Table 2 | The Fourier coefficients for the surrogate and LGN data follow a Gaussian distribution. We sampled the Fourier coefficients 128 times for each of the 2560 sine and cosine terms that we tested for each cell. Each distribution was tested with two standard tests for normality: the Shapiro–Wilk's test and the Lilliefors test. The percentage of distributions that passed each test at the $p < 0.05$ significance level was calculated for each cell, and the table gives the mean and standard deviation for the test results.

	Repeats (% passed)		Uniques (% passed)	
TEST	**SHAPIRO–WILK**	**LILLIEFORS**	**SHAPIRO–WILK**	**LILLIEFORS**
Surrogate data (12 cells)	95.3 ± 0.31	95.2 ± 0.34	95.3 ± 0.41	95.1 ± 0.3
LGN cells (9 cells)	94.9 ± 1.62	94.6 ± 0.35	93.9 ± 1.31	94.6 ± 0.45

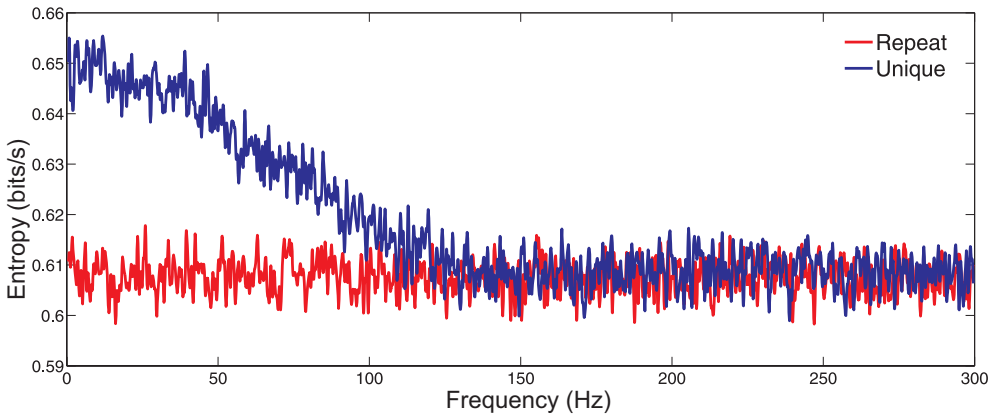


FIGURE 1 | Entropy per frequency conveyed by a single surrogate neuron. The signal-plus-noise entropy (derived from the unique stimuli) is shown in blue, and the noise entropy (from the repeated stimulus) is shown in red. The data shown are typical of data from other cells.

Single cell information

For the 12 model cells, the cumulative sum of information over frequency (Eq. A26) is given in **Figure 2** (left frame). We note that all the curves indeed finish their ascent as they approach 160 Hz. More detailed examination shows a feature that is not obvious: the output information rate of a cell reflects its input information rate, and the input information rate of a mixed, weighted mean input is less than that of a pure, unmixed input. This accounts for the observation that the second-fastest cell (cell 11, with a near even mixture) delivers information at only about half the rate of the fastest (cell 12).

Group information

We turn now to the information rate of a *group* of cells, firing in parallel in response to related stimuli. We proceed similarly to what is above, but use the multi-cell equation (Eq. A25) and its cumulative sum over frequencies. As a first exercise we start with the slowest-firing surrogate cell and then group it with the next-slowest, next the slowest 3 and so on up to the fastest; the set of cumulative curves we obtain from these groupings are shown in the left frame of **Figure 3**. Again we see that the accumulation of information appears to be complete earlier than the frame-rate frequency of 160 Hz.

REDUNDANCY AND SYNERGY AMONG NEURONS IN A POPULATION

Redundancy

The mutual information communicated by a group of cells typically falls below the sum of the mutual information amounts of its constituent members. This leads us to define a measure of information redundancy. The redundancy of a cell with respect to a group of cells can be intuitively described as the proportion of its information already conveyed by other members of the group. For example, if a cell is added to a group of cells and 100% of its information is novel, then it has 0 redundancy. If, on the other hand, the cell brings *no* new information to the group, then it contains only redundant information, and it therefore has redundancy 1. With this in mind, we define the redundancy of a cell C, after being added to a group G, as:

$$r_{c,g} = \left(I(c) - \left(I(g+c) - I(g)\right)\right)/I(c).$$

According to this formula, if all the information of the additional cell appears as added information in the new group, then that cell's redundancy is zero.

The procedure of information redundancy evaluation is general, and can be applied to the addition of any cell to any group of cells. Thus for the cell groups of **Figure 3**, we can evaluate the redundancy of each newly added cell not only upon its addition to the group but also thereafter. This is shown for the 70 resulting redundancies, in **Figure 4** (Left).

Synergy

When the total information conveyed by several neurons exceeds the sum of the individual ones, the neurons are synergistic (Gawne and Richmond, 1993; Schneidman et al., 2003; Montani et al., 2007). When this happens, our formula yields a negative redundancy value.

ANALYSIS OF MONKEY LGN SPIKE TRAINS

We now apply the same techniques to simultaneous laboratory recordings of 9 parvocellular cells from the LGN of a macaque monkey, responding to a common full-field naturalistic stimulus (van Hateren, 1997; Reinagel and Reid, 2000).

Figure 2 (right frame) shows the single cell cumulative information of these neurons as frequency increases. In two obvious ways their behavior differs from that of the Poisson model neurons. First, at low frequency there is a qualitative difference indicative of initially very small increment, which differs from the Poisson model's initial linear rise. Second, the real geniculate neurons show a substantial heterogeneity in the shape of their rise curves. For example, the second most informative cell (cell 8), has obtained half its information from frequencies below 40 Hz, while the most informative cell (cell 9) has obtained only 11% of its information from below that frequency.

The right frame of **Figure 3** shows for LGN cells the accumulating multineuron group information, while the left frame shows it for the surrogate data.

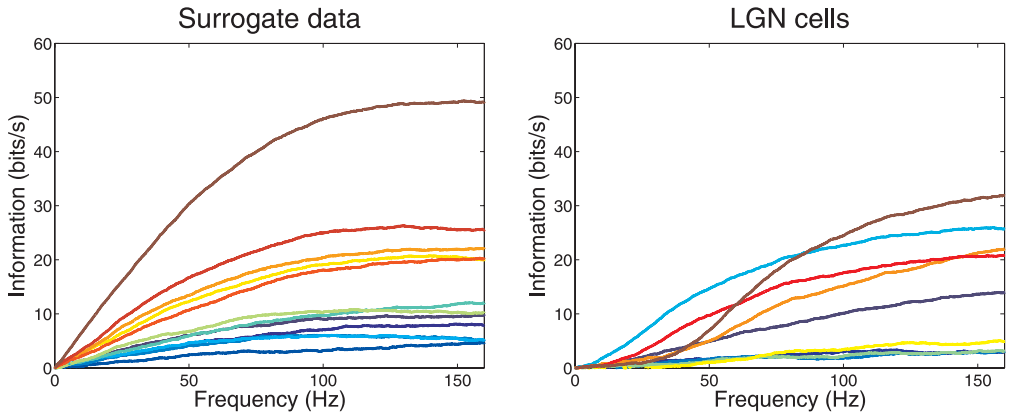


FIGURE 2 | Cumulative information rate vs frequency for 12 surrogate Poisson model neurons and 9 LGN cells. The firing rates of the various neurons in the two groups were similar.

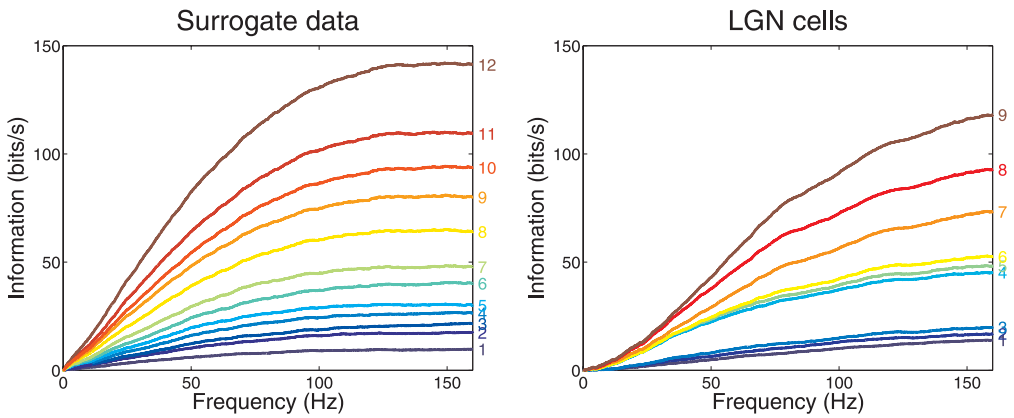


FIGURE 3 | Group information vs frequency for our Poisson model surrogate neurons and 9 LGN cells. The group size is indicated to the right of the cumulative curve for each group. The neurons were ranked according to their firing rate. The first group contained only the slowest firing neuron, and each new group was formed by adding the next ranking cell.

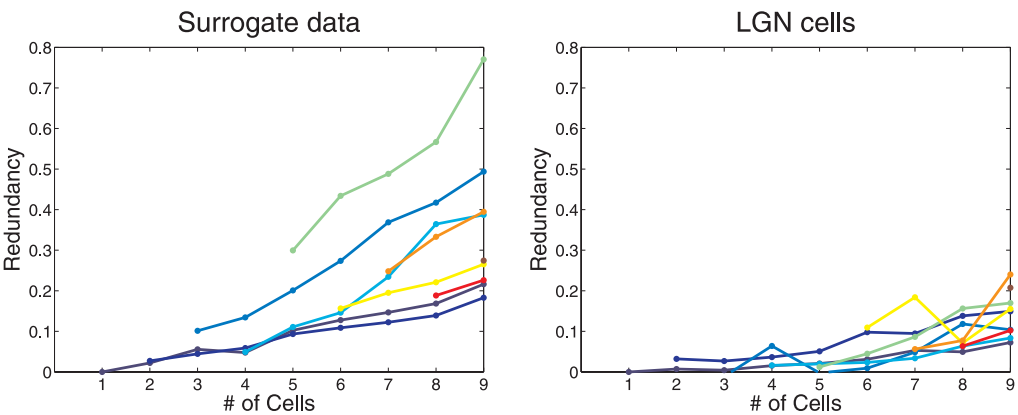


FIGURE 4 | Accumulating redundancy as more cells are added to a population. The cells are added in order of their mean firing rates, starting with the slowest firing cells, with each cell taking its turn as a starting point for a new population.

Redundancy in surrogate and real LGN neurons

Figure 4 (right frame) compares the redundancy over the 9 LGN cells with what was shown for the first 9 Poisson model neurons in **Figure 4** (left frame). The pair of sharp features at cells 4 and 7 might be attributed to difficulties in spike separation. Note that the redundancy of real neurons appears to be quite different from that of their Poisson model counterparts: as cluster size increases, real cells manifest a stronger tendency than our simulated neurons to remain non-redundant. This implies that the different LGN neurons are reporting with differences in emphasis on the various temporal features of their common stimulus.

DISCUSSION

THE VALIDITY OF THE GAUSSIAN ASSUMPTION

Our method exploits the theoretical prediction that the distribution of each stochastic Fourier coefficient of our data should be Gaussian. Our evidence supports this prediction. A standard visual check is to normalize a distribution by a *Z*-score transformation and plot its quantiles against those of a standard Gaussian. If the distribution is likewise Gaussian, the points will fall near a unit-slope straight line through the origin. **Figure 5** shows two typical cases, each with 128 points: surrogate data in the left frame and LGN cell data on the right. Both show good qualitative confirmation of the Gaussian assumption.

We have proceeded to apply to our numerous Fourier coefficient distributions two standard statistical tests for Gaussian distribution: the Shapiro–Wilk test and the Lilliefors test. Both are designed to confirm that a sample was drawn from a true Gaussian distribution in 95% of cases. **Table 2** shows that in almost all cases for both unique and repeat responses of our 12 surrogate and 9 LGN cells our distributions passed both tests at the 95% significance level.

SMALL SAMPLE BIAS

In the extraction of mutual information from spike data, traditional methods suffer from a bias due to the small size of the sample. We checked the Fourier method for such bias by dividing our sets of 128 runs into subsets of 64, 32 and 16 runs. The results for one surrogate cell (number 12) and one LGN cell (number 8) are shown in **Figure 6**. These results are typical, and show no clear small-sample

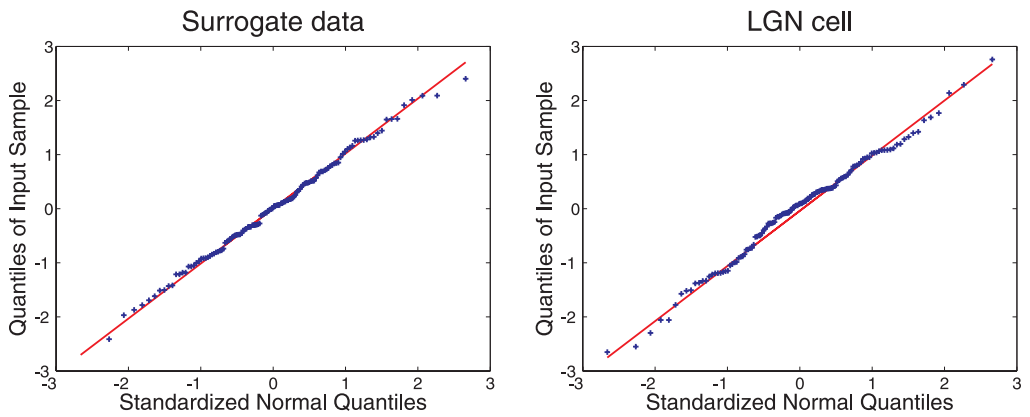


FIGURE 5 | Q–Q plots for the Fourier coefficients of one surrogate cell (#6) and one LGN cell (#4). The data are typical of data from other cells. The fact that the data points hug the $y = x$ line demonstrates the Gaussian nature of the distributions of the Fourier coefficients.

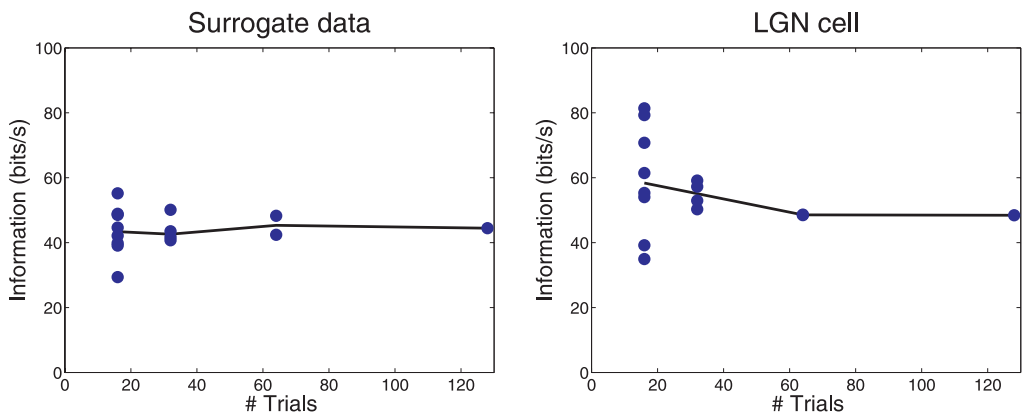


FIGURE 6 | The effect of the number of trials on information calculation. Data are from surrogate cell #12 and LGN cell #8, which were typical of other cells. Solid symbols show the information calculated from individual segments of the record. The solid line connects the medians of the samples. Note the rapid convergence of the information estimates as the number of trials increases.

bias. We also notice that, for these data, a sample of 64 runs gives a mutual information estimate reliable to better than ±10%. A summary of small-sample bias and estimated reliability for several recent techniques for calculating spike-train mutual information is given by Ince et al. (2009) (their Figure 1).

In addition to the number of data segments, the number of spikes used in estimating the mutual information is also an important factor, and we discus it further at the end of the Appendix.

SUMMARY AND CONCLUSIONS

We have presented a new method for calculating the amount of information transmitted by a neuronal population, and have applied it to populations of simulated neurons and of monkey LGN neurons. Since the method can be used also to calculate the information transmitted by individual cells, it provides an estimate of the redundancy of information among the members of the population. In addition, the method reveals the temporal frequency bands at which the communicated information resides.

The new method fills a gap in the toolbox of the modern neurophysiologist, who now has the ability to record simultaneously from many neurons. The methodology presented here might permit insights regarding the mutual interactions of neuronal clusters, an area that has been explored less than the behavior of single neurons or whole organisms.

APPENDIX

Suppose we have a stochastic numerical data-stream that we will call $u(t)$, and which becomes uncorrelated for two values of t that are separated by a time interval greater than a maximum correlation time-interval t^*. That is to say, if $t_2 - t_1 > t^*$, then $u(t_2)$ and $u(t_1)$ are *independent* random variables in the probability sense. Suppose now that in the laboratory, by running the probabilistically identical experiment repeatedly, we gather N realizations (samples) of $u(t)$, the n^{th} of which we will call $u^{(n)}(t)$. Suppose further that we collect each data sample over a time-span T that is large compared to the correlation time interval t^*.

We can represent each sample $u^{(n)}(t)$ to whatever accuracy we desire, as a discrete sequence of numbers in the following way. Over the time interval $t = 0$ to $t = T$, we choose a set of functions $\varphi_m(t)$ that are *orthonormal* in the sense that they have the property:

$$\int_0^T dt\varphi_q(t)\varphi_r(t) = \delta_{qr} (= 1 \text{ if } q = r, \text{ else} = 0). \tag{A1}$$

Then $u^{(n)}(t)$ may be represented as a weighted sum of these basis functions:

$$u^{(n)}(t) = \sum_q u_q^{(n)}\varphi_q(t) \tag{A2}$$

where the weighting coefficients $u_m^{(n)}$ may be evaluated from the data by,

$$u_m^{(n)} = \int_0^T dt\varphi_m(t)u^{(n)}(t). \tag{A3}$$

This claim can be verified if we substitute (Eq. A2) into (Eq. A3) and then use (Eq. A1) to evaluate the integral. Here our choice of the $\varphi_m(t)$ will be the conventional normalized sinusoids:

$$\varphi_m(t) = \begin{cases} \sqrt{2/T}\sin 2\pi((m+1)/2)(t/T) \text{ for } m \text{ odd} \\ \sqrt{2/T}\cos \pi m(t/T) \text{ for } m \text{ even} \end{cases} \tag{A4}$$

It is a straightforward exercise to show that these functions have the property required by (Eq. A1).

Now let us see what follows from $T >> t^*$. Divide the full time-span T into K sub-intervals by defining the division times:

$$t_k = (k/K)T \tag{A5}$$

and define the integrals over shorter sub-intervals:

$$A_{m,k}^{(n)} = \int_{t_{k-1}}^{t_k - t^*} dt\varphi_m(t)u^{(n)}(t) \tag{A6}$$

$$B_{m,k}^{(n)} = \int_{t_k - t^*}^{t_k} dt\varphi_m(t)u^{(n)}(t) \tag{A7}$$

from which (Eq. A3) tells us that the Fourier coefficient $u_m^{(n)}$ is given by,

$$u_m^{(n)} = \sum_k A_{m,k}^{(n)} + \sum_k B_{m,k}^{(n)}. \tag{A8}$$

But we note that the measure of the support of the integral (Eq. A7) is smaller than that of (Eq. A6) by the ratio $t^*/((T/K) - t^*)$, and if we can pick T long enough, we can make that ratio as close to zero as we choose. So the second sum in (Eq. A8) is negligible in the limit. But now note that, because they are all separated from each other by a correlation time, the individual terms in the first sum are realizations of independent random variables. If the distribution of an individual term in the sum is constrained in any one of a number of non-pathological ways, and if there are a sufficient number of members in the sum, then the central limit theorem states that the distribution of the sum approaches a Gaussian.

In the more general case, where we have several simultaneous correlated numerical data-streams, the argument runs exactly the same way. If, for many repeated samples, at a particular frequency we compute the Fourier coefficient for each, to estimate a multivariate probability density, then from a long enough time span, by the multivariate central limit theorem that density will approach a multivariate Gaussian. Simply because the notation is easier, we elaborate the univariate case first.

Specializing, for cell response we use the spike train itself, expressed as a sequence of δ-functions, so for the r^{th} realization $u^{(r)}(t)$ of the stochastic spike-train variable $u(t)$, we have:

$$u^{(r)}(t) = \sum_{n=1}^{N_r} \delta(t - t_{(r)n}) \tag{A9}$$

where $t_{(r)n}$ is the time of the n^{th} spike of the r^{th} realization, and N_r is the total number of spikes that the cell under discussion fires in that realization.

Substituting this and also (Eq. A4) into (Eq. A3) we see that the integral may be performed at once. In the cosine case of (Eq. A4) it is,

$$u_m^{(r)} = \sqrt{2/T}\sum_{n=1}^{N_r}\cos \pi m(t_{(r)n}/T) \tag{A10}$$

Before proceeding further we look back at Eq. A8 and note that, because a cosine is bounded between +1 and −1, every term in the sums of (Eq. A8) is bounded in absolute value by $\sqrt{2/T}$ times the number of spikes in that sub-interval. As real biology will not deliver a cluster of spikes overwhelmingly more numerous than the local mean rate would estimate, the *distribution* of each term in the stochastic sum cannot be heavy-tailed, and we may trust the central limit theorem.

Thus we may estimate that the probability density function for the stochastic Fourier coefficient variable u_m is of the form,

$$p_m(u_m) = (2\pi V_m)^{-1/2}\exp(-(u_m - \overline{u}_m)^2/2V_m). \tag{A11}$$

where,

$$\overline{u}_m = \langle u_m \rangle_{p_m} \cong \frac{1}{R}\sum_{r=1}^{R}(u_m^{(r)}), \tag{A12}$$

$$V_m = \langle (u_m^{(r)} - \overline{u}_m)^2 \rangle_{p_m} \cong \frac{1}{R-1}\sum_{r=1}^{R}(u_m^{(r)} - \overline{u}_m)^2. \tag{A13}$$

The right-hand-most expressions in (Eq. A12), (Eq. A13) testify that $\overline{u}_m$ and V_m can be estimated directly from the available laboratory data.

What is the information content carried by the Gaussian (Eq. A11)? The relevant integral may be performed analytically:

$$I(p_m) = -\int du_m(\ln p_m(u_m))p_m(u_m) = \frac{1}{2}\ln((2\pi e)V_m). \tag{A14}$$

For a signal with finite forgetting-time the stochastic Fourier coefficients (Eq. A10) at different frequencies are statistically independent of one another, so that the signal's full multivariate probability distribution in terms of Fourier coefficients is given by,

$$p(u_1, u_2, \quad) = \prod_m p_m(u_m). \tag{A15}$$

It is easily shown that if a multivariate distribution is the product of underlying univariate building blocks, then its information content is the sum of the information of its components, whence

$$I(p) = \sum_{m=0}^{M-1} I(p_m) = \frac{1}{2}\sum_{m=0}^{M-1}\ln((2\pi e)V_m). \tag{A16}$$

Observing (Eq. A13) we note that this can be evaluated from available laboratory data.

Generalization of the information rate calculation to the case of multiple neurons is conceptually straightforward but notationally messy due to additional subscripts. The rth realization's spike train from the qth neuron (out of a total of Q neurons) may be defined as a function of time $u_{(q)}^{(r)}(t)$ just as in (Eq. A9) above, and from our orthonormal set of sines and cosines we may find the Fourier coefficient $u_{(q)m}^{(r)}$. This number is a realization drawn from an ensemble whose multivariate probability density function we may call:

$$p_m(u_{(1)m}, u_{(2)m}, .., u_{(Q)m}). \tag{A17}$$

This density defines a vector center of gravity $\overline{u}_m$ whose Q components are of the form:

$$\overline{u}_{(q)m} = \langle u_{(q)m} \rangle_{p_m} \cong \frac{1}{R}\sum_{r=1}^{R} u_{(q)m}^{(r)}, \tag{A18}$$

and similarly it defines a covariance matrix V_m whose (q,s)th matrix element is given by,

$$\begin{aligned} V_{(q,s)m} &= \left\langle (u_{(q)m} - \overline{u}_{(q)m})(u_{(s)m} - \overline{u}_{(s)m}) \right\rangle_{p_m} \\ &\cong \frac{1}{R-1}\sum_{r=1}^{R}(u_{(q)m}^r - \overline{u}_{(q)m}^r)(u_{(s)m}^r - \overline{u}_{(s)m}^r). \end{aligned} \tag{A19}$$

This covariance matrix has a matrix inverse A_m:

$$A_m = V_m^{-1}. \tag{A20}$$

Clearly (Eq. A18) and (Eq. A19) are the multivariate generalizations of (Eq. A12) and (Eq. A13) above. The central limit theorem's multivariate Gaussian generalization of (Eq. A11) is,

$$p_m(u_{(1)m}, .., u_{(Q)m}) = ((2\pi)^Q \det V_m)^{(-1/2)}\exp\left(-\frac{1}{2}\sum_{q,s}(u_{(q)m} - \overline{u}_{(q)m})A_{(q,s)m}(u_{(s)m} - \overline{u}_{(s)m})\right). \tag{A21}$$

This expression becomes less intimidating in new coordinates $Z_{(q)}$ with new origin located at the center of gravity and orthogonally turned to diagonalize the covariance matrix (Eq. A19). We need not actually undertake this task. Call the eigenvalues of the covariance matrix

$$\lambda_{(1)m}, .., \lambda_{(Q)m}. \tag{A22}$$

Under the contemplated diagonalizing transformation, the double sum in the exponent collapses to a single sum of squared terms, and in the new coordinates p_m becomes,

$$\hat{p}_m(Z_1, .., Z_Q) = \prod_{q=1}^{Q}(2\pi\lambda_{(q)m})^{-1/2}\exp(-Z_q^2/2\lambda_{(q)m}), \tag{A23}$$

a form that is familiar from (Eq. A15) above. Its corresponding information is the sum of those of the individual terms of the product and is

$$I(p_m) = \frac{1}{2}\sum_{q=1}^{Q}\ln((2\pi e)\lambda_{(q)m}). \tag{A24}$$

Shannon (1949, chapter 4), in a formally rather analogous context, has noted that much care is needed in the evaluation of expressions similar to (Eq. A24) from laboratory data. The problem arises here if the eigenvalues approach zero (and their logarithms tend to $-\infty$) before the sum is completed. However, the information in signal-plus-noise in the mth coefficient, expressed by (Eq. A24) is

not of comparable interest to the information from signal alone. With some caution, this signal-alone information contribution may be obtained by subtracting from (Eq. A24) a similar expression for noise alone, taken from additional laboratory data in which the *same* stimulus was presented repeatedly. If we use 'μ' to annotate the eigenvalues of the covariance matrix which emerges from these runs, then the information difference of interest, following from (Eq. A24) is

$$I_m(\text{signal alone}) = \frac{1}{2}\sum_{q=1}^{Q}\left\{\ln((2\pi e)\lambda_{(q)m}) - \ln((2\pi e)\ \ _{(q)m})\right\} = \frac{1}{2}\sum_{q=1}^{Q}\ln\left(\frac{\lambda_{(q)m}}{\ \ _{(q)m}}\right). \quad \text{(A25)}$$

Equation A25 expresses the multi-cell information contributed by the *m*th frequency component. To obtain the total multi-cell information, it is to be summed over increasing *m* until further contributions become inappreciable.

An entirely analogous procedure applies to obtain the information of signal alone for an individual cell. Call the variance of the *m*th frequency component of the unique runs V_{mu}, and that of the repeat runs V_{mr}. Each will yield a total information rate expressed by (Eq. A16) above, and their difference, the information rate from signal alone, consequently will be:

$$I\left(\text{cell, signal alone}\right) = \frac{1}{2}\sum_{m=0}^{M-1}\ln\left(\frac{V_{mu}}{V_{mr}}\right). \quad \text{(A26)}$$

In the data analysis in the main text, the single-cell sums (Eq. A16), for both uniques and repeats, approached a common, linearly advancing value which they achieved near 160 Hz, which is the stimulus frame-rate. Consequently, the summation over frequency of signal only information was cut off at that frequency, both for single cells (see Eq. A26) and for combinations of cells.

In both the simulations and the experiments, each run was of $T = 8$ s duration. In consequence the orthonormalized sines and cosines of (Eq. A4) advanced by steps of 1/8 Hz.

EFFECT OF THE NUMBER OF RESPONSE SPIKES

With reference to small-sample bias, a further word is appropriate here regarding our methodology. If the number of runs is modest, the total number of spikes in response to the repeated stimulus may show a significant statistical fluctuation away from the total number of spikes in response to the unique runs. In this case, the asymptotic high-frequency entropy values, as seen in our **Figure 1**, will not quite coincide, and consequently the accumulated mutual information will show an artifactual small linear drift with increasing frequency. This introduces a bit of uncertainty in the cut-off frequency and in the total mutual information. This asymptotic drift may be turned into a more objective way to evaluate the total mutual information. In cases where the problem arises, we divide our repeat runs into two subsets: the half with the most spikes and the half with the least. Accumulating both mutual information estimates at high frequency, we linearly extrapolate both asymptotic linear drifts back to zero frequency, where they intersect at the proper value of mutual information.

ACKNOWLEDGMENTS

This work was supported by NIH grants *EY16224, EY16371, NIGM71558 and Core Grant EY12867.* We thank Drs. J. Victor, Y. Xiao and A. Casti for their help with this project.

REFERENCES

Bair, W., Zohary, E., and Newsome, W. T. (2001). Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J. Neurosci.* 21, 1676–1697.

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

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

Brainard, D. H. (1989). Calibration of a computer controlled color monitor. *Color Res. Appl.* 14, 23–34.

Brenner, N., Strong, S. P., Koberle, R., Bialek, W., and de Ruyter van Steveninck, R. R. (2000). Synergy in a neural code. *Neural Comput.* 12, 1531–1552.

Brown, E. N., Kass, R. E., and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 7, 456–461.

DiCaprio, R. A. (2004). Information transfer rate of nonspiking afferent neurons in the crab. *J. Neurophysiol.* 92, 302–310.

Fiset, P., Paus, T., Daloz, T., Plourde, G., Meure, P., Bonhomme, V., Haij-Ali, N., Backman, S. B., and Evans, A. (1999). Brain mechanisms of propofol-induced loss of conciousness in humans: a positron emission tomography study. *J. Neurosci.* 19, 5506–5513.

Gawne, T. J., and Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? *J. Neurosci.* 13, 2758–2771.

Ince, R. A. A., Petersen, R. S., Swan, D. C., and Panzeri, S. (2009). Python for information theoretic analysis of neural data. *Front. Neuroinformatics* 3:4. doi: 10.3389/neuro.11.004.2009.

Kaplan, E. (2007). "The M, K, and P streams in the primate visual system: what do they do for vision?" Chapter 1.16, in *The Senses* eds R. Masland and T. D. Albright (San Diego: Elsevier), 369–382.

Knight, B. W. (1972). Dynamics of encoding in a population of neurons. *J. Gen. Physiol.* 59, 734–766.

Montani, F., Kohn, A., Smith, M. A., and Schultz, S. R. (2007). The role of correlations in direction and contrast coding in the primary visual cortex. *J. Neurosci.* 27, 2338–2348.

Optican, L. M., and Richmond, B. J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. information theoretic analysis. *J. Neurophysiol.* 57, 162–178.

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, 11003–11013.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999.

Pillow, J. W., and Simoncelli, E. P. (2006). Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis. *J. Vis.* 6, 414–428.

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

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

Richmond, B. J., and Optican, L. M. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. quantification of response waveform. *J. Neurophysiol.* 57, 147–161.

Richmond, B. J., Optican, L. M., Podell, M., and Spitzer, H. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J. Neurophysiol.* 57, 132–146.

Rieke, F., Warland, D., Steveninck, R. D., and Bialek, W. (1997). *Spikes: Exploring the Neural Code.* Cambridge, MA: MIT Press.

Schneidman, E., Bialek, W., and Berry, M. J. (2003). Synergy, redundancy, and independence in population codes. *J. Neurosci.* 23, 11539–11553.

Shannon, C., and Weaver, W. (1949). *A Mathematical Theory of Communication.* Chicago, IL: University of Illinois Press.

Shannon, C. E. (1949). Communication in the presence of noise. *Proc. IEEE* 37, 10–21.

Shoham, S., Fellows, M. R., and Normann, R. A. (2003). Robust, automatic spike sorting using mixtures of multivariate t-distributions. *J. Neurosci. Methods* 127, 111–122.

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, 197–200.

Uglesich, R., Casti, A., Hayot, F., and Kaplan, E. (2009). Stimulus size dependence of information transfer from retina to thalamus. *Front. Syst. Neurosci.* 3:10. doi: 10.3389/neuro.06.010.2009.

van Hateren, J. H. (1997). Processing of natural time series of intensities by the visual system of the blowfly. *Vis. Res.* 37, 3407–3416.

Wandell, B. A. (1995). *Foundations of Vision*. Sunderland, MA: Sinauer Associates.

Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 December 2009; paper pending published: 31 December 2009; accepted: 29 March 2010; published online: 26 April 2010.

Citation: Yu Y, Crumiller M, Knight B and Kaplan E (2010) Estimating the amount of information carried by a neuronal population. Front. Comput. Neurosci. ***4****:10. doi: 10.3389/fncom.2010.00010*

frontiers in
COMPUTATIONAL NEUROSCIENCE

METHODS ARTICLE
published: 19 March 2010
doi: 10.3389/fncom.2010.00004

Quantifying auditory event-related responses in multichannel human intracranial recordings

Dana Boatman-Reich[1,2]*, Piotr J. Franaszczuk[1], Anna Korzeniewska[1], Brian Caffo[3], Eva K. Ritzl[1], Sarah Colwell[1] and Nathan E. Crone[1]

[1] *Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA*
[2] *Department of Otolaryngology, Johns Hopkins School of Medicine, Baltimore, MD, USA*
[3] *Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*

Edited by:
Philipp Berens, Max-Planck Institute for Biological Cybernetics, Germany; Baylor College of Medicine, USA

Reviewed by:
Bjoern Schelter, University of Freiburg, Germany
Jonathan Z. Simon, University of Maryland, USA
Moritz Grosse-Wentrup, Max-Planck Institute for Biological Cybernetics, Germany

***Correspondence:**
Dana Boatman-Reich, Department of Neurology, Epilepsy Division, Johns Hopkins School of Medicine, 600 North Wolfe Street, Meyer 2-147, Baltimore, MD 21287, USA.
e-mail: dboatma@jhmi.edu

Multichannel intracranial recordings are used increasingly to study the functional organization of human cortex. Intracranial recordings of event-related activity, or electrocorticography (ECoG), are based on high density electrode arrays implanted directly over cortex, combining good temporal and spatial resolution. Developing appropriate statistical methods for analyzing event-related responses in these high dimensional ECoG datasets remains a major challenge for clinical and systems neuroscience. We present a novel methodological framework that combines complementary, existing methods adapted for statistical analysis of auditory event-related responses in multichannel ECoG recordings. This analytic framework integrates single-channel (time-domain, time–frequency) and multichannel analyses of event-related ECoG activity to determine statistically significant evoked responses, induced spectral responses, and effective (causal) connectivity. Implementation of this quantitative approach is illustrated using multichannel ECoG data from recent studies of auditory processing in patients with epilepsy. Methods described include a time–frequency matching pursuit algorithm adapted for modeling brief, transient cortical spectral responses to sound, and a recently developed method for estimating effective connectivity using multivariate autoregressive modeling to measure brief event-related changes in multichannel functional interactions. A semi-automated spatial normalization method for comparing intracranial electrode locations across patients is also described. The individual methods presented are published and readily accessible. We discuss the benefits of integrating multiple complementary methods in a unified and comprehensive quantitative approach. Methodological considerations in the analysis of multichannel ECoG data, including corrections for multiple comparisons are discussed, as well as remaining challenges in the development of new statistical approaches.

Keywords: electrocorticography, auditory, matching pursuit, multivariate autoregressive modeling, epilepsy, statistical testing

INTRODUCTION

Multichannel intracranial recordings are used increasingly to investigate the functional organization of human cortex. Intracranial EEG recordings, known as electrocorticography (ECoG), use electrodes implanted for clinical purposes, including seizure localization and surgical planning for treatment of intractable epilepsy. This clinical circumstance provides a rare opportunity to record focal neuronal population activity directly from cortex. ECoG recordings offer excellent temporal resolution (1 ms) and the proximity of recording electrodes to underlying cortical sources enhances spatial resolution, signal-to-noise ratio, and sensitivity to a broad range of EEG frequencies. Recent ECoG studies have investigated cortical sensory (auditory, visual), motor, language, and cognitive systems (Crone et al., 2006, 2009; Miller et al., 2007; Brugge et al., 2009; Jacobs and Kahana, 2009; Sinai et al., 2009). ECoG recordings are usually obtained from large numbers of electrodes, yielding high dimensional data sets.

In this methods paper, we propose a novel quantitative framework that integrates multiple existing methods for analyzing high dimensional ECoG data sets. This quantitative approach is used to determine the statistical significance of event-related responses in intracranial recordings. We adapted this quantitative approach for auditory ECoG studies conducted at our epilepsy center. ECoG is useful for investigating the functional organization of auditory cortex and is used clinically for pre-surgical functional mapping and, more recently, for brain–computer interfaces (Howard et al., 2000; Lachaux et al., 2007; Brugge et al., 2009; Hong et al., 2009; Sinai et al., 2009). We will examine how complementary methods can be combined to evaluate statistically significant changes in multiple aspects of auditory event-related ECoG activity: evoked responses, spectral responses, event-related (causal) connectivity, and spatial distribution (normalization). Each method is illustrated with examples from recent auditory ECoG studies. We begin with a brief overview of intracranial recording methods and cortical auditory event-related responses. We discuss the advantages of using multiple complementary methods (e.g., single-channel and multichannel) to analyze the same ECoG data sets. Methodological issues, including multiple comparisons, as well as future directions for development of new statistical approaches are also discussed.

INTRACRANIAL RECORDING METHODS

RECORDING ELECTRODES

Intracranial recordings are obtained with subdural or stereotactic depth electrodes. Subdural electrodes are positioned on the lateral surface of cortex; depth electrodes are inserted through cortex to record from deeper structures, such as hippocampus. Recordings can be obtained intraoperatively by moving electrodes to different locations on the exposed cortex, or extraoperatively by leaving implanted electrodes indwelling for up to 10 days of monitoring. Although microelectrodes have been used for single-neuron recording studies (Howard et al., 1996; Schwartz et al., 2000; Ojemann et al., 2002; Gelbard-Sagiv et al., 2008), most clinical centers use macroelectrodes to record from neuronal populations. At our center, intracranial auditory recordings are usually obtained extraoperatively with subdural and depth macroelectrodes (Crone et al., 2001a; Boatman and Miglioretti, 2005; Sinai et al., 2009).

ELECTRODE PLACEMENT

For extraoperative recordings, subdural and depth electrodes are implanted by craniotomy under general anesthesia. Typically, electrodes are implanted over one hemisphere where the seizure focus is suspected based on clinical data. Subdural electrodes consist largely of platinum-iridium disks, 2–3 mm in diameter, spaced 5–10 mm apart center-to-center and embedded in 1.5-mm-thick arrays of medical grade silastic. Common subdural electrode array configurations are 4×5, 6×8 or 8×8 grids and 1×8 or 2×8 strips (**Figure 1**). Most patients have multiple grids and/or strips implanted to ensure adequate spatial sampling for seizure localization, with the total number of recording electrodes per patient at our center typically between 48 and 184 (maximum to date). Depth electrodes are one-dimensional arrays typically of 1×4 or 1×8 contacts, 2-mm in diameter, inserted through gyri or sulci. While subdural electrodes record primarily from gyral structures because they are located over the cortical surface, depth electrodes can record from both gyral and sulcal structures. Intracranial recordings can also be made from electrodes implanted in the epidural space. Although signal amplitude is reduced by the dura mater and epidural electrodes cannot cover as many brain areas as can subdural electrodes, they potentially can be implanted less invasively and can offer an important alternative to subdural electrodes in patients with severe subdural adhesions from prior surgery. These electrodes have been used for presurgical evaluations for intractable epilepsy, and their use has also been considered for brain–machine interface applications (Barnett et al., 1990; Blount et al., 2008; Slutzky et al., 2008). In this paper, we focus on intracranial recordings using subdural electrodes.

Intracranial electrode configuration and placement are determined individually, based on each patient's clinical circumstances. Many of our patients have subdural electrode coverage of the superior temporal gyrus corresponding to auditory cortex (Boatman et al., 2000; Miglioretti and Boatman, 2003; Boatman, 2004, 2006; Boatman and Miglioretti, 2005). During implantation, electrode grids and strips are sutured to the overlying dura to prevent movement during closure of the craniotomy. Post-implantation CT scans are obtained the following day to confirm electrode locations. After implantation, patients are monitored in the neurological critical care unit and then admitted to the epilepsy monitoring unit for ECoG recordings and monitoring. Auditory ECoG studies are usually initiated 3–5 days after electrode implantation, when surgery-related edema and discomfort are reduced. All patient research participants provide informed written consent in compliance with our Institutional Review Board. Patients are tested individually at bedside in private rooms with measured ambient noise levels ≤45 dB SPL. Recordings are made from the same subdural electrodes used clinically for seizure localization; they introduce no additional risk to the patient and do not interfere with the ongoing clinical video recordings of patients' seizures for localization of the epileptic zone. Once sufficient clinical information has been obtained, patients return to surgery for removal of the electrodes and possible resection for treatment of seizures.

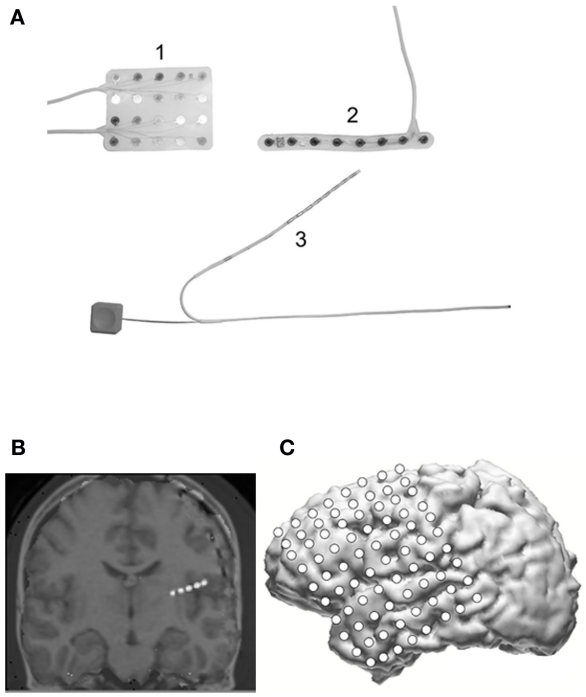


FIGURE 1 | (A) Shows three different electrode configurations used for intracranial recordings: (1) a 4×5 subdural electrode array (Ad-Tech, Racine, WI, USA); (2) a 1×8 electrode strip; and (3) an 8-contact depth electrode. **(B)** A coronal 3D MRI reconstruction showing depth electrodes (white) in right temporal lobe; **(C)** a 3D MRI reconstruction with 8×8 subdural electrode array and additional 2×8 strip implanted over the lateral left hemisphere.

RECORDING PARAMETERS

ECoG recodings are obtained using standard clinical parameters. The ECoG signal is amplified (Schwarzer amplifier) at a channel gain of 1408 and recorded digitally from all channels (Stellate Systems Inc.) at a minimum sampling rate of 1000 Hz with a bandwidth of 0.1–350 Hz (6 dB/octave). We routinely use a referential montage in which all subdural electrodes are referenced to a single intracranial electrode. A benefit of referential recordings is that they can be remontaged readily for analysis, in contrast to bipolar recordings (see Signal Pre-processing). Ideally, the electrode

selected for the reference has minimal electrical artifact and electrographic abnormalities and is distal to the recording region of interest (e.g., superior temporal gyrus). Although it is not possible to have an entirely inactive reference, choosing a reference distal to recording sites of interest can reduce its potential contributions and the need for spatial reformatting. Extracranial reference electrodes are not used to avoid contamination by muscle activity that has prominent spectral energy in high frequencies (e.g., gamma band). Markers for stimulus onset times are recorded simultaneously to ECoG marker channels.

SIGNAL PRE-PROCESSING

The continuous ECoG signal is pre-processed for event-related analysis. Pre-processing is performed to identify and exclude from analysis channels and trials with artifact and to remontage the recording data. The continuous ECoG recording is first inspected visually to identify and reject channels with excessive artifact or epileptiform activity. The continuous ECoG signal is then segmented into individual trials containing pre-stimulus baseline and post-stimulus intervals; these are then visually inspected to reject trials with artifact or epileptiform activity. Review by an epileptologist or clinical neurophysiologist is helpful to ensure correct identification of channels and trials with artifact. At our center, an epileptologist also routinely inspects the intracranial EEG prior to recording to rule out the presence of excessive spiking or epileptiform activity that can reduce the quality of the recordings. Once channels and trials with artifact have been excluded, the remaining channels can be remontaged for event-related analysis.

For our ECoG studies, we remontage to a common average reference (Sinai et al., 2005, 2009). For each sample of the ECoG signal in a given trial, an average of the voltages in all channels, excluding those with artifact or frequent epileptic discharges, is subtracted from the voltage in each individual channel. This spatial reformatting reduces variations in signal amplitude across the recording array that result from differences in distance between active electrodes and the reference electrode. Although this procedure is sometimes used in scalp EEG studies to approximate a neutral reference, this cannot be assumed, particularly in the case of intracranial recordings. The choice of reference electrode should be considered carefully so that noise and other prominent electrical activity are not inadvertently introduced into the signal. More complex reformatting, such as a Laplacian or local average reference, are not usually performed in ECoG studies. This is because they are difficult to implement with intracranial arrays, requiring exclusion of edge electrodes or use of a spline to approximate sites off the array. Moreover, these procedures were originally developed for scalp EEGs to approximate local sources, effectively functioning as high-pass spatial filters and, therefore, may not be necessary or appropriate for ECoG recordings. Although volume conduction from distant sources (i.e., far field potentials) can occur, signals recorded with intracranial electrodes are dominated by local sources within a few millimeters of the contacts such that signal features in adjacent electrodes are often very different (Crone et al., 2001a; Sinai et al., 2005, 2009). Nevertheless, these procedures, as well as simpler alternatives (bipolar derivations), are important alternatives to consider, particularly when volume conduction from distant sources is suspected. Additional signal pre-processing for multichannel connectivity analyses is described separately below (see Signal Pre-processing for ERC Analysis).

CORTICAL AUDITORY EVENT-RELATED RESPONSES

Cortical auditory event-related responses are electrophysiology-based measures of neural activity generated, in response to sound, by neural sources in primary and non-primary auditory cortex located in the superior temporal gyrus of both cerebral hemispheres. We will focus on three types of cortical auditory event-related activity in ECoG signals: evoked responses; spectral (induced) responses; and multichannel event-related connectivity.

EVOKED AUDITORY RESPONSES

Evoked responses, also known as evoked potentials or ERPs, are synchronized, low-voltage, typically low-frequency (<50 Hz) electrical signals with latencies and amplitudes phase-locked to a stimulus. Because of their low amplitude, trial averaging in the time domain is used to extract evoked responses and identify individual component peaks (positive, negative). One of the earliest and largest cortical evoked responses is the vertex-negative N1 that peaks in adults around 75–120 ms after stimulus onset and is an automatic, transient response to sound onset or change, with generators in primary and non-primary auditory cortex (Scherg and von Cramon, 1986; Naatanen and Picton, 1987; Godey et al., 2001). The N1 is embedded between two positive peaks – the P1 and the P2– forming a three-component evoked complex known as the P1-N1-P2. Later cortical auditory evoked responses include the N2, occurring approximately 200 ms post-stimulus onset (Halgren et al., 1998; Hong et al., 2009); the mismatch negativity that reflects pre-attentive detection of stimulus differences (Tiitinen et al., 1994; Naatanen, 2001; Naatanen et al., 2007); and the P3 (or P300) response that has been investigated extensively in studies of auditory attention and other higher level cognitive and language functions (Knight et al., 1989; Polich and Kok, 1995). A variety of auditory stimuli can be used to elicit cortical event-related responses, ranging from simple sinusoidal tones to complex speech (Sinai et al., 2009). Similarly, a number of different paradigms can be used to elicit cortical auditory event-related responses including passive listening tasks and active discrimination tasks (Crone et al., 2001b; Sinai et al., 2009). The choice of stimulus and paradigm is determined largely by the research hypothesis to be tested. Dependent variables in auditory ERP studies include peak latency (ms) and amplitude (dB).

SPECTRAL RESPONSES

It is well established that auditory stimuli also induce event-related changes in ECoG spectral power that are not phase-locked to the stimulus (Crone et al., 2001a; Edwards et al., 2005; Lachaux et al., 2007; Sinai et al., 2009). A variety of induced spectral responses, once considered 'noise' in the analysis of evoked responses, are now associated with perceptual and cognitive processing. Because spectral responses are not phase-locked to a stimulus, they are not evident in the averaged evoked waveform. To identify event-related spectral power changes, time–frequency analyses are used for averaging in the frequency domain rather than in the time domain. A number of different time–frequency methods are used to measure event-related changes in the ECoG spectrum, including short-time Fourier

transform, wavelet transform, and matching pursuit (MP) (Mallat and Zhang, 1993). Scalp recording studies have associated changes in a variety of EEG frequency bands with task-related cortical processing, including increases and decreases in theta (4–7 Hz), alpha (8–13 Hz), and beta (14–20 Hz) oscillations under different functional task conditions (Klimesch et al., 1993; Neuper and Pfurtscheller, 2001; Jensen and Tesche, 2002; Struber and Herrmann, 2002). Previous studies have also identified higher frequencies, including gamma (≥30 Hz), as potential indices of task-related cortical processing (Crone et al., 1998; Tallon-Baudry and Bertrand, 1999; Edwards et al., 2005; Sinai et al., 2005; Lachaux et al., 2007). Event-related gamma activity has been associated with auditory, visual, and motor functions (Pantev et al., 1995; Tallon-Baudry and Bertrand, 1999; Pfurtscheller et al., 2003; Sinai et al., 2009). The same stimuli and experimental paradigms used to elicit cortical auditory evoked responses are used to induce changes in spectral power. Modulation of spectral intensity is measured in units of natural log power change.

EVENT-RELATED CONNECTIVITY

Recent advances in signal processing have engendered investigations of event-related functional interactions in the cortical networks associated with sensory, motor, cognitive, and language functions. Two main types of functional network interactions are recognized: functional connectivity and effective connectivity. Functional connectivity is defined as the temporal relations (coherences) between distant cortical regions, without reference to their directionality (causality). Effective connectivity refers to the causal interactions of cortical networks (Friston et al., 1994; Astolfi et al., 2004; Sporns et al., 2007). A number of multichannel analysis methods have been developed to probe the dynamic interactions of auditory and other cortical functional networks, including Granger causality (Oya et al., 2007; Gow et al., 2009), dynamic causal modeling (Friston et al., 2005; David et al., 2006; Garrido et al., 2007), independent component analysis (Onton et al., 2006), direct Directed Transfer Function (dDTF) (Korzeniewska et al., 2003), Short-time Directed Transfer Function (SDTF) (Ginter et al., 2001, 2005), and more recently Short-time direct Directed Transfer Function (SdDTF) (Korzeniewska et al., 2008). We will focus on the SdDTF method which was developed at our center for evaluating multichannel causal interactions over brief periods (milliseconds) and is well suited for studying cortical sound processing. SdDTF originates from directed transfer function (DTF) (Kaminski and Blinowska, 1991; Franaszczuk et al., 1994), which is based on the concept of Granger causality. SdDTF uses multiple trials/repetitions (multiple realizations of the same stochastic process) to measure the dynamics of event-related functional interactions between cortical sites, using short time windows (Ding et al., 2000).

ANALYSIS OF AUDITORY EVOKED RESPONSES

TIME-DOMAIN AVERAGING

Auditory evoked responses are derived by averaging in the time domain because their latencies and amplitudes are time- and phase-locked to the stimulus. In contrast, background electrophysiological activity is not phase-locked and, therefore, is reduced by phase cancellation. The main goal of most clinical studies is to identify the largest evoked response for measurement (amplitude, latency). The largest response is often identified visually, without statistical testing. Existing analysis methods, such as independent component analysis, were developed to address the poor spatial resolution of scalp recordings through advanced source localization and signal de-noising (Makeig et al., 1997; Delorme and Makeig, 2004). Because the spatial resolution of intracranial recordings is considerably better than that of scalp recordings, these methods may not be necessary or applicable.

Recent ECoG studies have begun using statistical testing to compare event-related responses to the baseline signal (Edwards et al., 2005; Towle et al., 2008; Sinai et al., 2009). This is useful for verifying waveform detection and for reducing potential biases associated with reliance on visual identification. This approach is also helpful for determining the spatial distribution of cortical evoked responses associated with different experimental conditions (Lachaux et al., 2007; Towle et al., 2008; Sinai et al., 2009). Although there is no standard method for measuring baseline ECoG, the two most common approaches are computing the mean amplitude, based on a random sample of a fixed number of time points, and resampling of the time-series (Edwards et al., 2005; Sinai et al., 2009). Differences in response latency and amplitude can also be measured as a function of experimental conditions (stimulus, task) using linear regression with generalized estimating equations to account for correlation within subjects, as previously described (Liang and Zeger, 1986; Boatman and Miglioretti, 2005). Comparing the timing and size of evoked responses across multiple channels provides useful information on the spatial–temporal profiles of auditory cortical responses. However, performing multiple comparisons also increases the likelihood for false rejections. To address this issue, correction methods such as the Bonferroni method and false discovery rate (FDR) are increasingly used in ECoG studies. We discuss the problem of multiple comparisons and correction methods in more detail below (see Multiple Comparisons).

SPECTRAL ANALYSIS

To quantify event-related changes in spectral composition, the ECoG signal is segmented into temporal epochs and transformed to the frequency domain for averaging across experimental trials. There are a number of different algorithms for converting the signal into the frequency domain, including discrete Fourier transforms, wavelets, and complex demodulation. Each method offers trade-offs between time and frequency resolution on the one hand and computational transparency and efficiency on the other. We use a time–frequency MP algorithm (Mallat and Zhang, 1993; Franaszczuk et al., 1998; Durka et al., 2001; Ray et al., 2003). MP is an iterative algorithm for adaptive time–frequency estimates of signal power. The MP method is well suited for analysis of non-stationary changes in the ECoG signal, and combines advantages of other time–frequency decomposition approaches – including short-time Fourier transform and wavelet transform – with enhanced time–frequency resolution, as demonstrated previously (Ray et al., 2003; Sinai et al., 2005, 2009). The MP method is implemented in C, based on the original software (Mallat and Zhang, 1993), and runs under Linux on a cluster of computer nodes (software program available upon request).

Spectral analyses of event-related electrophysiological responses often distinguish between phase-locked and non-phase-locked signal components. Phase-locked components are obtained by

averaging across trials in the time domain, yielding traditional evoked potentials. When signals are averaged in the frequency domain, the resulting time–frequency averages include both phase-locked and non-phase-locked components. A variety of approaches can be used to isolate the phase-locked components in order to emphasize the non-phase-locked components of electrophysiological responses. The efficacy of these approaches depends largely on the validity of the phase-locked components, which are themselves somewhat of a methodological construct. A simple, but arguably simplistic, way to try to isolate non-phased-locked components is to subtract the time-domain-averaged signal (i.e., evoked potential) from each trial prior to averaging in the frequency domain (Crone et al., 2001b). A similar approach is that of computing the inter-trial variance (Kalcher and Pfurtscheller, 1995). Because the amplitudes of phase-locked components are typically much smaller than the ongoing raw signal, their contributions to the spectral analysis results are likely to be small. However, large inter-trial variation in the amplitude and latency of the evoked potential can introduce spurious energies when subtracted from the raw signal (Truccolo et al., 2002). More advanced methods, including single-trial time–frequency analyses, may reduce the need for this procedure in the future. For now, the best approach may be to perform spectral analyses of signals with and without isolated phase-locked components, and in combination with time–frequency decomposition of the time-averaged evoked response itself (Trautner et al., 2006). While this approach can help to elucidate the contributions of phase-locked and non-phase-locked components, it is important to recognize that inherent methodological limitations remain.

MATCHING PURSUIT

The MP method decomposes the ECoG signal into a linear combination of time–frequency functions termed 'atoms', drawn from a large dictionary of functions well localized in the time–frequency plane. We implement the MP method using a dictionary of sine functions that have well-defined frequencies; Dirac delta functions that are localized in time; and sine-modulated Gaussians – Gabor functions. Gabor functions are characterized by the highest combined time–frequency resolution based on the uncertainty principle in time–frequency analysis that states that $\sigma_f\sigma_t \geq 1/2$ where σ_f and σ_t represent spread of the function in frequency and time, respectively. It can be shown that equality is achieved only for Gabor functions (i.e., modulated Gaussian functions) (Mallat and Zhang, 1993). The atom representing the maximum energy of the signal (i.e., the largest inner product with the signal) is selected first; atoms in the dictionary representing the maximum energy of the residual are then determined iteratively. After M-th iteration the signal $f(n)$ is expressed as:

$$f(n) = \sum_{m=0}^{M-1} \left\langle R^m f, g_m \right\rangle g_m(n) + R^M f(n), \quad (1)$$

where $R^m f$ is the residual after the m-th iteration, g_m is the atom selected in m-th iteration, n is the digitized signal sample number, and $\langle R^m f, g_m \rangle$ denotes the inner product of residual $R^m f$ and atom g_m. The time–frequency energy distribution is then computed by summing the Wigner–Ville distribution of the Gabor atoms expressed as:

$$E_M f(n,k) = \sum_{m=0}^{M-1} \left| \left\langle R^m f, g_m \right\rangle \right|^2 E_V g_m(n,k), \quad (2)$$

where $E_M f(n,k)$ represents energy of signal f in discrete time n and discrete frequency k after M steps of iteration, and $E_V g_m(n,k)$ represents the Wigner–Ville distribution of atom g_m. $E_V g_m(n,k)$ is represented as an ellipsoid in two-dimensional time–frequency plane for Gabor atoms, as a horizontal line for sines, and as a vertical line for Dirac deltas. The presence of electrical artifact (line noise) in the recordings is represented in the decomposition by sine or Gabor atoms, with a central frequency around 60 Hz in the United States (50 Hz in Europe) and its harmonics, and is typically excluded from summation in energy computation.

Time–frequency decomposition is performed for each trial separately. The lengths of the pre-stimulus and post-stimulus epochs are determined largely by the parameters (e.g., inter-stimulus intervals) of the experimental recording paradigm. For each frequency, the baseline (pre-stimulus) power is computed by averaging over all baseline time points within a trial and over all trials. To test the null hypothesis that event-related spectral power changes do not differ from baseline, estimates of pre-stimulus (baseline) and post-stimulus spectral power in each post-stimulus time point are compared. We use a logarithmic transformation and the Student's t-test to assess statistical significance of the differences (Zygierewicz et al., 2005). The t statistic is computed as:

$$t_{n,k} = \frac{\bar{E}[n,k] - \bar{B}[k]}{s_E}, \quad (3)$$

where $\bar{E}(n,k)$ is the average of $\log[E_M(n,k)]$ for post-stimulus time-points n and frequency k over N trials, $\bar{B}(k)$ is the average of $\log[E_M(j,k)]$ over baseline points j and N trials, and s_E is a weighted estimator of the standard deviation. The statistics $t_{n,k}$ follow Student's t distribution with $N(1 + K) - 2$ degrees of freedom, where N is the number of trials and K is the number of baseline time points.

Figure 2 shows results of single-channel MP analysis of spectral responses to two different stimuli (speech, tones) recorded from the same lateral temporal lobe site in one patient. Of interest is the observation that both simple tones and complex speech stimuli induced high frequency (gamma) spectral responses at sites in non-primary auditory association cortex. This finding challenges the traditional view that non-primary auditory areas are involved only in processing complex sounds.

To correct for multiple within-subject comparisons, the Bonferroni correction or FDR is applied, as discussed below (see Multiple Comparisons) and as previously described (Zygierewicz et al., 2005; Sinai et al., 2009). The resulting time–frequency energy distribution reflects the magnitude and statistical significance of energy changes over time. Time–frequency points (pixels) representing statistically significant changes from baseline can also be plotted across the frequency range by experimental conditions (stimulus, task), as shown in **Figure 2**.

Quantifying differences in spectral responses across experimental conditions (stimulus, task) poses additional challenges. Simple comparisons of spectral responses in the same time–frequency pixel in different experimental conditions can be readily performed by t-test (**Figure 3**). However, these parametric tests do not capture visible differences in the relative size, morphology (shape) and timing of two (or more) spectral responses, as seen in **Figure 3**. Quantifying these differences in spectral responses will require new statistical approaches.

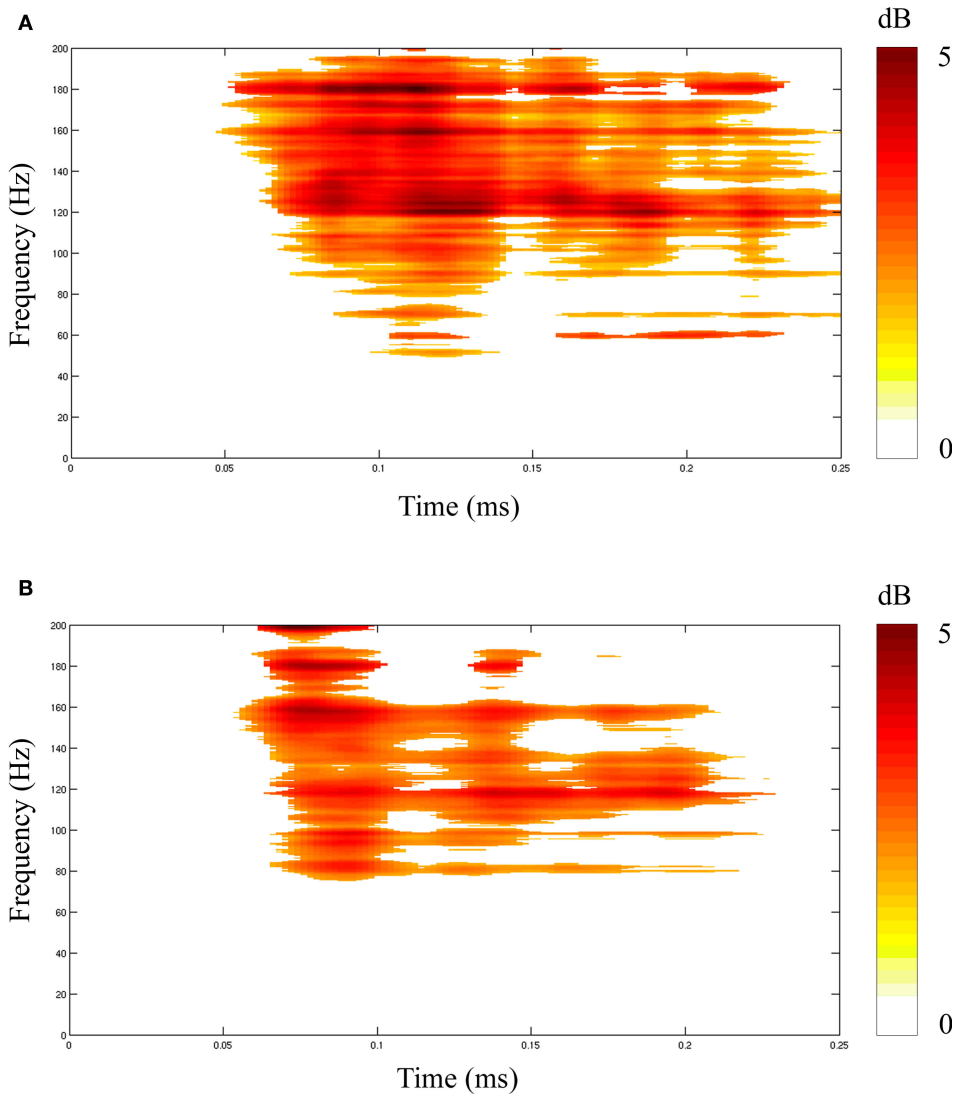


FIGURE 2 | Results of matching pursuit analysis showing statistically significant spectral power changes (0–200 Hz) occurring 0–250 ms following onset of speech (A) and tone (B) stimuli and recorded from the same electrode site located on the convexity of the posterior superior temporal gyrus in one patient. Scale bar to right of each panel shows color scale of spectral power changes (dB). Largest spectral power changes are shown in red.

METHODOLOGICAL CONSIDERATIONS

The length of pre-stimulus (baseline) and post-stimulus epochs is determined in part by the experimental protocol and, in particular, by the inter-stimulus interval. For our auditory ECoG studies, we use relatively short inter-stimulus intervals (~1–2 s) which allow us to record larger numbers of trials. However, using shorter time windows can make it more difficult to detect power changes in lower frequencies (alpha, beta).

The MP method is useful for studying non-phase-locked, event-related changes in ECoG signals. This approach is well suited for capturing the brief (milliseconds), rapidly changing neural responses characteristic of cortical sound processing. We have used this approach in our recent studies to characterize spectral responses to different auditory stimuli (tones, speech) in auditory association cortex (Ray et al., 2003; Sinai et al., 2009). Studies from our center and others have shown that spectral and time-domain analyses are complementary, each providing important clinical information and new insights into the functional organization of the human cortical auditory system (Crone et al., 2001a; Edwards et al., 2005; Lachaux et al., 2007; Towle et al., 2008; Sinai et al., 2009).

EVENT-RELATED CAUSAL AND EFFECTIVE CONNECTIVITY

Previous intracranial auditory recording studies have identified statistically significant event-related changes in ECoG spectral power using single-channel time–frequency analyses (Crone et al., 2001a; Lachaux et al., 2007; Sinai et al., 2009). Non-phased-locked changes in spectral power, once considered 'noise', are now thought to be neural indices of regional and distributed cortical processing, providing a useful tool for probing the functional organization of cortical networks (Engel and Singer, 2001; Singer, 1993). As a

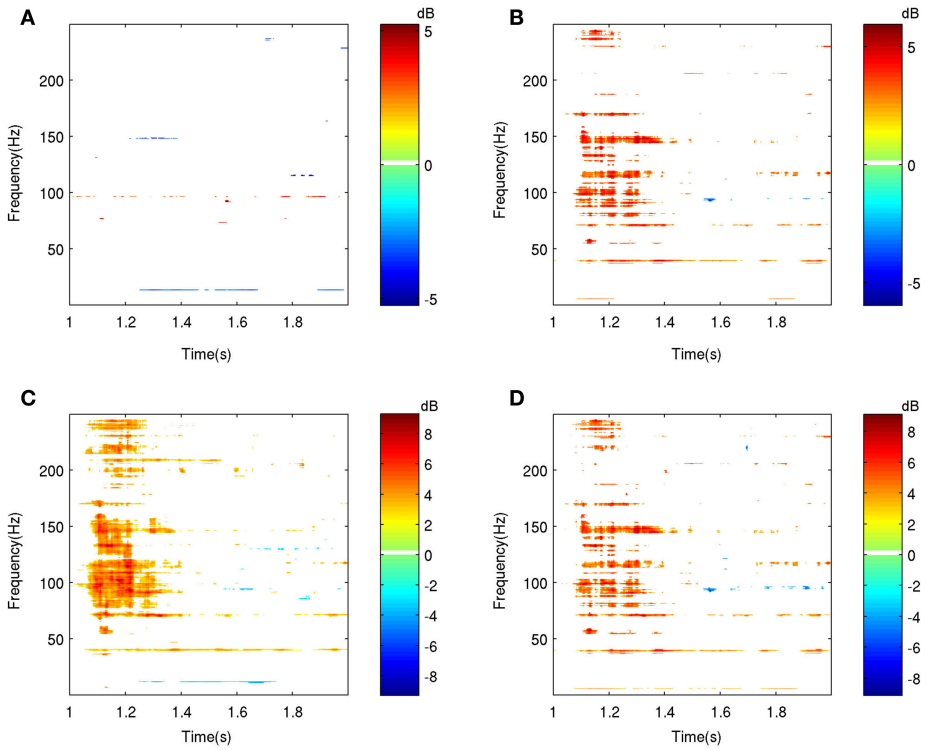


FIGURE 3 | Statistical comparisons of event-related spectral power responses (dB) elicited with two different auditory stimuli (tones, speech) from the same ECoG channel (electrode) on the lateral superior temporal gyrus. Spectral responses that differed significantly from their respective baselines are shown for **(A)** tones and **(B)** speech. Comparisons of two spectral responses (**A** versus **B**) showed significant differences when the response to tones was subtracted from the response to speech, as shown in **(C,D)**. Results of two statistical comparisons are shown for: a *t*-test assuming unequal variances **(C)** and a paired *t*-test for *Z*-scores **(D)**. False discovery rate correction for multiple comparisons was applied for both tests. The similarity of test results indicates similar variances, likely reflecting the common data sources (channel,subject).

result, there is now considerable interest in investigating effective neural connectivity based on the dynamic patterns of event-related propagation of the non-phase locked activity. Event-related causality (ERC) is a new method for measuring event-related changes in causal interactions between multi-electrode recording sites, to estimate the effective connectivity of cortical networks engaged by functional tasks (Korzeniewska et al., 2008). The ERC method measures statistically significant event-related changes in the direction, strength, and spectral content of direct electrophysiological interactions between brain sites and their timing. In the following sections, we describe the multichannel ERC method and its application to ECoG data, including auditory event-related recordings.

EVENT-RELATED CAUSALITY

ERC is based on the concept of Granger causality, which was originally developed for economic modeling and predictions (Granger, 1969). Granger causality postulates that an observed time series $x_k(t)$ causes another time series $x_l(t)$ if knowledge of $x_k(t)$'s past significantly improves prediction of $x_l(t)$. This approach was implemented in multiple time series by fitting a multivariate autoregressive (MVAR) model, and has been used recently to study the dynamics of causal interactions between neural populations for signals assumed to be either stationary (Brovelli et al., 2005; Krichmar et al., 2005; Seth, 2005; Cadotte et al., 2008, 2009; Anderson et al., 2009; Keil et al., 2009), or non-stationary (Freiwald et al., 1999; Hesse et al., 2003). By using frequency decomposition of Granger's time domain (Geweke, 1982) it is possible to examine spectral properties of Granger causality (sometimes referred as Granger–Geweke causality), which is useful for neurophysiological signals, where frequency domain is often of interest. The Granger causality technique is a 'model-free' measure of causal interactions in that it is not based on *a priori* assumptions about anatomical or functional connections. However, it is based on a statistical linear model and cannot describe non-linear causal interactions. The concept of Granger causality led to development of multiple related methods, including structural analysis (Bernasconi and Konig, 1999); partial directed coherence (Sameshima and Baccala, 1999; Baccala and Sameshima, 2001a,b; Schelter et al., 2006); and DTF (Kaminski and Blinowska, 1991; Franaszczuk et al., 1994; Kaminski et al., 2001; Astolfi et al., 2005; Kaminski and Liang, 2005). A number of these methods have been compared previously (Kus et al., 2004; Eichler, 2005; Winterhalder et al., 2005; Schlogl and Supp, 2006; Astolfi et al., 2007b). In particular, a study by Kaminski et al. (2001) showed equivalence of DTF and bivariate Granger causality. Other methods that are not based on the Granger causality concept have also been used to determine functional connectivity, including calculations of evoked potential covariances (Gevins et al., 1995); adaptive phase estimation (Schack et al., 1999); effective information (Tononi and

Sporns, 2003); the imaginary part of coherency (Nolte et al., 2004); and directed information transfer (Hinrichs et al., 2008). In this paper we will discuss only the SdDTF method, which is a modification of the DTF method and therefore also a linear Granger-like causality measure.

The MVAR model assumes that the values of multiple time series from K recording sites/channels – vector $\vec{x} = \{x_1,...,x_K\}$ – at time t, depend on p previous values of the time series, and the random components vector $\vec{e}$. When the MVAR model is fitted to ECoG signals from K channels, they are treated as one multivariate stochastic process, expressed as:

$$\vec{x}(t) = -\sum_{j=1}^{p} A_j \vec{x}(t-j) + \vec{e}(t), \tag{4}$$

where A_j is a $K \times K$ MVAR coefficients matrix and p is the model order. To determine the value of model order p, the Akaike Information Criterion is applied (Akaike, 1974). The MVAR model coefficients were computed using a Yule–Walker algorithm implemented in C (Franaszczuk et al., 1985). Because ECoG activity may be understood in terms of rhythms and oscillations, it is useful to describe the spectral properties of their signals. For this purpose the MVAR equation may be transformed to the frequency domain (Marple, 1987) as:

$$X(f) = H(f)E(f), \tag{5}$$

where

$$H(f) = \left(\sum_{j=0}^{p} A_j e^{-i2\pi j f \Delta t} \right)^{-1}, \tag{6}$$

$H(f)$ is the transfer function of the multichannel system, f is frequency, and Δt is the sampling interval. The element h_{kl} of the matrix $H(f)$ describes the transfer function between the k-th output and the l-th input of the system. If the element h_{kl} of H is equal to 0, the hypothesis that $x_k(t)$ causes $x_l(t)$ can be rejected. The matrix is not symmetric if any of the channel pairs (k,l) have unequal flows in both directions. As such, the directional properties of a multichannel system may be interpreted as Granger causal relationships, signal flows, or activity transfers. If H is symmetric, directionality cannot be determined. The direct transfer function was developed as a normalized version of H matrix (Kaminski and Blinowska, 1991; Franaszczuk et al., 1994; Kaminski et al., 2001; Astolfi et al., 2005; Kaminski and Liang, 2005). The DTF method has also been used to study activity flow in amnesic and Alzheimer's patients (Babiloni et al., 2009); Parkinson's patients (Androulidakis et al., 2008; Lalo et al., 2008); and spinal cord injury patients (Astolfi et al., 2006), and in studies of seizure onset and neural circuitry (Franaszczuk et al., 1994; Franaszczuk and Bergey, 1998; Ge et al., 2007); wake-sleep transitions (De Gennaro et al., 2004, 2005); working memory (Edin et al., 2007); memory encoding and retrieval (Babiloni et al., 2006); and animal behavior (Korzeniewska et al., 1997). Recently, DTF and related methods have also been used to investigate causal influences in functional MRI (fMRI) data (Deshpande et al., 2006, 2008; Hinrichs et al., 2006; Sato et al., 2008; Wilke et al., 2009), and to develop brain computer interfaces (Shoker et al., 2005).

To capture the dynamics of ERC, various modifications of MVAR model fitting can be applied (Astolfi et al., 2007a, b; Wilke et al., 2007). The SDTF (Ding et al., 2000), a modification of the DTF method, uses short, overlapping time windows that are shifted along the signals when there are multiple task repetitions (considered as a realization of the same stochastic process), or trials, to track brief changes in the flow of activity between brain regions (Ginter et al., 2001, 2005; Kaminski et al., 2005; Kus et al., 2006, 2008; Philiastides and Sajda, 2006; Korzeniewska et al., 2008).

Granger causality and DTF methods identify both direct and indirect relationships between signals. For example, for three signals related as follows: x1 → x2 → x3, these methods will show not only flows x1 → x2 and x2 → x3 but also x1 → x3 (indirect flow). To detect only direct relationships, a partial coherence function can be utilized. By multiplying this function with DTF, the dDTF is obtained which describes only direct flows (Korzeniewska et al., 2003). However, the partial coherence function can also yield spurious relationships, as when two non correlated signals are added to form a third signal. This will result in spurious partial coherence between to non correlated signals: the so called 'marrying parents of a joint child effect' (Schelter et al., 2006). However, in this case DTF will show no flow and dDTF will avoid the spurious effect. The recently developed SdDTF method involves a synthesis of both the SDTF and the dDTF collectively (Korzeniewska et al., 2008), in the form:

$$\zeta_{k,l} = \frac{|h_{k,l}(f)||\chi_{k,l}(f)|}{\sqrt{\sum_f \sum_{k,l} |h_{k,l}(f)|^2 |\chi_{k,l}(f)|^2}}, \tag{7}$$

where χ_{kl} are elements of partial coherence matrix. The SdDTF function determines whether a signal component at a given frequency in channel k is shifted in time with respect to a signal component of the same frequency in channel l, and whether the shifted components are coherent and are not explained by components of other channels. SdDTF takes values from 0 to 1. Zero indicates a lack of direct causal relationships. The non-zero values of SdDTF are interpreted as a flow of activity from one channel to another, that is, $\zeta_{kl}(f) > 0$ indicates flow of activity from channel l to channel k(l → k). The temporal evolution of causality estimates can then be obtained by calculating them in a short window that is shifted along the signal of interest, as previously described (Korzeniewska et al., 2008).

The interpretation of event-related causal interactions is constrained by the available measurements. As in all scientific inference, missing information can lead to false interpretation. In multichannel analyses, it is important to include measurements from all brain regions that are responsible for the analyzed task. When neural networks associated with functional processing are only partially represented, spurious causalities may result (Eichler, 2005; Krichmar et al., 2005). Removing, adding or replacing crucial recording sites from the analysis is most likely to produce artificial causalities (Eichler, 2005). Conversely, inclusion, deletion, or replacement of recording sites that are not crucial for the analyzed system may not substantially change the patterns of causal interactions (Korzeniewska et al., 2008). This issue can be addressed by using approaches like partial directed coherence, dDTF, and SdDTF – all of which emphasize direct flows or interactions – and

by increasing the number of channels. Nevertheless, it is important to have relatively comprehensive coverage of regions known to be functionally important, such as the superior temporal gyrus, for studying auditory processing. This is also illustrated in the application of SdDTF to auditory ECoG data described below (see Estimating ERC in auditory event-related ECoG): the patient had multiple electrodes covering the superior temporal gyrus and recording sites selected for inclusion were identified based on previous analyses of auditory event-related power spectra.

In drawing conclusions from these analyses, several limitations warrant consideration. As in any scientific investigation, we are limited to the set of recorded signals and these could be influenced by other processes not detected in the analysis. For example in the network x1 → x2 → x3, a fourth undetected process could be involved such that x1 → x4 → x3. This limitation underscores the importance of carefully choosing recording sites for analysis and ensuring adequate representation of all regions associated with the function under investigation. The second limitation of methods based on Granger causality, including dDTF, is the inability to correctly identify cyclical interactions (for an excellent discussion of these issues, see Eichler, 2006).

Causal interactions can be both linear and non-linear in brain systems. Previous studies have suggested that non-linear mechanisms may play an important role in the functional connectivity of large-scale neural networks (Friston, 1997; Schanze and Eckhorn, 1997; Bekisz and Wrobel, 1999; Breakspear and Terry, 2002a,b; Senkowski et al., 2007). ERC is a linear method and does not provide information about the nature of the causality (linear or non-linear). Nevertheless, linear methods may be sensitive to both linear and non-linear causal interactions (Freiwald et al., 1999; Chavez et al., 2003; Gourevitch et al., 2006). Indeed, MVAR models can be used to describe non-linear systems (Franaszczuk and Bergey, 1999). The detection of dependencies by linear methods does not require that those dependencies are linear (Freiwald et al., 1999). Thus, the ERC method cannot determine if the observed activity flow changes are due to linear or non-linear dynamics. However, it has been shown that higher-degree non-linearity models do not provide a clear advantage over linear ones (Barbero et al., 2009). A recent study using a non-linear Granger causality approach (Gourevitch et al., 2006) showed that functions similar to SdDTF (directed coherence, partial directed coherence) appear to correctly identify linear linkages even if the autoregressive components are non-linear. On the other hand, non-linear Granger causality can yield interesting results for complex systems, but remains dependent on the parameters of the method (order and scale chosen). Linear methods can correctly identify frequency-specific causal interactions if the analysis includes the relevant frequencies. In the functioning brain, it is likely that there are always causal interactions between neural populations in multiple brain regions. Therefore, to identify task-specific patterns of interaction, it is necessary to examine changes in those baseline interactions that correlate with a task.To evaluate the statistical significance of event-related changes in SdDTF (i.e., ERC), we implement a statistical test to compare pre-stimulus (baseline) with post-stimulus SdDTF values. Specifically, a semi-parametric regression model is applied to SdDTF values calculated from pre- and post-stimulus periods. The windowing strategy described earlier can be viewed as a first step in smoothing the time-dependent SdDTF function, especially when the analysis windows are overlapping. However some of the noise inherent in the original signal will be resistant to this smoothing method. Hence, we employ a formal bivariate smoothing model that takes into account both the frequency *f*, and the temporal window *t*, which is defined as:

$$Y_{f,t} = g(f,t) + \varepsilon_{f,t}, \tag{8}$$

where $g(f,t)$ is an unspecified function representing the actual SdDTF function and ε_{ft} are independent $N(0,\sigma_\varepsilon^2)$ random variables capturing the white noise around the signal. There are many nonparametric approaches to bivariate smoothing, but here we use a penalized thin-plate spline model for $g(.,.)$. The model was implemented in R using the SemiPar software package[1]. The method and its implementation in R has been described previously in detail (Ruppert et al., 2003).

The SdDTF is a non-stationary function, both in baseline and post-stimulus periods, accounting for the non-stationarity of the baseline signal and represents a recent improvement over previous approaches (Korzeniewska et al., 2008). The mean SdDTF value of each pre-stimulus baseline window is compared with the mean SdDTF value of each post-stimulus window using a *t*-test designed for the null hypothesis of zero differences between the SdDTF means. We conclude that there is significant event-related change in causal interactions within a given post-stimulus time window if the SdDTF value for this window is significantly different from all SdDTF values in the baseline period. If the SdDTF value for the post-stimulus time *T* is significantly higher than all values of SdDTF for every time *t* of the baseline period, we say that there is a significant increase in causal interaction. Our goal was to test for every frequency *f*, and for every baseline/stimulus pair of time windows (t,T), whether $g(f,t) = g(f,T)$. More precisely, the implicit null hypothesis was expressed as:

$$\begin{aligned} H_{0,f,T} : g(f,t_1) &= g(f,T) \text{ or } g(f,t_2) \\ &= g(f,T) \text{ or} \ldots \text{or } g(f,t_n) = g(f,T) \end{aligned} \tag{9}$$

with the corresponding alternative:

$$\begin{aligned} H_{A,f,T} : g(f,t_1) &\neq g(f,T) \text{ and } g(f,t_2) \\ &\neq g(f,T) \text{ and} \ldots \text{and } g(f,t_n) \neq g(f,T). \end{aligned} \tag{10}$$

These hypotheses were tested by constructing a joint 95% confidence interval for the differences $g(f,t) - g(f,T)$ for $t = t_1, \ldots, t_n$. Let $\hat{g}(f,t), \hat{\sigma}_g(f,t)$ be the penalized spline estimator of $g(f,t)$ and its associated estimated standard error in each baseline time window. Similarly, let $\hat{g}(f,T), \hat{\sigma}_g(f,T)$ be the penalized spline estimator of $g(f,T)$ and its associated estimated standard error in each post-stimulus time window. Since the penalized spline functions are fitted locally, the residuals are assumed to be independent at points well separated in time and randomly distributed. We can also assume that for every baseline/stimulus pair of time windows (t,T):

$$\frac{\hat{g}(f,t) - g(f,T) - \hat{g}(f,t) + g(f,T)}{\sqrt{\hat{\sigma}^2(f,t) + \hat{\sigma}_g^2(f,T)}} \sim N(0,1) \tag{11}$$

[1]http://www.uow.edu.au/~mwand/SemiPar.html

approximates a standard normal distribution. We confirmed these assumptions with the Kolmogorov–Smirnov normality test. A joint confidence interval with at least 95% coverage probability for $g(f,t) - g(f,T)$ is defined as:

$$\hat{g}(f,t) - \hat{g}(f,T) \pm m_{.95}\sqrt{\hat{\sigma}^2(f,t) + \hat{\sigma}_g^2}(f,T), \tag{12}$$

where $m_{.95}$ is the 97.5% quantile of the distribution:

$$\mathrm{MAX}(t_n, T_n) = \max_{t_1 \le t \le t_n,\ T_1 \le T \le T_n,\ f_1 \le f \le f_m} \left| N_{t,T,f} \right|, \tag{13}$$

where $N_{t,T,f}$ are independent $N(0,1)$ random variables. This test rejected $H_{0,f,T}$ if 0 was not contained in any of the corresponding confidence intervals. To account for multiple comparisons, either a Bonferroni correction or the less conservative FDR can be implemented. The choice of correction method will depend on whether there is greater concern about incorrectly assigning statistical significance to a particular pattern, as in an initial exploratory analysis, or about failing to detect statistically relevant patterns (for detailed discussion see Korzeniewska et al., 2008). By definition, the ERC method provides an estimate of the directions and magnitudes of statistically significant event-related changes in direct activity propagation between brain sites, as a function of frequency. In other words, ERC corresponds to SdDTF, but is masked according to the statistical significance of event-related changes in SdDTF.

ERC METHODOLOGICAL CONSIDERATIONS

The number of data samples and length of the time window are two important considerations in applying the ERC method. A sufficient number of data samples are needed for the MVAR model to fit appropriately the recording data. Similarly, the length of the data analysis window should be sufficiently short to allow the data to be treated as stationary, but not so small that it precludes measuring jitter in the recorded signal across trials. It is recommended that the number of parameters be <10% of the total number of data samples. The number of data samples should also be several times greater than the number of channels (K). As in previous studies, we estimate the sufficient number of data samples by the inequality:

$$\frac{K(p+1)}{N_s n_t} < 0.1, \tag{14}$$

where N_s is the length of the moving window (e.g., the number of samples per recording epoch) and n_t is the total number of trials. Selection of recording channels can be guided by results of the single-channel MP analyses (described in Spectral Analysis). The rationale for this is that event-related causal interaction between ECoG signals is more likely to occur at sites where an event-related increase in signal energy is evident.

SIGNAL PRE-PROCESSING FOR ERC ANALYSIS

The raw ECoG time series is first pre-processed as for single-channel spectral analyses (see Signal Pre-processing). Remontaging to a common average reference (see Signal Pre-processing) is useful for removing unrelated global activity prior to ERC analysis (Yao et al., 2005, 2007; Ludwig et al., 2009). For ERC analysis, pre-processing is important to remove artifact, including high-frequency noise, to select specific frequency bands for analysis, and to remove phase-locked activity from the signal. To accomplish the first two objectives, the ECoG signal is digitally band pass-filtered and down-sampled. Signals can be filtered to include a single frequency range or multiple frequency ranges. However, it is important to ensure that the filter does not change the signals' phase properties and that the filter's impulse response is short.

For ERC analysis, the third purpose of signal pre-processing is to remove the phase-locked activity. As discussed earlier (see Spectral Analysis), the resulting non-phase-locked activity, previously considered 'noise' in ERP studies, contains task-relevant information (Kalcher and Pfurtscheller, 1995; Ding et al., 2000) that cannot be inferred solely from the ERP (Crone et al., 2001a; Senkowski and Herrmann, 2002; Senkowski et al., 2007). Moreover, causality analyses with and without subtraction of the ensemble average have revealed spurious causality responses when subtraction was not performed (Oya et al., 2007). To remove phase-locked components that may obscure non-phase-locked activity and to meet MVAR model requirements, the mean signal values in each window are computed and subtracted from the signal. This results in a zero mean signal in each window, which is required for fitting the MVAR model (Eq. 4) To normalize signal amplitudes across channels, the signal in each window is then divided by its standard deviation. This normalization allows comparison of flow changes between different stages of task processing and different channel pairs independent of the relative amplitudes of the signals (Ding et al., 2000).

ESTIMATING ERC IN AUDITORY EVENT-RELATED ECoG

Figure 4 illustrates results of a recent ERC analysis of auditory event-related responses from an adult patient who had a focal right parietal dysplasia, with complex partial seizures, and who had a right subdural electrode grid implanted for pre-surgical monitoring. The top panel depicts a transmission matrix of statistically significant event-related changes in the flow of activity between electrode sites during the first 200 ms after presentation of a speech syllable (/da/; 300 ms). A number of relevant transmissions can be seen. The location of the most prominent flows occurs in the first 150 ms at recording sites on the lateral posterior superior temporal gyrus, corresponding to auditory areas known to be critical for processing complex sounds, including speech (Miglioretti and Boatman, 2003; Boatman et al., 2000; Boatman, 2006; Sinai et al., 2009). The relationships between sites of sound processing are illustrated in the bottom panel of **Figure 4**. The arrows represent integrals of changes in causal interactions during the time course of sound processing. The color and width of the arrows represent the magnitude of integrals, over the analyzed period, of statistically significant ERC values. The cluster of arrows focused on the posterior superior temporal gyrus and inferior parietal cortex are consistent with the proposed local processing networks for complex sounds in auditory association cortex (Crone et al., 2001a; Boatman, 2004, 2006; Boatman and Miglioretti, 2005; Edwards et al., 2005; Lachaux et al., 2007). The directionality and magnitude of the changes in causal interactions within this local processing network can be represented schematically, as shown in **Figure 5**. These results illustrate the

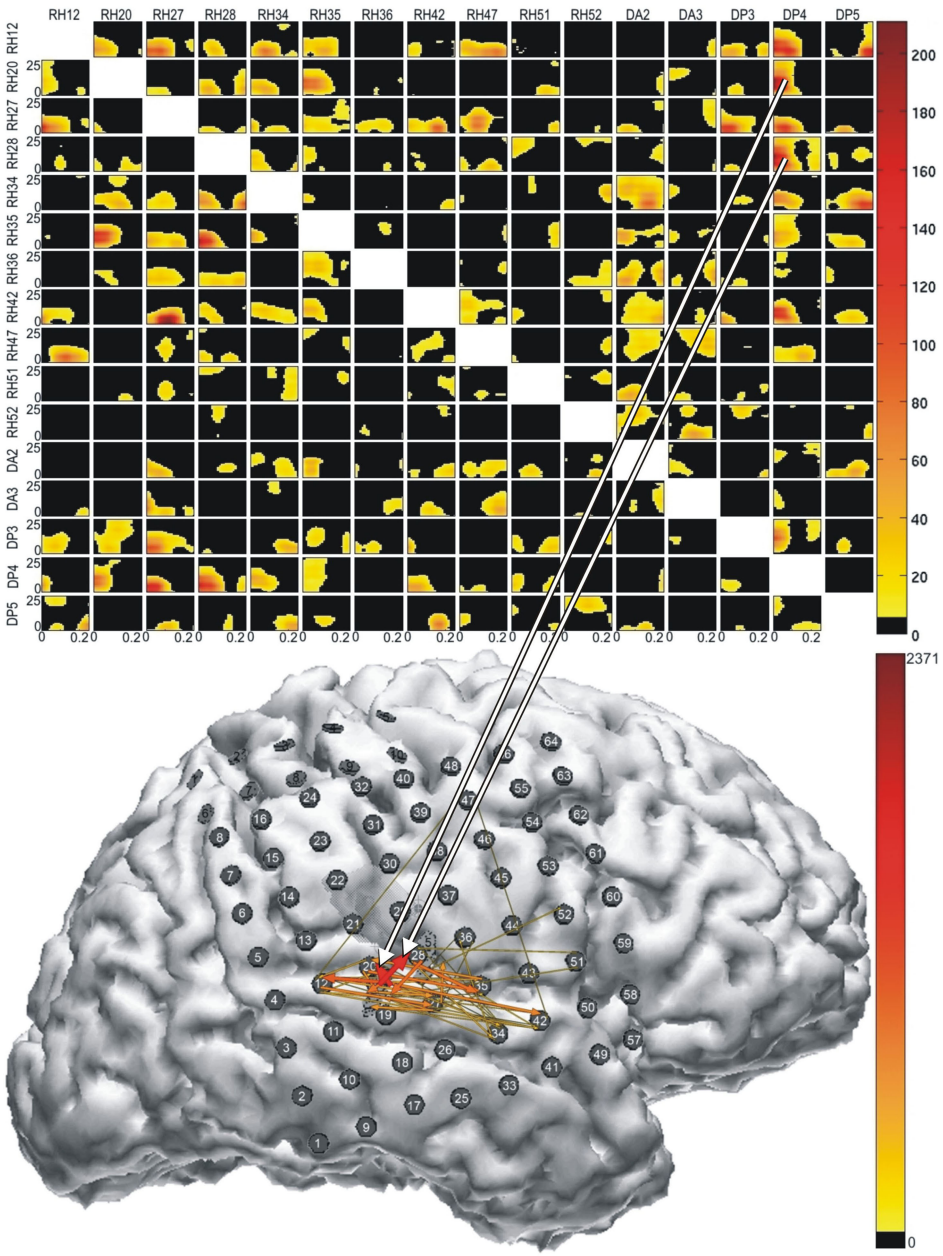


FIGURE 4 | Results of multichannel event-related causality (ERC) analysis of ECoG signals recorded from a patient listening to speech sounds (syllables). Top panel shows ERC matrix. Direction of ERC flow is from the electrode labeled above each column to electrode labeled at left of each row. Time is on the horizontal axis (1–250 ms); frequency (0–25 Hz) is on the vertical axis. Color scale of ERC values (0 to max) is shown to right of array. Black indicates time–frequency points with no significant difference between SdDTF values after stimulus and SdDTFs for baseline. Colored time–frequency points indicate increases in post-stimulus SdDTF relative to baseline SdDTF (red > orange > yellow). Bottom panel depicts integrals of ERC for frequency range 0–25 Hz calculated for the same ECoG data set. Arrows indicate directionality of ERC. Width and color of arrows represent values of ERC integrals. Color scale is at the right. For clarity, only integrals for event-related flow increases are shown. White projecting lines show how matrix cells in top panel correspond to electrode recording channels in bottom panel.

utility of multichannel ERC analyses, which provide information about effective connectivity between cortical sites that cannot be obtained from single-channel analyses. We view these two methodological approaches as largely complementary; each provides important information about the functional organization of the cortical auditory system.

SPATIAL NORMALIZATION

Electrode placement for intracranial monitoring is determined by each patient's clinical circumstances, resulting in restricted spatial sampling in individual patients and variability across patients. The ability to compare electrode locations across patients has become a challenge as ECoG studies have expanded from single

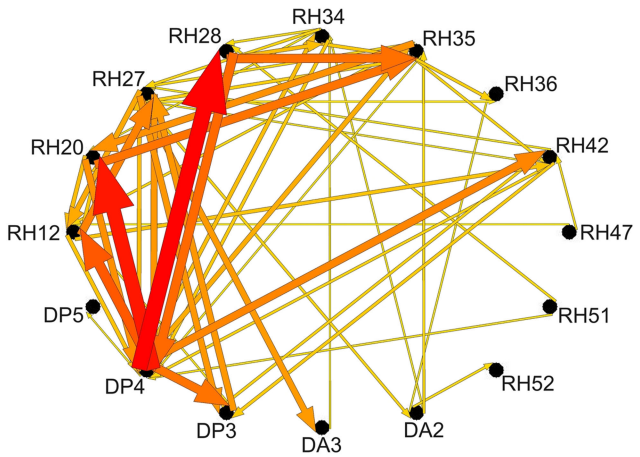


FIGURE 5 | Schematic representation of ERC integrals for frequency range 0–25 Hz as shown in bottom panel of Figure 4. Labels around the outside represent electrode channels shown in **Figure 4**. Arrows indicate directionality of ERC. Arrow width and color represent the values of ERC integrals shown in **Figure 4**.

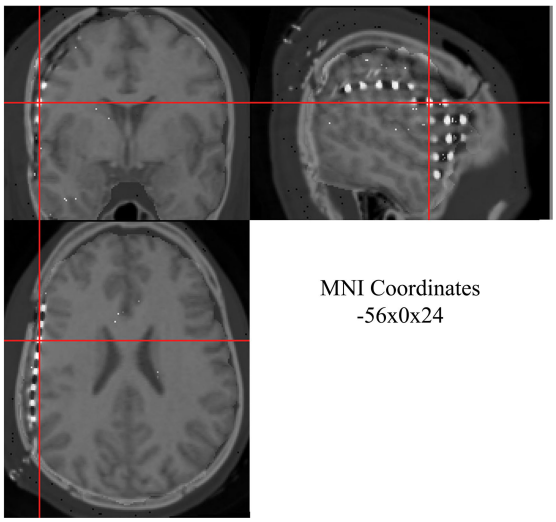


FIGURE 6 | Co-registered 3D MRI and CT images from one patient displayed in MRIcro. Three linked views are shown with same subdural electrode selected at crosshair: coronal (top left), saggital (top right), and axial (bottom left). Normalized MNI coordinates for selected subdural electrode are displayed in bottom right window.

case reports to include larger numbers of subjects. Volumetric three-dimensional (3D) MRI scans are obtained routinely before electrode implantation surgery and 3D CT scans are often used for post-implantation imaging. To localize electrodes in individual patients, the pre-implantation MRI and post-implantation CT scans must be co-registered. To compare electrode locations across patients (groups), individual 3D electrode positions are then transformed to a common reference space. The Talairach and Montreal Neurological Institute (MNI) 3D coordinate systems are standard reference systems for reporting brain locations in functional neuroimaging studies. Here we describe a semi-automated method to determine the 3D locations of intracranial electrodes (Ritzl et al., 2007). This method uses two freely available software programs – SPM and MRIcro – to co-register individual CT and MRI images and then transform electrode locations to a standard 3D reference space (Talairach, MNI) for group comparisons.

DATA PRE-PROCESSING

The pre-implantation volumetric MRI (1–1.8 mm coronal slices) and post-implantation CT (1 mm axial slices) scans are acquired in digital imaging and communication (DICOM) format. The MRI and corresponding CT data are then converted to Analyze format using MRIcro[2].

CO-REGISTRATION AND NORMALIZATION

The CT data are automatically co-registered onto MRI data from the same patient using SPM8[3] and a six-parameter rigid body transformation (Ritzl et al., 2007). The pre-implantation MRI is then normalized onto the standard MNI brain representation included in SPM8, using default normalization parameters. The 3D CT scan is then normalized using parameters derived from normalization of the 3D MRI.

NORMALIZED ELECTRODE COORDINATES

The MRI, CT, and 3D co-registered data can be displayed in spatially linked windows in MRIcro (**Figure 6**). This facilitates visualization and selection of individual electrodes. MRIcro automatically displays MNI coordinates of selected data points (electrodes). Talairach coordinates can be derived from MNI coordinates using the MATLAB mni2tal function[4].

Advantages of this semi-automatic normalization approach include: (1) it uses freely available software programs; (2) it is useful for combining different imaging data sets, including fMRI; (3) co-registration is automatic, thereby avoiding human error; and (4) it can also be used to localize depth electrodes implanted in deeper brain structures including the hippocampus. This method was developed for extraoperative ECoG studies in which electrodes are implanted. Other approaches have been developed for localization of electrode positions during intraoperative recording studies, including co-registration of electrode locations derived from infrared probes with pre-surgical MRI scans (Edwards et al., 2005).

The normalized electrode data may undergo further statistical modeling. For example, we have used template mixture modeling, a Bayesian hierarchical framework derived from normalized electrode coordinates, to quantify within- and between-patient variability in the distribution of cortical auditory responses (Miglioretti and Boatman, 2003; Boatman and Miglioretti, 2005).

ECoG METHODOLOGICAL AND STATISTICAL CONSIDERATIONS

LIMITATIONS OF ECoG STUDIES

A potential limitation of the intracranial (ECoG) method is that electrodes are usually implanted only over one hemisphere (seizure side), precluding recording from both hemispheres in

[2]http://www.sph.sc.edu/comd/rorden/mricro.html

[3]http://www.fil.ion.ucl.ac.uk/spm/software/spm8/

[4]http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach

the same patient. Some patients have strips implanted on the contralateral side for improved lateralization, but this is less common. Likewise, implanted electrodes rarely cover an entire hemisphere, further restricting spatial sampling within patients. There is also considerable individual variability in electrode placement across patients, and the spatial resolution of electrode arrays is high enough that important anatomical distinctions can exist between similarly placed arrays. This may pose additional challenges for statistical group comparisons. Another potential limitation is that patients who undergo invasive recordings usually have longstanding neurological disorders that may result in atypical functional organization. To increase the generalizability of results, we routinely screen patients beforehand to detect functional abnormalities, including hearing loss and auditory dysfunction (Boatman and Miglioretti, 2005; Sinai et al., 2009). Another potential concern is that the reliability of ECoG recordings has yet to be determined. This is particularly problematic since recordings are often done over multiple sessions (days), and changes in clinical status due to seizures or medications are likely to occur. Studies are underway at our center to examine test-retest reliability of different event-related response measures. Finally, recent studies have suggested that EEG recordings of gamma activity may be contaminated by ocular and muscle artifact. Specifically, it has been shown that high frequency responses in scalp EEG are influenced by micro-saccades (Yuval-Greenberg et al., 2008) and that recordings from the temporal pole region may influenced by myogenic artifact due to the proximity of extraocular muscles (Jerbi et al., 2009). These potential limitations need to be taken into consideration in the interpretation of ECoG findings.

MULTIPLE COMPARISONS

The multiple comparisons problem arises in ECoG studies because the event-related response of interest is measured at a large number of electrodes and time points requiring multiple statistical comparisons. Large numbers of statistical comparisons come with the potential to falsely reject the null hypothesis due to chance associations. The family-wise error rate is the probability of falsely concluding there is an effect (e.g., difference). The multiple comparisons problem can be resolved by controlling the family-wise error rate at a specified alpha level (e.g., 0.05). However, it is not possible to control the family-wise error rate by means of standard statistical methods that operate at the level of single samples (e.g., *t*-test).

Two correction methods are widely used in ECoG studies: the Bonferroni correction and the FDR (Benjamini and Hochberg, 1995). The Bonferroni correction restricts the so-called family-wise error rate (i.e., the probability of at least one false rejection under the null hypothesis) by dividing the type I error rate by the total number of comparisons performed. This procedure is very conservative because it ignores correlations in the hypothesis test outcomes and bounds the family-wise error rate, a criterion that is generally too strict to be practical for modern high-throughput studies such as ECoG. To address this issue, several modifications to the Bonferroni method have been developed, including the Holm-Bonferonni method that controls family-wise error rate at the α level, thereby allowing more opportunity for rejection of the null hypothesis (Holm, 1979). Alternatively, FDR is the expected proportion of falsely rejected null hypotheses for a specified threshold. The original work by Benjamini and Hochberg (1995) and recent work (Storey, 2002) has shown how to develop thresholding rules that bound FDR, not unlike the rules by which the Bonferroni correction bounds the family-wise error rate. While the FDR procedure tends to be less conservative than the Bonferroni, both methods have been used to determine the statistical significance of event-related responses in multichannel ECoG data (Durka et al., 2004; Edwards et al., 2005; Sinai et al., 2009). Because these two correction methods have different purposes, they are therefore not mutually exclusive.

In our time–frequency studies, the Bonferroni and FDR have yielded similar results. One potentially useful strategy is to combine both methods in a two-stage process: first implement the FDR method to identify data trends and then apply the Bonferroni method to verify the results. A promising new approach for handling multiple comparisons in ECoG data involves applying non-parametric permutation testing to estimate statistical significance (Maris and Oostenveld, 2007; Jacobs and Kahana, 2009) – a procedure that is gaining wide acceptance in neuroimaging studies (Nichols and Holmes, 2002). When applying these non-parametric tests, it is important to use a sufficiently large number of permutations to achieve convergence to asymptotic values. As long as test results continue to change when the number of permutations is increased, they are considered not yet reliable.

CONCLUSIONS

We propose a comprehensive analytic framework that combines multiple, complementary methods for evaluating the statistical significance of event-related responses in ECoG data sets. We demonstrated the utility of this approach for intracranial auditory mapping studies. The individual methods described have been used in ECoG studies of sensory, motor, language, and cognitive functions (Ray et al., 2003; Sinai et al., 2005; Canolty et al., 2007; Miller et al., 2007; Oya et al., 2007; Jacobs and Kahana, 2009) as well as studies of cortical abnormalities, including seizures (Franaszczuk et al., 1994, 1998). The combination of multiple complementary single-channel and multichannel methods in a comprehensive unified framework is novel and potentially more powerful than the traditional single-method approach. This methodological framework may also be useful for analyzing intracortical recordings of local field potentials in animal studies. Future directions include development of new statistical approaches for quantifying differences in the temporal-spectral shape of event-related responses across subjects and experimental conditions (stimulus, task) and for integration of multimodal brain mapping data, including fMRI and whole-head magnetoencephalography (MEG).

ACKNOWLEDGMENTS

The analytic framework and ECoG data described in this paper were developed with support from NIH grants RO1-DC05645 (Dana Boatman-Reich), K24-DC010028 (Dana Boatman-Reich), RO1-NS040596 (Nathan E. Crone), and RO1-NSO60910 (Brian Caffo). Special thanks to Paras Bhatt for assistance with the figures.

REFERENCES

Akaike, H. (1974). New Look at the Statistical-Model Identification. *IEEE Trans. Automatic Control* 19, 716–723.

Anderson, K. L., Rajagovindan, R., Ghacibeh, G. A., Meador, K. J., and Ding, M. (2009). Theta oscillations mediate interaction between prefrontal cortex and medial temporal lobe in human memory. *Cereb. Cortex* (in press). doi: 10.1093/cercor/bhp223.

Androulidakis, A. G., Mazzone, P., Litvak, V., Penny, W., Dileone, M., Gaynor, L. M., Tisch, S., Di Lazzaro, V., and Brown, P. (2008). Oscillatory activity in the pedunculopontine area of patients with Parkinson's disease. *Exp. Neurol.* 211, 59–66.

Astolfi, L., Cincotti, F., Mattia, D., Babiloni, C., Carducci, F., Basilisco, A., Rossini, P. M., Salinari, S., Ding, L., Ni, Y., and others. (2005). Assessing cortical functional connectivity by linear inverse estimation and directed transfer function: simulations and application to real data. *Clin. Neurophysiol.* 116, 920–932.

Astolfi, L., Cincotti, F., Mattia, D., De Vico Fallani, F., Colosimo, A., Salinari, S., Marciani, M.G., Ursino, M., Zavaglia, M., Hesse, W., Witte, H., and Babiloni, F. (2007a). Time-varying cortical connectivity by adaptive multivariate estimators applied to a combined foot-lips movement. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2007, 4402–4405.

Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccala, L. A., De Vico, F. F., Salinari, S., Ursino, M., Zavaglia, M., Ding, L., Edgar, J. C., Miller, G. A., He, B., and Babiloni, F. (2007b). Comparison of different cortical connectivity estimators for high-resolution EEG recordings. *Hum. Brain Mapp.* 28, 143–157.

Astolfi, L., Cincotti, F., Mattia, D., Mattiocco, M., De Vico Fallani, F., Colosimo, A., Marciani, M. G., Hesse, W., Zemanova, L., Lopez, G. Z., Kurths, J., Zhou, C., and Babiloni, F. (2006). Estimation of the time-varying cortical connectivity patterns by the adaptive multivariate estimators in high resolution EEG studies. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 1, 2446–2449.

Astolfi, L., Cincotti, F., Mattia, D., Salinari, S., Babiloni, C., Basilisco, A., Rossini, P. M., Ding, L., Ni, Y., He, B., Marciani, M. G., and Babiloni, F. (2004). Estimation of the effective and functional human cortical connectivity with structural equation modeling and directed transfer function applied to high-resolution EEG. *Magn. Reson. Imaging* 22, 1457–1470.

Babiloni, F., Ferri, R., Binetti, G., Vecchio, F., Brisoni, G. B., Lanuzza, B., Miniussi, C., Nobili, F., Rodriguez, G., Rundo, F., Cassarino, A., Infarinato, F., Cassetta, E., Salinari, S., Eusebi, F., and Rossini, P. M. (2009). Directionality of EEG synchronization in Alzheimer's disease subjects. *Neurobiol. Aging* 30, 93–102.

Babiloni, C., Vecchio, F., Cappa, S., Pasqualetti, P., Rossi, S., Miniussi, C., and Rossini, P. M. (2006). Functional frontoparietal connectivity during encoding and retrieval processes follows HERA model. A high-resolution study. *Brain Res. Bull.* 68, 203–212.

Baccala, L. A., and Sameshima, K. (2001a). Overcoming the limitations of correlation analysis for many simultaneously processed neural structures. *Prog. Brain Res.* 130, 33–47.

Baccala, L. A., and Sameshima, K. (2001b). Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* 84, 463–474.

Barbero, A., Franz, M., van Drongelen, W., Dorronsoro, J. R., Scholkopf, B., and Grosse-Wentrup, M. (2009). Implicit wiener series analysis of epileptic seizure recordings. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 1, 5304–5307.

Barnett, G. H., Burgess, R. C., Awad, I. A., Skipper, G. J., Edwards, C. R., and Luders, H. (1990). Epidural peg electrodes for the presurgical evaluation of intractable epilepsy. *Neurosurgery* 27, 113–115.

Bekisz, M., and Wrobel, A. (1999). Coupling of beta and gamma activity in corticothalamic system of cats attending to visual stimuli. *Neuroreport* 10, 3589–3594.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* 57, 289–300.

Bernasconi, C., and Konig, P. (1999). On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. *Biol. Cybern.* 81, 199–210.

Blount, J. P., Cormier, J., Kim, H., Kankirawatana, P., Riley, K., and Knowlton, R. (2008). Advances in intracranial monitoring. *Neurosurg. Focus* 25, E18. doi: 10.3171/FOC/2008/25/E18.

Boatman, D. (2004). Cortical bases of speech perception: functional lesion studies. *Cognition* 92, 47–65.

Boatman, D. (2006). Cortical auditory systems: speech and other complex sounds. *Epilepsy Behav.* 8, 494–503.

Boatman, D., Hart, J., Gordon, B., Selnes, O., Miglioretti, D., and Lenz, F. (2000). Transcortical sensory aphasia: revisited and revised. *Brain* 123, 1634–1642.

Boatman, D., and Miglioretti, D. (2005). Cortical sited critical for speech perception in normal and impaired listeners. *J. Neurosci.* 25, 5475–5480.

Breakspear, M., and Terry, J. R. (2002a). Detection and description of nonlinear interdependence in normal multichannel human EEG data. *Clin. Neurophysiol.* 113, 735–753.

Breakspear, M., and Terry, J. R. (2002b). Topographic organization of nonlinear interdependence in multichannel human EEG. *Neuroimage* 16, 822–835.

Brovelli, A., Lachaux, J. P., Kahane, P., and Boussaoud, D. (2005). High gamma frequency oscillatory activity dissociates attention from intention in the human premotor cortex. *Neuroimage* 28, 154–164.

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., and Howard, M. A. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *J. Neurophysiol.* 102, 2358–2374.

Cadotte, A. J., DeMarse, T., He, P., and Ding, M. (2008). Causal measures of structure and plasticity in simulated and living neural networks. *PLoS ONE* 7, e3355. doi: 10.1371/journal/pone.0003355.

Cadotte, A. J., Mareci, T. H., DeMarse, T. B., Parekh, M. B., Rajagovindan`, R., Ditto, W. L., Talathi, S. S., Hwang, D. U., and Carney, P. R. (2009). Temporal lobe epilepsy: anatomical and effective connectivity. *IEEE Trans. Neural Syst. Rehabil. Eng.* 17, 214–223.

Canolty, R. T., Soltani, M., Dalal, S. S., Edwards, E., Dronkers, N. F., Nagarajan, S. S., Kirsch, H. E., Barbaro, N. M., and Knight, R. T. (2007). Spatiotemporal dynamics of word processing in the human brain. *Front. Neurosci.* 1, 185–196.

Chavez, M., Martinerie, J., and Le Van Quyen, M. (2003). Statistical assessment of nonlinear causality: application to epileptic EEG signals. *J. Neurosci. Methods* 124, 113–28.

Crone, N. E., Boatman, D., Gordon, B., and Hao, L. (2001a). Induced electrocorticographic gamma activity during auditory perception. *Clin. Neurophysiol.* 112, 565–582.

Crone, N. E., Hao, L., Hart, J., Boatman, D., Lesser, R. P., Irizarry, R., and Gordon, B. (2001b). Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology* 57, 2045–2053.

Crone, N. E., Korzeniewska, A., Ray, S., and Franaszczuk, P. J. (2009). "Cortical function mapping with intracranial EEG". in *Quantitative EEG Analysis Methods and Application*, eds S. Tong and N. V. Thakor (Norwood, MA: Artech House), 355–366.

Crone, N. E., Miglioretti, D., Gordon, B., and Lesser, R. P. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* 121, 2301–2315.

Crone, N. E., Sinai, A., and Korzeniewska, A. (2006). High-frequency gamma oscillations and human brain mapping with electrocorticography. *Prog. Brain Res.* 159, 275–295.

David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., and Friston, K. J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. *Neuroimage* 30, 1255–1272.

De Gennaro, L., Vecchio, F., Ferrara, M., Curcio, G., Rossini, P. M., and Babiloni, C. (2004). Changes in fronto-posterior functional coupling at sleep onset in humans. *J. Sleep Res.* 13, 209–217.

De Gennaro, L., Vecchio, F., Ferrara, M., Curcio, G., Rossini, P. M., and Babiloni, C. (2005). Antero-posterior functional coupling at sleep onset: changes as a function of increased sleep pressure. *Brain Res. Bull.* 65, 133–140.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.

Deshpande, G., Hu, X., Stilla, R., and Sathian, K. (2008). Effective connectivity during haptic perception: a study using Granger causality analysis of functional magnetic resonance imaging data. *Neuroimage* 40, 1807–1814.

Deshpande, G., LaConte, S., Peltier, S., and Hu, X. (2006). Directed transfer function analysis of fMRI data to investigate network dynamics. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 1, 671–674.

Ding, M., Bressler, S. L., Yang, W., and Liang, H. (2000). Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biol. Cybern.* 83, 35–45.

Durka, P. J., Ircha, D., Neuper, C., and Pfurtscheller, G. (2001). Time–frequency microstructure of event-related electro-encephalogram desynchronisation and synchronisation. *Med. Biol. Eng. Comput.* 39, 315–321.

Durka, P. J., Zygierewicz, J., Klekowicz, H., Ginter, J., and Blinowska, K. J. (2004). On the statistical significance of event-related EEG desynchronization and synchronization in the time–frequency plane. *IEEE Trans. Biomed. Eng.* 51, 1167–1175.

Edin, F., Klingberg, T., Stodberg, T., and Tegner, J. (2007). Fronto-parietal connection asymmetry regulates working

memory distractibility. *J. Integr. Neurosci.* 6, 567–596.

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., and Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J. Neurophysiol.* 94, 4269–4280.

Eichler, M. (2005). A graphical approach for evaluating effective connectivity in neural systems. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 360, 953–967.

Eichler, M. (2006). "Graphical modeling of dynamic relationships in multivariate time series". in *The Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, eds B. Schelter, M. Winterhalder and J. Timmer (Weinheim: Wiley-VCH), 335–372.

Engel, A. K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25.

Franaszczuk, P. J., and Bergey, G. K. (1998). Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures. *Brain Topogra.* 11, 13–21.

Franaszczuk, P. J., and Bergey, G. K. (1999). An autoregressive method for the measurement of synchronization of interictal and ictal EEG signals. *Biol. Cybern.* 81, 3–9.

Franaszczuk, P. J., Bergey, G. K., Durka, P. J., and Eisenberg, H. M. (1998). Time–frequency analysis using the matching pursuit algorithm applied to seizures originating from the mesial temporal lobe. *Electroencephalogr. Clin. Neurophysiol.* 106, 513–521.

Franaszczuk, P. J., Bergey, G. K., and Kaminski, M. (1994). Analysis of mesial temporal seizure onset and propagation using the directed transfer function method. *Electroencephalogr. Clin. Neurophysiol.* 91, 413–427.

Franaszczuk, P. J., Blinowska, K. J., and Kowalczyk, M. (1985). The application of parametric multichannel spectral estimates in the study of electrical brain activity. *Biol. Cybern.* 51, 239–247.

Freiwald, W. A., Valdes, P., Bosch, J., Biscay, R., Jimenez, J. C., Rodriguez, L. M., Rodriguez, V., Kreiter, A. K., and Singer, W. (1999). Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. *J. Neurosci. Methods.* 94, 105–119.

Friston, K. J. (1997). Another neural code? *Neuroimage* 5, 213–220.

Friston, K. J., Penny, W., and David, O. (2005). Modeling brain responses. *Int. Rev. Neurobiol.* 66, 89–124.

Friston, K. J., Tononi, G., Reeke, G. N., Sporns, O., and Edelman, G. M. (1994). Value-dependent selection in the brain: simulation in a synthetic neural model. *Neuroscience* 59, 229–243.

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., and Friston, K. J. (2007). Dynamic causal modelling of evoked potentials: a reproducibility study. *Neuroimage* 36, 571–580.

Ge, M., Jiang, X., Bai, Q., Yang, S., Gusphyl, J., and Yan, W. (2007). Application of the directed transfer function method to the study of the propagation of epilepsy neural information. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2007, 3266–3269.

Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., and Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 322, 96–101.

Gevins, A., Leong, H., Smith, M. E., Le, J., and Du, R. (1995). Mapping cognitive brain function with modern high-resolution electroencephalography. *Trends Neurosci.* 18, 429–436.

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

Ginter, J., Blinowska, K. J., Kaminski, M., and Durka, P. J. (2001). Phase and amplitude analysis in time–frequency space – application to voluntary finger movement. *J. Neurosci. Methods* 110, 113–124.

Ginter, J., Blinowska, K. J., Kaminski, M., Durka, P. J., Pfurtscheller, G., and Neuper, C. (2005). Propagation of EEG activity in the beta and gamma band during movement imagery in humans. *Methods Inf. Med.* 44, 106–113.

Godey, B., Schwartz, D., de Graaf, J. B., Chauvel, P., and Liegeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clin. Neurophysiol.* 112, 1850–1859.

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

Gow, D. W., Keller, C. J., Eskandar, E., Meng, N., and Cash, S. S. (2009). Parallel versus serial processing dependencies in the perisylvian speech network: a Granger analysis of intracranial EEG data. *Brain Lang.* 110, 43–48.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.

Halgren, E., Marinkovic, K., and Chauvel, P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalogr. Clin. Neurophysiol.* 106, 156–164.

Hesse, W., Möller, E., Arnold, M., and Schack, B. (2003). The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. *J. Neurosci. Methods* 124, 27–44.

Hinrichs, H., Heinze, H. J., and Schoenfeld, M. A. (2006). Causal visual interactions as revealed by an information theoretic measure and fMRI. *Neuroimage* 31, 1051–1060.

Hinrichs, H., Noesselt, T., and Heinze, H. J. (2008). Directed information flow: a model free measure to analyze causal interactions in event related EEG-MEG-experiments. *Hum. Brain Mapp.* 29, 193–206.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.

Hong, B., Guo, F., Liu, T., Gao, X., and Gao, S. (2009). N200-speller using motion-onset visual response. *Clin. Neurophysiol.* 120, 1658–1666.

Howard, M. A., Volkov, I. O., Abbas, P. J., Damasio, H., Ollendieck, M. C., and Granner, M. A. (1996). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res.* 724, 260–264.

Howard, M. A., Volkov, I. O., Mirsky R., Garell P. C., Noh, M. D., Granner, M. A.,Damasio, H., Steinschneider, M., Reale, R. A., Hind, J. E., and Brugge, J. F. (2000). Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol.* 416, 79–92.

Jacobs, J., and Kahana, M. J. (2009). Neural representations of individual stimuli in humans revealed by gamma-band electrocorticographic activity. *J. Neurosci.* 29, 10203–10214.

Jensen, O., and Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *Eur. J. Neurosci.* 15, 1395–1399.

Jerbi, K., Ossandon, T., Hamame, C. M., Senova, S., Dalal, S. S., Jung, J., Minotti, L., Bertrand, O., Berthoz, A., Kahane, P., and Lachaux, J. P. (2009). Task-related gamma-band dynamics from an intracerebral perspective: review and implications for surface EEG and MEG. *Hum. Brain Mapp.* 30, 1758–1771.

Kalcher, J., and Pfurtscheller, G. (1995). Discrimination between phase-locked and non-phase-locked event-related EEG activity. *Electroencephalogr. Clin. Neurophysiol.* 94, 381–384.

Kaminski, M., and Blinowska, K. J. (1991). A new method of the description of the information flow in the brain structures. *Biol. Cybern.* 65, 203–210.

Kaminski, M., Ding, M., Truccolo, W. A., and Bressler, S. L. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.* 85, 145–157.

Kaminski, M., and Liang, H. (2005). Causal influence: advances in neurosignal analysis. *Crit. Rev. Biomed. Eng.* 33, 347–430.

Kaminski, M., Zygierewicz, J., Kus, R., and Crone, N. (2005). Analysis of multichannel biomedical data. *Acta. Neurobiol. Exp. (Wars)* 65, 443–452.

Keil, A., Sabatinelli, D., Ding, M., Lang, P. J., Ihssen, N., and Heim, S. (2009). Re-entrant projections modulate visual cortex in affective perception: evidence from Granger causality analysis. *Hum. Brain Mapp.* 30, 532–540.

Klimesch, W., Schimke, H., and Pfurtscheller, G. (1993). Alpha frequency, cognitive load and memory performance. *Brain Topogr.* 5, 241–251.

Knight, R. T., Scabini, D., Woods, D. L., and Clayworth, C. C. (1989). Contributions of temporal-parietal junction to the human auditory P3. *Brain Res.* 502, 109–116.

Korzeniewska, A., Crainiceanu, C. M., Kus, R., Franaszczuk, P. J., and Crone, N. E. (2008). Dynamics of event-related causality in brain electrical activity. *Hum. Brain Mapp.* 29, 1170–1192.

Korzeniewska, A., Kasicki, S., Kaminski, M., and Blinowska, K. J. (1997). Information flow between hippocampus and related structures during various types of rat's behavior. *J. Neurosci. Methods* 73, 49–60.

Korzeniewska, A., Manczak, M., Kaminski, M., Blinowska, K. J., and Kasicki, S. (2003). Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method. *J. Neurosci. Methods* 125, 195–207.

Krichmar, J. L., Seth, A. K., Nitz, D. A., Fleischer, J. G., and Edelman, G. M. (2005). Spatial navigation and causal analysis in a brain-based device modeling cortical-hippocampal interactions. *Neuroinformatics* 3, 197–221.

Kus, R., Blinowska, K. J., Kaminski, M., and Basinska-Starzycka, A. (2008). Transmission of information during Continuous Attention Test. *Acta Neurobiol. Exp. (Wars)* 68, 103–112.

Kus, R., Ginter, J. S., and Blinowska, K. J. (2006). Propagation of EEG activity during finger movement and its imagination. *Acta. Neurobiol. Exp. (Wars)* 66, 195–206.

Kus, R., Kaminski, M., and Blinowska, K. J. (2004). Determination of EEG activity propagation: pair-wise versus multichannel estimate. *IEEE Trans. Biomed. Eng.* 51, 1501–1510.

Lachaux, J. P., Jerbi, K., Bertrand, O., Minotti, L., Hoffmann, D., Schoendorff, B., and Kahane, P. (2007). A blueprint for real-time functional mapping via human intracranial recordings. *PLoS ONE* 2, e1094. doi: 10.1371/journal.pone.0001094.

Lalo, E., Thobois, S., Sharott, A., Polo, G., Mertens, P., Pogosyan, A., and Brown, P. (2008). Patterns of bidirectional communication between cortex and basal ganglia during movement in patients with Parkinson disease. *J. Neurosci.* 28, 3008–3016.

Liang, K. Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22.

Ludwig, K. A., Miriani, R. M., Langhals, N. B., Joseph, M. D., Anderson, D. J., and Kipke, D. R. (2009). Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *J. Neurophysiol.* 101, 1679–1689.

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. USA* 94, 10979–10984.

Mallat, S., and Zhang, Z. (1993). Matching pursuits with time–frequency dictionaries. *IEEE Trans. Signal Process* 41, 3397–3415.

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190.

Marple, S. (1987). *Digital Spectral Analysis with Applications*. Englewood Cliffs: Simon & Schuster.

Miglioretti, D., and Boatman, D. (2003). Modeling variability in the cortical representation of complex sound perception. *Exp. Brain Res.* 153, 382–387.

Miller, K. J., den Nijs, M., Shenoy, P., Miller, J. W., Rao, R. P., and Ojemann, J. G. (2007). Real-time functional brain mapping using electrocorticography. *Neuroimage* 37, 504–507.

Naatanen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38, 1–21.

Naatanen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590.

Naatanen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425.

Neuper, C., and Pfurtscheller, G. (2001). Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates. *Int. J. Psychophysiol.* 43, 41–58.

Nichols, T. E., and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin. Neurophysiol.* 115, 2292–2307.

Ojemann, G. A., Schoenfield-McNeill, J., and Corina, D. P. (2002). Anatomic subdivisions in human temporal cortical neuronal activity related to recent verbal memory. *Nat. Neurosci.* 5, 64–71.

Onton, J., Westerfield, M., Townsend, J., and Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neurosci. Biobehav. Rev.* 30, 808–822.

Oya, H., Poon, P. W., Brugge, J. F., Reale, R. A., Kawasaki, H., Volkov, I. O., and Howard, M. A. III. (2007). Functional connections between auditory cortical fields in humans revealed by Granger causality analysis of intracranial evoked potentials to sounds: comparison of two methods. *Biosystems* 89, 198–207.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., and Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalogr. Clin. Neurophysiol.* 94, 26–40.

Pfurtscheller, G., Woertz, M., Supp, G., and Lopes da Silva, F. H. (2003). Early onset of post-movement beta electroencephalogram synchronization in the supplementary motor area during self-paced finger movement in man. *Neurosci. Lett.* 339, 111–114.

Philiastides, M. G., and Sajda, P. (2006). Causal influences in the human brain during face discrimination: a short-window directed transfer function approach. *IEEE Trans. Biomed. Eng.* 53, 2602–2605.

Polich, J., and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol. Psychol.* 41, 103–146.

Ray, S., Jouny, C. C., Crone, N. E., Boatman, D., Thakor, N. V., and Franaszczuk, P. J. (2003). Human ECoG analysis during speech perception using matching pursuit: a comparison between stochastic and dyadic dictionaries. *IEEE Trans. Biomed. Eng.* 50, 1371–1373.

Ritzl, E. K., Wohlschlaeger, A. M., Crone, N. E., Wohlschlaeger, A., Gingis, L., Bowers, C. W., and Boatman, D. (2007). Transforming electrocortical mapping data into standardized common space. *Clin. EEG Neurosci.* 38, 132–136.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.

Sameshima, K., and Baccala, L. A. (1999). Using partial directed coherence to describe neuronal ensemble interactions. *J. Neurosci. Methods* 94, 93–103.

Sato, J. R., da Graca Morais Martin, M., Fujita, A., Mourao-Miranda, J., Brammer, M. J., and Amaro, E. (2008). An fMRI normative database for connectivity networks using one-class support vector machines. *Hum. Brain Mapp.* 30, 1068–1076.

Schack, B., Rappelsberger, P., Weiss, S., and Moller, E. (1999). Adaptive phase estimation and its application in EEG analysis of word processing. *J. Neurosci. Methods* 93, 49–59.

Schanze, T., and Eckhorn, R. (1997). Phase correlation among rhythms present at different frequencies: spectral methods, application to microelectrode recordings from visual cortex and functional implications. *Int. J. Psychophysiol.* 26, 171–189.

Schelter, B., Winterhalder, M., Eichler, M., Peifer, M., Hellwig, B., Guschlbauer, B., Lucking, C. H., Dahlhaus, R., and Timmer, J. (2006). Testing for directed influences among neural signals using partial directed coherence. *J. Neurosci. Methods* 152, 210–219.

Scherg, M., and von Cramon, D. (1986). Evoked dipole source potentials of the human auditory cortex. *Electroencephalogr. Clin. Neurophysiol.* 65, 344–360.

Schlogl, A., and Supp, G. (2006). Analyzing event-related EEG data with multivariate autoregressive parameters. *Prog. Brain Res.* 159, 135–147.

Schwartz, T. H., Haglund, M. M., Lettich, E., and Ojemann, G. A. (2000). Asymmetry of neuronal activity during extracellular microelectrode recording from left and right human temporal lobe neocortex during rhyming and line-matching. *J. Cogn. Neurosci.* 12, 803–812.

Senkowski, D., Gomez-Ramirez, M., Lakatos, P., Wylie, G. R., Molholm, S., Schroeder, C. E., and Foxe, J. J. (2007). Multisensory processing and oscillatory activity: analyzing non-linear electrophysiological measures in humans and simians. *Exp. Brain Res.* 177, 184–195.

Senkowski, D., and Herrmann, C. S. (2002). Effects of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. *Clin. Neurophysiol.* 113, 1742–1753.

Seth, A. K. (2005). Causal connectivity of evolved neural networks during behavior. *Network* 16, 35–54.

Shoker, L., Sanei, S., and Sumich, A. (2005). Distinguishing between left and right finger movement from EEG using SVM. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 5, 5420–5423.

Sinai, A., Bowers, C. W., Crainiceanu, C. M., Boatman, D., Gordon, B., Lesser, R. P., Lenz, F., and Crone, N. E. (2005). Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming. *Brain* 128, 1556–1570.

Sinai, A., Crone, N. E., Wied, H. M., Franaszczuk, P. J., Miglioretti, D., and Boatman-Reich, D. (2009). Intracranial mapping of auditory perception: event-related responses and electrocortical stimulation. *Clin. Neurophysiol.* 120, 140–149.

Singer, W. (1993). Synchronization of cortical activity and its putative role in information-processing and learning. *Annu. Rev. Phys.* 55, 349–374.

Slutzky, M. W., Jordan, L. R., and Miller, L. E. (2008). Optimal spatial resolution of epidural and subdural electrode arrays for brain–machine interface applications. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 3771–3774.

Sporns, O., Honey, C. J., and Kotter, R. (2007). Identification and classification of hubs in brain networks. *PLoS ONE* 2, e1049. doi: 1371/journal.pone.0001049.

Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. B Stat. Methodol.* 64, 479–498.

Struber, D., and Herrmann, C. S. (2002). MEG alpha activity decrease reflects destabilization of multistable percepts. *Brain Res. Cogn. Brain Res.* 14, 370–382.

Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci.* 3, 151–162.

Tiitinen, H., May, P., Reinikainen, K., and Naatanen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 372, 90–92.

Tononi, G., and Sporns, O. (2003). Measuring information integration. *BMC Neurosci.* 4, 31.

Towle, V. L., Yoon, H. A., Castelle, M., Edgar, J. C., Biassou, N. M., Frim, D. M., Spire, J. P., and Kohrman, M. H.

(2008). ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain* 131, 2013–2027.

Trautner, P., Rosburg, T., Dietl, T., Fell, J., Korzyukov, O. A., Kurthen, M., Schaller, C., Elger, C. E., and Boutros, N. N. (2006). Sensory gating of auditory evoked and induced gamma band activity in intracranial recordings. *Neuroimage* 32, 790–798.

Truccolo, W. A., Ding, M., Knuth, K. H., Nakamura, R., and Bressler, S. L. (2002). Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity. *Clin. Neurophysiol.* 113, 206–226.

Wilke, C., Ding, L., and He, B. (2007). An adaptive directed transfer function approach for detecting dynamic causal interactions. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 4949–4952.

Wilke, M., Lidzba, K., and Krageloh-Mann, I. (2009). Combined functional and causal connectivity analyses of language networks in children: a feasibility study. *Brain Lang.* 108, 22–29.

Winterhalder, M., Schelter, B., Hesse, W., Schwab, K., Leistritz, L., Klan, D., Bauer, R., Timmer, J., and Witte, H. (2005). Comparison directed of linear signal processing techniques to infer interactions in multivariate neural systems. *Signal Processing* 85, 2137–2160.

Yao, D., Wang, L., Arendt-Nielsen, L., and Chen, A. C. (2007). The effect of reference choices on the spatio-temporal analysis of brain evoked potentials: the use of infinite reference. *Comput. Biol. Med.* 37, 1529–1538.

Yao, D., Wang, L., Oostenveld, R., Nielsen, K. D., Arendt-Nielsen, L., and Chen, A. C. (2005). A comparative study of different references for EEG spectral mapping: the issue of the neutral reference and the use of the infinity reference. *Physiol. Meas.* 26, 173–184.

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., and Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron* 58, 429–441.

Zygierewicz, J., Durka, P. J., Klekowicz, H., Franaszczuk, P. J., and Crone, N. E. (2005). Computationally efficient approaches to calculating significant ERD/ERS changes in the time–frequency plane. *J. Neurosci. Methods* 145, 267–276.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 November 2009; paper pending published: 19 December 2009; accepted: 04 March 2010; published online: 19 March 2010.

Citation: Boatman-Reich D, Franaszczuk PJ, Korzeniewska A, Caffo B, Ritzl EK, Colwell S and Crone NE (2010) Quantifying auditory event-related responses in multichannel human intracranial recordings. Front. Comput. Neurosci. ***4****:4. doi: 10.3389/fncom.2010.00004*

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 26 May 2010
doi: 10.3389/fncom.2010.00014

Directed coupling in local field potentials of macaque V4 during visual short-term memory revealed by multivariate autoregressive models

Gregor M. Hoerzer[1]*†, Stefanie Liebe[2]*†, Alois Schloegl[3], Nikos K. Logothetis[2] and Gregor Rainer[2,4]

[1] *Institute for Theoretical Computer Science, Graz University of Technology, Graz, Austria*
[2] *Department of Physiology of Cognitive Processes, Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany*
[3] *Institute for Human-Computer Interfaces, Graz University of Technology, Graz, Austria*
[4] *Department of Medicine/Physiology, University of Fribourg, Fribourg, Switzerland*

Edited by:
Nicolas Brunel, Centre National de la Recherche Scientifique, France

Reviewed by:
Hualou Liang, Drexel University, USA
Mingzhou Ding, University of Florida, USA

****Correspondence:***
Gregor M. Hoerzer, Institute for Theoretical Computer Science, Graz University of Technology, Inffeldgasse 16b/I, A-8010 Graz, Austria.
e-mail: gregor@igi.tugraz.at
Stefanie Liebe, Department of Physiology of Cognitive Processes, Max-Planck-Institute for Biological Cybernetics, Spemannstrasse 38, D-72070 Tübingen, Germany.
e-mail: stefanie.liebe@tuebingen.mpg.de

†Gregor M. Hoerzer and Stefanie Liebe have contributed equally to this work.

Processing and storage of sensory information is based on the interaction between different neural populations rather than the isolated activity of single neurons. In order to characterize the dynamic interaction and transient cooperation of sub-circuits within a neural network, multivariate autoregressive (MVAR) models have proven to be an important analysis tool. In this study, we apply directed functional coupling based on MVAR models and describe the temporal and spatial changes of functional coupling between simultaneously recorded local field potentials in extrastriate area V4 during visual memory. Specifically, we compare the strength and directional relations of coupling based on generalized partial directed coherence (GPDC) measures while two rhesus monkeys perform a visual short-term memory task. In both monkeys we find increases in theta power during the memory period that are accompanied by changes in directed coupling. These interactions are most prominent in the low frequency range encompassing the theta band (3–12 Hz) and, more importantly, are asymmetric between pairs of recording sites. Furthermore, we find that the degree of interaction decreases as a function of distance between electrode positions, suggesting that these interactions are a predominantly local phenomenon. Taken together, our results show that directed coupling measures based on MVAR models are able to provide important insights into the spatial and temporal formation of local functionally coupled ensembles during visual memory in V4. Moreover, our findings suggest that visual memory is accompanied not only by a temporary increase of oscillatory activity in the theta band, but by a direction-dependent change in theta coupling, which ultimately represents a change in functional connectivity within the neural circuit.

Keywords: multivariate autoregressive model, local field potential, partial directed coherence, visual short term memory, macaque monkey, theta frequency range, V4, multi-channel recordings

INTRODUCTION

Cortical oscillatory activity measured from local field potential (LFP) recordings or electroencephalogram (EEG) is a widespread neuronal phenomenon and is considered to underlie the communication of local and distant neural populations throughout the brain (Fries, 2005). Different parameters of oscillations in distinct frequency bands often show correlations with various aspects of sensory information processing (Buzsaki and Draguhn, 2004). A prominent example is the modulation of gamma synchrony in visual cognition, for example in tasks involving the manipulation of visual attention (Fries et al., 2001), binocular rivalry (Gail et al., 2004) or object recognition (Supp et al., 2007).

In contrast to visual processing, several studies revealed a specific role of theta oscillations (3–12 Hz) in mnemonic processing, for example in spatial memory in rodents (Okeefe, 1993; Buzsaki, 2005), working memory in humans (Klimesch, 1999; Raghavachari et al., 2001, 2006) and visual short-term memory in non-human primates (Rainer et al., 2004; Lee et al., 2005). In the latter study, neuronal oscillations in the theta band in extrastriate area V4 have been shown to mediate the coding and maintenance of relevant visual information within short-term memory. Thus, theta oscillations in V4 could provide a possible mechanism for supporting and coordinating cross-neuronal interactions within neuronal ensembles during visual memory. However, physiological evidence for directed oscillatory interactions in the theta-frequency range during short-term memory has not been obtained yet.

A description of the interaction patterns of oscillatory processes is provided by different measures that quantify various aspects of functional coupling. For example, some measures such as the phase-locking value (Lachaux et al., 1999, 2000) provide insights into the instantaneous phase-relationship between two oscillatory processes and are derived from Wavelet- or Hilbert transform-based methods. In contrast, coupling measures derived from multivariate autoregressive (MVAR) models are becoming increasingly important as they capture not only instantaneous interactions between neural signals, but can give insights into the causal relationship between oscillations as well as the direction of their interaction. Thus, MVAR models are powerful in capturing the complex nature of oscillatory interactions and their role in neural processing.

Multivariate autoregressive models are a generalization of univariate autoregressive (AR) models, which were among the first methods that were applied to EEG data to reveal the spectral properties of brain signals already in the late 1960s (Zetterberg, 1969). MVAR models are able to take the interactions of multiple simultaneously recorded brain signals into account. A large set of coupling measures in the frequency domain such as coherency (Nunez et al., 1997, 1999), Directed Transfer Function (DTF; Kaminski and Blinowska, 1991) or Partial Directed Coherence (PDC; Baccala and Sameshima, 2001) as well as variants of these and similar measures can be derived using the MVAR model parameters (Schlögl and Supp, 2006; Porcaro et al., 2009) and the implementation of coupling analyses is readily achieved by various toolboxes (Cui et al., 2008; Schlögl and Brunner, 2008).

Importantly, DTF and PDC, unlike coherency, assess the directionality of couplings between signals, i.e., they measure the direction of information flow between different channels. Both measures are based on the concept of Granger causality (Granger, 1969), which can be informally stated as follows: if the observation of a time series $x(t)$ significantly improves the prediction of a time series $y(t)$, $x(t)$ "Granger-causes" $y(t)$. PDC differs from DTF by having the ability to reveal exclusively direct couplings, which means that it does not assess indirect couplings via intermediate sites. For example, if the model incorporates three observed channels, with a connection structure $A \rightarrow B \rightarrow C$, PDC is not expected to show a connection from A to C. It is important to note that Granger causality is not identical to physical causality, but is a statistical measure reflecting the improvement of predictability of one signal based on the information of another.

Previously, AR models have been applied to EEG data and LFP data for various brain areas and frequency bands of interest and have revealed important insights into the functional relations between neuronal assemblies involved in sensorimotor behavior, sensory integration and visual attention (Bressler et al., 1999, 2007; Liang et al., 2000, 2001, 2003; Brovelli et al., 2004; Chen et al., 2006; Supp et al., 2007; Anderson et al., 2009; Kayser and Logothetis, 2009).

In the present study, we applied MVAR modeling to simultaneous LFP recordings from multiple electrodes in V4 while monkeys performed a visual identification task. MVAR models have been used previously to examine causal influences in area V4 in order to elucidate physiological mechanisms underlying neuronal oscillations in the alpha frequency range (i.e., 10–15 Hz) (Bollimunta et al., 2008).

In our study, our goal was to exploit the advantages of MVAR models in order to investigate the directed functional relationship between multiple sources underlying theta oscillations during visual memory in V4. In order to gain insights into the direct interaction between multiple oscillatory components (i.e., bypassing coupling due to indirect influences) our MVAR models incorporated LFP activity of more than two simultaneously recorded channels. In addition, we evaluated the temporal and spatial dynamics of these direct interactions and provide a first description of causal and directed oscillatory coupling in the theta-frequency range during visual memory.

MATERIALS AND METHODS

In the following, we describe the procedure that was used for our analysis. Afterwards, the experimental procedures for the data acquisition are described.

PREPROCESSING

Local field potential data was preprocessed using standard techniques, as described for example in Ding et al. (2000). First, we resampled the data to a frequency f_s of 200 Hz. This sampling rate is low enough to be able to use a sufficiently low MVAR model order while being high enough for an adequate representation of the frequency bands of interest. Then, we used a 50 Hz notch filter to suppress the electrical supply line noise. Afterwards, the data was normalized by subtracting the mean waveform across trials (grand-averaged mean waveform) from each single trial and subsequently dividing the result by the standard deviation across trials. This is necessary to remove first order instationarities from the data and to set the ensemble mean of the resulting data set to zero. We did not apply the same normalization procedure using the temporal mean and standard deviation for each separate trial, which is also frequently proposed, because this can lead to an underestimation of the low frequency components in which we were particularly interested.

MULTIVARIATE AUTOREGRESSIVE MODELING

To assess coupling between different LFP channels, we separately generated linear MVAR models of the data for each recording session and each time interval of interest. The MVAR model can be expressed as:

$$\underline{y}(t) = \sum_{p=1}^{P} \mathbf{A}_p \underline{y}(t-p) + \underline{x}(t).$$

The model tries to predict the data at sample t from a linear combination of the P previous samples of all M channels. Here, $\underline{y}(t)$ is the vector of M simultaneously observed LFP recordings, P is the model order stating the number of preceding samples that are used to predict the data at sample t, and the innovation process $\underline{x}(t)$ (sometimes addressed as the "residual error" or "prediction error", see Schlögl, 2000; Supp et al., 2007 for comments) is assumed to be a multivariate white noise process and is equal to the difference between the model prediction and the actual data. In order to estimate the model parameter matrices A_p that weight the previous samples of the time series to predict the current one such that the mean quadratic error is minimized, we use the Burg-type method of Vieira–Morf (Marple, 1987) which, according to Schlögl (2006), is expected to provide the most accurate estimates of the model parameters. We used 250 ms windows for the time-frequency analysis, with an overlap of 200 ms for subsequent time intervals (Ding et al., 2000 called this procedure an Adaptive MVAR or AMVAR approach), and 1 s windows for the assessment of statistical significance of coupling and change in coupling between the two investigated task conditions (cf. Experimental Task). Note that the model assumes the data to be stationary, which is usually not the case for longer time segments of electrophysiological data, but for the short time intervals that are investigated in this study, the data is assumed to be quasi-stationary. We used the freely available open source Matlab implementation of the `BioSig` Toolbox for biomedical signal processing (Schlögl and Brunner, 2008) for our analysis, which can be found at `http://biosig.sf.net/`.

There exists a number of criteria for estimating the optimal model order for each data set such as the Akaike Information Criterion (AIC; Akaike, 1974) or Schwarz's Bayesian Information Criterion (BIC; Schwarz, 1978) which try to estimate the optimal model order for the MVAR model.

Both criteria take the goodness of fit to the empirical data into account, but also penalize for increasing numbers of free parameters to avoid overfitting to the data. Note that smaller values indicate better model orders. Unfortunately, the optimal model order is usually not consistent for different criteria and different data sets. We tried to estimate the optimal model order (in the range between 1 and 50, which reflects the length of the 250 ms windows we used for the time-frequency analysis) by using these measures, but the results did not show consistent local minima and qualitatively decreased with increasing model order instead (see **Figure 1**). We compared models of order 20 and 40 for the 250 ms windows and found the resulting average power spectra and couplings to be qualitatively consistent. Therefore, we used a model order P of 20 for every data set, which corresponds to a time window of 100 ms given the sampling frequency of 200 Hz. This model order reflects a tradeoff between spectral resolution (specifically, we make clear that the model order does not determine the spectral resolution, which is in fact infinite, but instead it determines the number of observed frequency components for each pair of channels, which is $P/2$, and relates to the "frequency resolution" in this sense; Schlögl and Supp, 2006) and overparametrization and approximately corresponds to the model orders used in similar approaches. For example, Brovelli et al. (2004) used a model order of 10 (corresponding to a 50 ms window) for analyzing beta oscillations, Supp et al. (2007) revealed couplings in the gamma frequency range using a model order of 15 (30 ms), and Kayser and Logothetis (2009) and Anderson et al. (2009) studied oscillations including the theta range using model orders of 6 (60 ms) and 17 (85 ms), respectively. Additionally, this model order fulfills all the requirements stated in Schlögl and Supp (2006) to obtain a sufficient model of the data. Furthermore, one should note that slight changes in the model order do not lead to arbitrarily large changes in the prediction error, but it is still an important parameter for the correct estimation of the couplings (Schlögl, 2000).

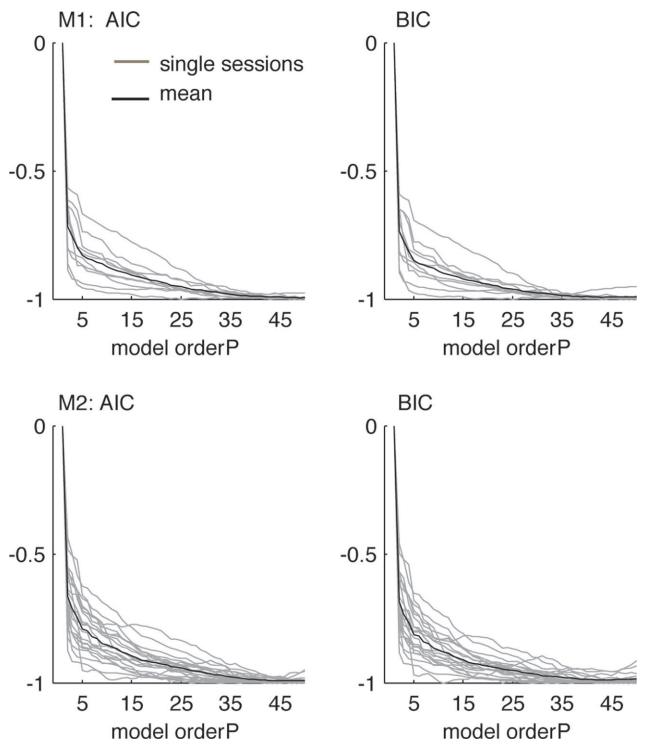


FIGURE 1 | Evaluation of model orders using Akaike Information Criterion (left, AIC) and Schwarz's Bayesian Information Criterion (right, BIC), normalized between maxima and minima of each session for the 1-s data from the delay condition. Gray lines indicate single sessions, black lines correspond to the average over all sessions. Criteria did not show consistent local minima, but qualitatively decreased with increasing model order up to $P = 50$. Smaller values indicate better model orders. For the data from both animals and all sessions, the model order $P = 20$ was chosen as a tradeoff between frequency resolution and overparametrization.

GENERALIZED PARTIAL DIRECTED COHERENCE

As mentioned earlier, we used generalized partial directed coherence (GPDC; Baccala et al., 2007) for our analysis, which is a slightly adapted version of PDC with better variance stabilization properties. Analysis of the validity of this coupling measure using simulated and real data for which the ground truth is known as well as a comparison to DTF and other measures can be found elsewhere (Baccala and Sameshima, 2001; Kus et al., 2004; Pereda et al., 2005; Gourevitch et al., 2006; Porcaro et al., 2009). Moreover, Porcaro et al. (2009) indicated that PDC is the most suitable method for this kind of analysis based on their results on MEG data.

Generalized partial directed coherence is derived by first transforming the MVAR model from the time domain into the frequency domain to obtain the frequency representation of the model parameters:

$$\mathbf{A}(f) = \boldsymbol{I} - \sum_{p=1}^{P} \mathbf{A}_p \mathrm{e}^{-2\pi i p(f/f_s)},$$

where $\boldsymbol{I}$ refers to the M-dimensional identity matrix and f_s is the sampling frequency. Note that in this equation, $i^2 = -1$.

Then, GPDC_{ij} (which reflects the coupling from channel j to channel i) is calculated to be:

$$\mathrm{GPDC}_{ij}(f) = \frac{\frac{1}{\sigma_i}|A_{ij}(f)|}{\sqrt{\sum_{k=1}^{M}\frac{1}{\sigma_k^2}|\mathbf{A}_{kj}(f)|^2}},$$

where σ_i^2 refers to the variance of the innovation process $x_i(t)$. GPDC_{ij} is normalized in the interval [0, 1], with increasing values for stronger interactions at particular frequencies, and sums up to one for each frequency component over all destination channels including the channel itself. The idea is to calculate the degree of influence of channel j to channel i with respect to the total influence of j on all channels. Note that this normalization procedure of (G)PDC was recently criticized (Schelter, 2009) because of some difficulties in comparing interaction strengths for different

frequencies. As the values $A_{ij}(f)$ and $A_{ji}(f)$ are not necessarily identical, directionality of coupling is obtained. As $GPDC_{ij}$ has to be interpreted as the remaining amount of coupling that can not be assigned to the influence on other channels, we excluded self-coupling of channel *j* to itself for the subsequent analysis.

EXPERIMENTAL TASK

Two adult male rhesus monkeys (*Macacca mulatta*) participated in the experiments. All studies were approved by local authorities and were in full compliance with applicable guidelines (EUVD 86/609/EEC) for the care and use of laboratory animals. The behavioral task of the monkeys was a delayed matching to sample task. The monkey was seated in front of a screen at a distance of approximately 110 cm. An initial tone indicated the potential start of a trial. The monkey initiated a trial-start by grasping a lever and fixating on a small fixation spot on the center of the screen (baseline period). After 1500 ms, a first stimulus appeared on the screen for 250 ms, the so-called sample stimulus. As sample stimuli we used different natural images. The stimuli that were used in all of the experiments were chosen from the Corel-Photo-CD "Corel Professional Photos" comprising a collection of natural images showing birds, flowers, monkeys and butterflies in their natural surroundings. The images used in this study were randomly selected. All images were manipulated by Fourier techniques that have been described in detail elsewhere (Liebe et al., 2009). The sample stimulus was followed by a delay period of 1500 ms during which the monkey held fixation. After the delay, a second stimulus, the so-called test stimulus, was presented. The monkeys were rewarded for a lever release whenever the test stimulus matched the sample stimulus. Whenever the test stimulus did not match the sample, the monkeys' task was to withhold the lever release until, after a brief delay of 200 ms, a second test stimulus appeared, that always matched the sample. This procedure ensured that the monkey had to initiate a behavioral response on every trial. The monkeys were rewarded with juice for every correct trial. Within one session, the different trial types were randomly interleaved. Stimuli were $7° \times 7°$ in size, at 24-bit color depth, and presented at the center of gaze on a 21″ monitor (ViewSonic P810) with linear luminance response as well as linear response at separate color channels (gamma corrected).

ELECTROPHYSIOLOGY

Local field potentials were recorded from recording chambers placed on the surface of the skull based on stereotaxic coordinates allowing vertical access to the dorsal region of extrastriate area V4. The Hoarsley–Clark coordinates for the center of the recording chambers for monkey 1 were AP: −6.5, ML: −29.7. For monkey 2 the chamber coordinates were AP: −5.2, ML: −29.9. The implantation as well as surgical procedures used are described in detail in Lee et al. (2005). Neural signals were measured using two custom made micro drives mounted on a plastic grid (Crist Instruments, Hagerstown, MD, USA). In each recording session 4–6 tungsten microelectrodes (UEWLGDSMNN1E, FHC Inc., Bowdoinham, ME, USA) were manually lowered down into the cortex in pairs with a minimal separation between electrodes of 0.5 mm. The impedance of the microelectrodes was approximately 1 MΩ. The signal from each electrode was preamplified (factor 20, Thomas Recording, Giessen, Germany) using the recording chamber as the external reference. The analog signal was then filtered and amplified (BAK electronics, Germantown, MD, USA) to extract the LFP responses. After an additional waiting period of at least 1 h we started the recordings. The LFP was obtained by band-pass filtering the signal between 0.1 and 300 Hz and digitizing with a sampling rate of 4464 Hz. One unit of the analog-to-digital converter corresponds to 5 μV.

We recorded LFP activity from 44 channels in 10 sessions from monkey 1 and 86 channels in 20 sessions from monkey 2. This resulted in 202 channel pairs for monkey 1 and 398 channel pairs for monkey 2. For each monkey, the minimum number of channels per session was 3, the maximum number of channels was 6. The spatial distribution of all recorded channels for monkey 1 and 2 can be found in **Figure 2**. For each recording site, its location is defined by two dimensions (anterior to posterior, and medial to lateral) based on the recording grid placed within the recording chamber. In order to measure coupling as a function of distance between recording sites, we calculated the Euclidian distance between two sites based on their respective locations along the two dimensions. The minimal distance between sites was 0.5 mm (i.e., sites directly neighboring each other within the grid), the maximal distance we obtained was 4 mm.

STATISTICAL ANALYSIS

In order to be able to calculate confidence intervals that can be used to evaluate the significance of differences in coupling between different time intervals, we used a bootstrapping procedure that samples with replacement from the original trial set in order to generate bootstrap samples of the same size as the original data, but with different subsets of trials in them (Efron and Tibshirani, 1993). For each regarded data set of 1 s (last second of baseline and delay period), a set of 1000 bootstrap samples was generated. These bootstrap samples were then independently used to calculate the MVAR models as stated above and to estimate the couplings between the simultaneously recorded LFP channels with their respective confidence intervals. Change in coupling was considered significant if both the 0.01st and 99.9th percentile of the bootstrap distribution was above (increase) or below (decrease) the average baseline level.

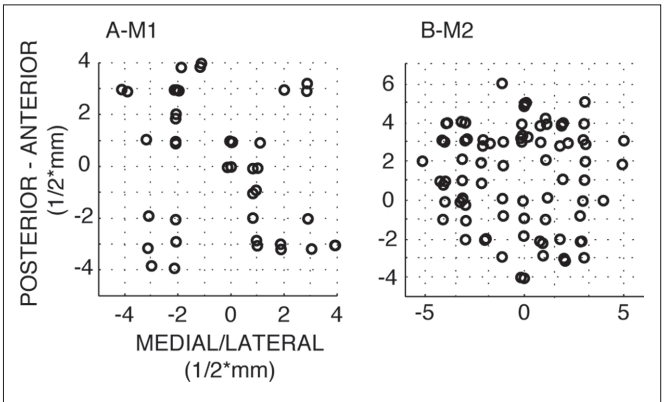


FIGURE 2 | Recording locations along the medial–lateral and anterior posterior direction for monkey 1 and 2. 1 unit corresponds to 1/2 mm. Note that symbols are slightly jittered at their recording locations for better visualization.

Significance of coupling strength compared to the hypothesis that there was no coupling at all was assessed using a shuffling procedure. For each recording channel, trials were independently permuted repeatedly to obtain 1000 shuffled samples. MVAR model estimation was then also applied to these data sets.

For assessing the statistical significance of the effects of coupling as a function of distance between recording sites, we used a shuffling procedure that randomly shuffles the coupling values over distances to obtain 10^4 shuffled samples. Statistical significance of the real rank correlation was then calculated with respect to this distribution.

RESULTS

POWER SPECTRA

First, we examined the frequency content of induced oscillations during different periods of the visual memory task. Previously it had been found that there is enhanced power in the theta band during the delay period of the task in V4 (Rainer et al., 2004; Lee et al., 2005). We first sought to confirm these findings and compared the power spectrum for the delay period (i.e., across last 1000 ms before the onset of the test stimulus) to the power spectrum obtained from the 1000 ms time interval preceding the onset of the sample stimulus ("baseline"). **Figure 3** shows the median amplitude spectra of LFP activity across all recorded channels (**Figure 3A**) derived using a Morlet wavelet based approach (Tallon-Baudry and Bertrand, 1999; Graimann and Pfurtscheller, 2006). In both monkeys, the power spectra showed a local peak in the theta-frequency range during the delay period (black) which is absent during the pre-stimulus baseline period (gray). **Figure 3B** shows the distribution of power at the peak frequency within the theta range for the baseline vs. delay period and illustrates a significant change in theta power during the delay compared to the baseline.

GENERAL COUPLING ANALYSIS

Based on the occurrence of enhanced theta power during the delay period, we analyzed coupling strength between the different recording sites using GPDC obtained from MVAR modeling. We were interested in whether the enhanced theta power we observed during the delay period of the task coincides with directed coupling in the theta band. Thus, we first examined GPDC coupling as a function of

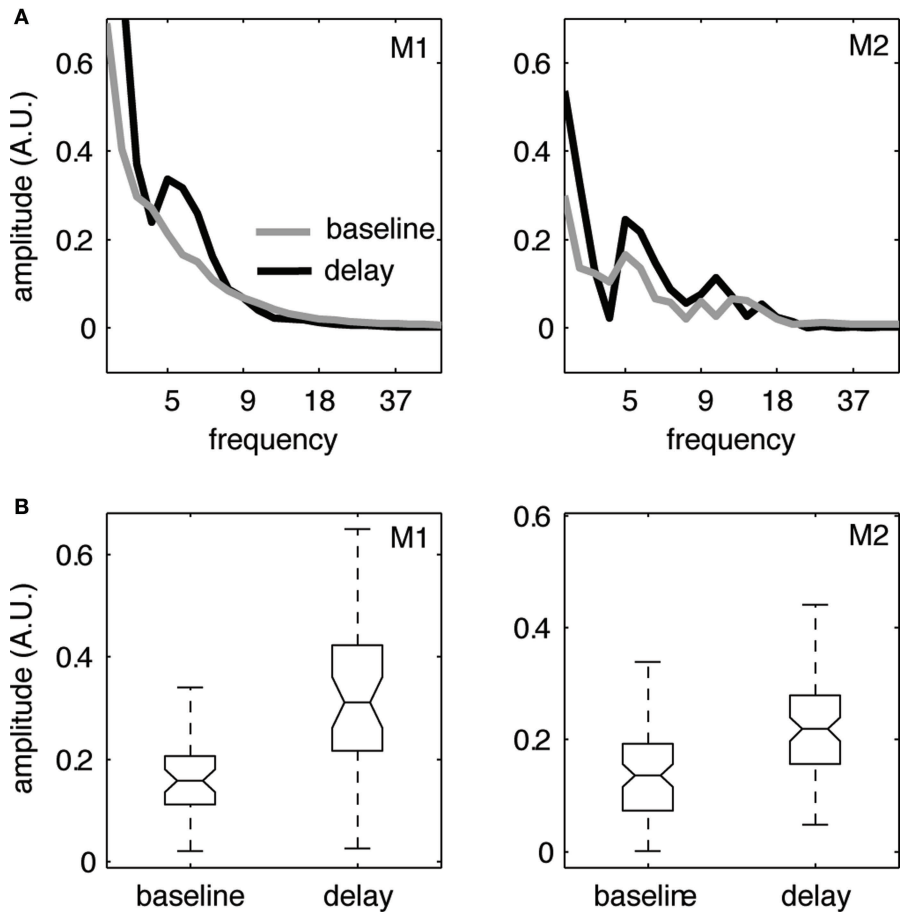


FIGURE 3 | Enhanced power in the theta band during the delay period of the task. (A) Amplitude spectrum of LFP activity (median across all recorded channels in V4, $N = 44/86$ for monkeys 1/2, respectively) during the pre-stimulus baseline period (gray) and the delay period (black) of the delayed matching to sample task for frequencies from 1 to 40 Hz. We find enhanced power in the theta band (4–10 Hz) during the delay period for both monkeys. **(B)** Boxplots showing the distribution of power at the peak frequency within the theta range for the baseline vs. delay period. Both monkeys show a significant increase in theta power from baseline to delay (Wilcoxon signed rank test $Z = 6.95/4.47$, $p < 0.01$ for monkey 1/2, respectively).

frequency during delay (see **Figure 4A**). Similar to the power spectra, we observed local peaks in GPDC coupling within the theta range (3–12 Hz) for both monkeys, albeit at slightly different frequencies. For monkey 1 the average peak frequency for highest coupling within the theta range was 4.33 ± 3.6 Hz (mean across sessions, ±1 SD) and was located well within the range of the maximum power peak frequency at 5.86 ± 0.81 Hz (see **Figure 3**). For the second monkey the average peak frequency for highest coupling was larger (8.35 ± 3.8 Hz), and also higher than the average peak of power (5.51 ± 1.77 Hz), but not significantly higher ($p > 0.05$). Similarly, inspection of **Figure 3** shows that although the peak in power for monkey 2 is around 5 Hz, we find elevated power during the delay up to 10 Hz. Thus, the peak frequencies at theta power and theta coupling were overall similar. Note that these and subsequent results are based on the models that were fitted to 1 s time intervals in baseline and delay conditions (equivalent time intervals as for power spectra).

Subsequently, we assessed changes in coupling between the baseline and the delay period for several frequency bands that have been traditionally implicated in the interaction of oscillatory components during sensation and cognition and also follow conventional definitions of theta and beta bands [theta (3–12 Hz), beta (20–35 Hz), gamma (40–80 Hz); Buzsaki, 2006]. **Figure 4B** shows the median absolute difference in coupling between baseline and delay period (left). The graph shows that the degree of change significantly decreases with increasing frequencies with the largest coupling change occurring in the theta band. Likewise, the proportion of pairs that show significant changes in coupling is highest in the theta range compared to the other frequency bands (right). In the theta band, 116 of 202 pairs showed significant changes in coupling (57%, $p < 0.001$) in monkey 1, in monkey 2 235 pairs showed significant changes in coupling (59%, $p < 0.001$). In both monkeys we found significant increases as well as decreases in

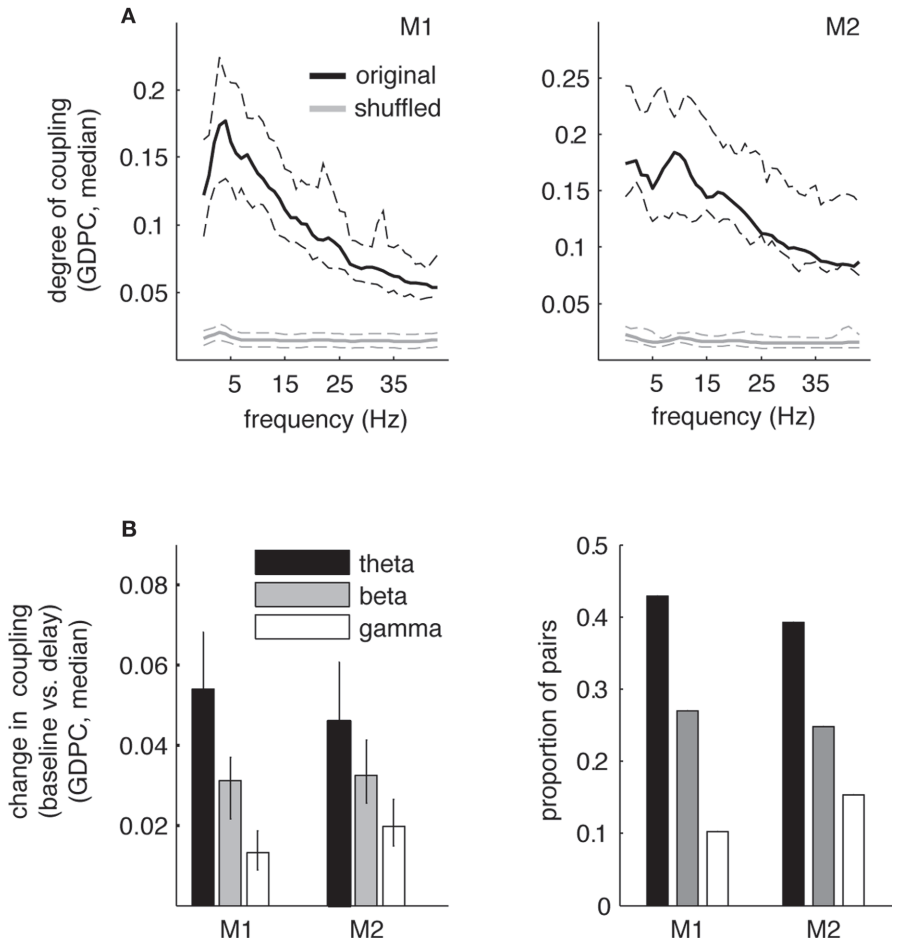


FIGURE 4 | Generalized partial directed coherence coupling in theta band during the delay period of the task. (A) Median degree of coupling (GPDC) as a function of frequency during the delay period (black) across all pairs in V4 for monkeys 1 and 2 for non-shuffled (black) and shuffled pairs. For both monkeys we find a peak in coupling in the theta-frequency range (3–12 Hz), although peak coupling occurred at slightly different frequencies (around 5 Hz for monkey 1, around 9 Hz for monkey 2). Dashed lines correspond to ±34th percentiles of values around the median. Note that for the shuffled data, overall coupling was found to be around 0.02 for the whole frequency range. **(B)** Median absolute difference in degree of coupling (GPDC) (left) and proportion of pairs showing a significant change ($p < 0.001$) in coupling between pre-stimulus baseline and delay period for different frequency bands (right). Error bars correspond to ±34th percentiles around the median. In both monkeys, the median change in coupling between baseline and delay is highest in the theta band (non-parametric ANOVA Kruskal–Wallis test, $\chi^2 = 11.4/4.23$, $p < 0.01$). Likewise, the proportion of pairs showing significant changes in coupling is also highest in the theta band compared to the other frequency bands (χ^2 test for comparison of proportions, $\chi^2 > 8.2$, $p < 0.001$ for all comparisons).

theta coupling during the delay when compared to the baseline. Specifically, in monkey 1 74 pairs showed significant increases, and 42 pairs showed significant decreases in coupling. In contrast, in monkey 2 89 pairs showed increases and 146 pairs decreases. Thus, monkey 1 shows significantly more increases than monkey 2 and vice versa ($\chi^2 = 18.5$, $p < 0.01$). One factor that might contribute to this difference is the different distribution of electrode spacing between the animals, with monkey 1 showing significantly larger distances between electrodes than monkey 2 (mean [median] distance 4.6 [4]/3.5 [3] for monkey 1/2, respectively; ranksum-test, $Z = 5.23$, $p < 0.01$). This is supported by several facts. First, for the smallest distance between electrodes, i.e., the distance that is identical and therefore comparable between the animals (unit 1, or 0.5 mm), the proportion of increases vs. decreases is similar between the monkeys, i.e., statistically identical (50%/27% increases, $Z = 3.6$, $p > 0.05$). Second, for the smallest distance we find an identical proportion of increases and decreases (i.e., 50/50) in monkey 1. Third, the proportion of significant decreases is slightly enhanced for smaller distances (50% at distance 1 vs. 20% at distance 4 for monkey 1, and 72% vs. 53% for distances 1 and 3 for monkey 2) and likewise the proportion of increases reduced at smaller distances. As the distances between electrodes are significantly lower in monkey 2 compared to monkey 1, the percentage of decreases should be higher in monkey 2, and vice versa. Ultimately, due to the limitations in spatial sampling, the differences in spatial configuration can only give an indication of why we find differences in the proportion of significantly increased vs. decreased coupling between the monkeys. In summary, our findings demonstrate that significant directed interactions between LFP within V4 during visual memory predominantly occur in the theta-frequency range and the frequencies at which highest coupling occurs are comparable to the frequency range of power increases during the delay period. Based on these results we further investigated the time course and directionality of theta coupling during the delay period.

TIME COURSE AND DIRECTIONALITY OF COUPLING

To illustrate the time course of theta coupling during the task, we used moving windows comprising time intervals of 250 ms (with an overlap of 200 ms) and fitted MVAR models to these individual windows. **Figure 5** shows representative time courses of theta coupling in single recording pairs as well as the time course of coupling in the opposite direction (left/right graphs, respectively). These examples represent channel pairs with a significant ($p < 0.001$) increase or decrease in GPDC during the delay period compared to the baseline period and were chosen based on the previous analyses using coupling measures obtained from 1 s windows (see also Materials and Methods).

In all examples, theta increases and decreases occur shortly after the offset of the sample stimulus and are sustained throughout the entire 1500 ms long delay period. Interestingly, in all selected pairs we find differences in coupling strength and even opposing effects between pair directions, for example a significant increase in theta coupling in one, and a significant decrease in theta coupling in the opposite direction (see graph B example for monkey 1). To investigate this asymmetry across all channel pairs in more detail, we first computed the median coupling across all pairs showing significant increases or decreases in the theta band. We subsequently selected all pairs with the respective opposite direction and computed the median coupling across these pairs. The reasoning behind the procedure is as follows: if all channels show significant changes in theta coupling in both directions (i.e., from channel X to channel

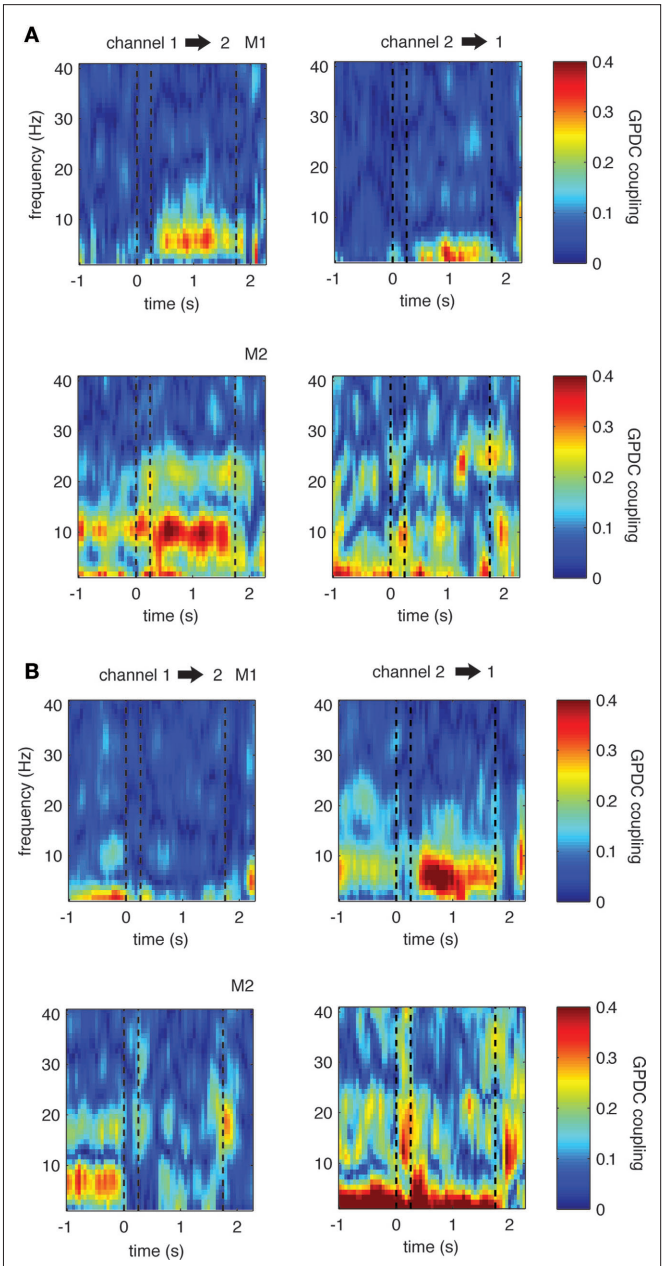


FIGURE 5 | Single example LFP channel pairs showing significant changes in coupling during the delay period compared to baseline. **(A)** One example for each monkey in which there is significant increase in coupling in the direction from channel 1 to 2 (left). **(B)** Two examples in which there is significant decrease in theta in one direction (left). The opposite directions show less strong or even opposing trends, indicating that coupling is not symmetric between sites. Dashed lines represent the on- and offset of the sample stimulus, as well as the onset of the first test stimulus during the trials, from left to right, respectively. This convention is also used in the subsequent figures.

Y and from channel Y to channel X), each channel pair will be represented in both groups. Consequently the median coupling across the channel pairs would be the same.

However, this is not the case. **Figure 6** displays the resulting median coupling strength over all site pairs showing significant couplings within the theta range (left) and their respective opposite direction (right): for both monkeys we find asymmetrical, i.e., more unidirectional increases and decreases in theta coupling during the delay. This result is further illustrated in **Figure 7** that shows the ratio of median coupling between pairs of channels and their opposite directions separately for each monkey (**Figures 7A,B**). Furthermore, **Figure 7C** shows coupling strength in the delay for sites with significant changes from baseline to delay in the theta band vs. the couplings in the opposite direction. Similarly to the observed asymmetries in coupling values, significant proportions of pairs (64/48% monkey 1/2, $Z > 15.4$, $p < 0.001$) show significant *increases* in one direction only and significant proportions of pairs (57/58% monkey 1/2, $Z > 18.6$, $p < 0.001$) of the pairs show significant *decreases* in only one direction. Overall, in three out of four cases, the majority of pairs showed significant changes of coupling in one channel pair direction, but not the other.

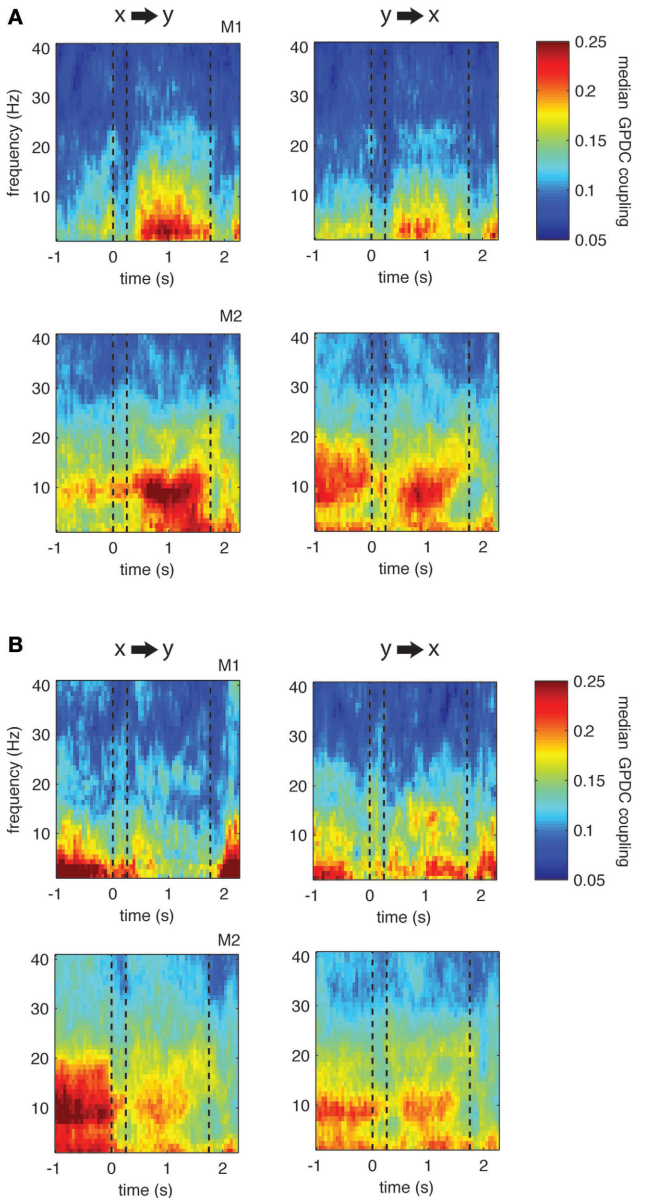


FIGURE 6 | Grand-Median directed coupling across all pairs showing significant increase (A) or decrease (B) during delay compared to baseline ($p < 0.001$, left column) as well as the median coupling for the opposite direction of channel pairs (right column). Note that if there is significant increase (decrease) in both directions (i.e., from *x* to *y* and vice versa), both channel pair directions will contribute to the median for both directions of interaction. Otherwise, if all pairs would show significant increase (decrease) in both directions, left and right plots would be identical.

Taken together, these findings illustrate that theta coupling during the delay is not symmetric between channel pairs and provide evidence for a complex interaction involving both directionally dependent increases and decreases in coupling during visual memory. In the following we examine a different aspect of these coupling phenomena, namely their dependence on the spatial layout of the different oscillatory components. Our recording setup allowed us to simultaneously measure the activity of up to 6 LFP electrodes that were spatially distributed across a cortical surface area of approximately 6 mm × 6 mm. Therefore, within one session, electrode locations varied in spatial position and distance to each other.

RELATION OF COUPLING STRENGTH AND DISTANCE BETWEEN RECORDING SITES

Figure 8 illustrates the dependence of absolute directed coupling and changes of coupling on the distance between electrodes, with higher direct coupling occurring at lower distances (both monkeys: $\rho_s = -0.52/-0.32$, $p < 0.0001$). Similar effects were found for the changes in coupling (i.e., decrease M1: rank correlation coefficient $\rho_s = -0.24$, $p < 0.05$, M2: $\sigma_s = -0.31$, $p < 0.01$ and increase M1: $\rho_s = -0.12$, $p = 0.1$, M2: $\sigma_s = -0.2$, $p < 0.05$) during delay with respect to the baseline). Note that the decrease of change in coupling with higher distance in monkey 1 does not reach a significance level of $p < 0.05$ for increases in coupling, but is at trend level.

Our results indicate that both the strength of coupling and the change in coupling from the baseline to the delay condition are stronger for smaller distances between site pairs. This dependence could be found despite the differences in electrode spacing between the animals (see also General Coupling Analysis). Thus, not only absolute coupling but also the dynamic changes in coupling are a local phenomenon within the neural network. Our findings are consistent with earlier reports for example from V1 recordings of in the macaque showing that pairwise spectral coherence in LFP activity between electrodes decreases as a function of receptive field distance (which is related to spatial distance; Frien and Eckhorn, 2000) and from recordings of several sites of the human cortex (Raghavachari et al., 2006) and extend these previous results using directed coupling measures.

DISCUSSION

Oscillatory activity in neural networks as measured by EEG or LFP recordings is a widespread phenomenon of neural behavior and is thought to arise from the synchronous activity of

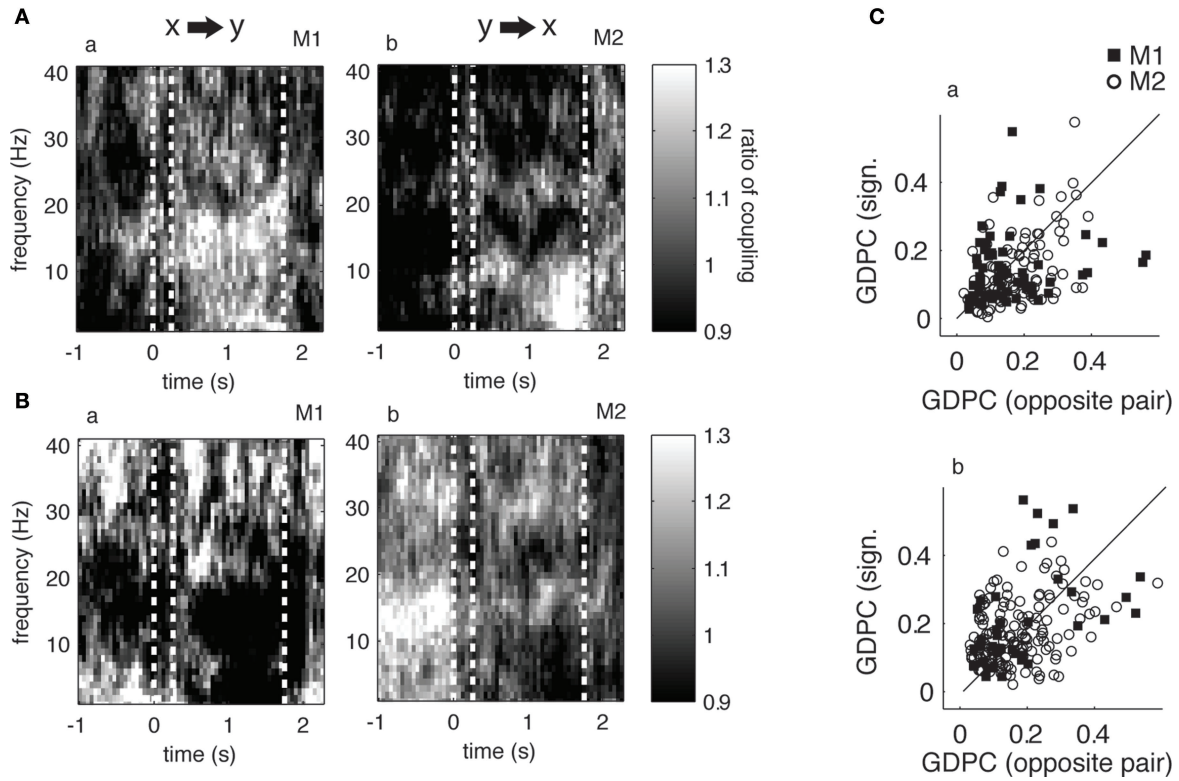


FIGURE 7 | Ratio of median coupling between pairs showing significantly enhanced (A) and reduced (B) directed coupling during baseline and their opposite pairs (i.e., coupling in opposite direction). Plots a–b show the ratio of median coupling across the time course of the trial for different frequencies for monkey 1 and 2, respectively. A ratio larger (or smaller) than one indicates a stronger median coupling in one than the opposite direction, thus the ratio is an indicator of how "directed" or "symmetric" the coupling is between pairs of channels. Graphs **(C)** plot the degree of coupling of significant pairs and their opposite directions (*y*-axis pairs chosen based on sign difference between delay and baseline, *x*-axis shows opposite direction) in the theta band. If the degree of coupling is not symmetric between directions, data points deviate from the diagonal. Also note that if there was a significant increase (decrease, respectively) in both directions for a single pair, two data points are shown (one below, one above the diagonal if the respective increase/decrease is not fully symmetric).

neuronal populations at various spatial and temporal scales (Salinas and Sejnowski, 2001; Buzsaki and Draguhn, 2004; Fries, 2005). In many studies it has been shown that oscillations in different frequency bands are important for neural computations. Synchronous activity can, for example, establish and support temporal relationships between different elements within a neural network depending on context, stimulus or behavioral state (Tallon-Baudry, 2009; Uhlhaas et al., 2009) or represent information about sensory events that can not be inferred from spiking activity alone (Montemurro et al., 2008). Thus, oscillatory activity serves the precisely timed cooperation between neural ensembles and could also provide temporal windows that allow for the selective routing and gating of information in an efficient manner (Mizuseki et al., 2000; Fries et al., 2001; Salinas and Sejnowski, 2001). Another important characteristic of neuronal oscillations is that oscillations at different frequencies are thought to subserve different behavioral and cognitive functions. Prominent examples are the involvement of gamma oscillations (>40 Hz) in visual and attention-related processes (Fries et al., 2001; Keil et al., 2001), the role of beta oscillations (15–35 Hz) in sensorimotor tasks (Murthy and Fetz, 1992) and the importance of theta oscillations in memory-related processing (Okeefe, 1993; Rainer et al., 2004; Lee et al., 2005; Raghavachari et al., 2006).

The quantification of neural synchrony has traditionally been carried out using measures that assess the pairwise and instantaneous correlation in either amplitude or phase between two neural signals, for example using cross-correlation analysis between spike trains of multiple neurons (Aertsen and Arndt, 1989), spike-field coherence between the spiking activity of neurons and LFP activity (Fries et al., 2001; Pesaran et al., 2002) or phase-locking analysis of simultaneously recorded LFP or EEG data (Lachaux et al., 1999, 2000). However, despite the fact that these measures assess the strength with which two neural processes are coupled, they fail to provide information on several aspects of synchronization that can be important to fully describe their interaction, for example the direction of coupling between neural elements.

Here, MVAR models have proven to be efficient tools for assessing the direction of coupling and can be more appropriate to capture the complexity of oscillatory dynamics as synchrony between neuronal ensembles changes across time or behavioral conditions. However, it is important to note that unipolar signals are vulnerable to volume conducted far-field effects and issues related to the usage of a common reference against which all differences in electrical potential are measured. Both factors might lead to adversely affected measures of coupling strengths, and elaborate methods for resolving these issues completely (besides using bipolar signals) need still be found

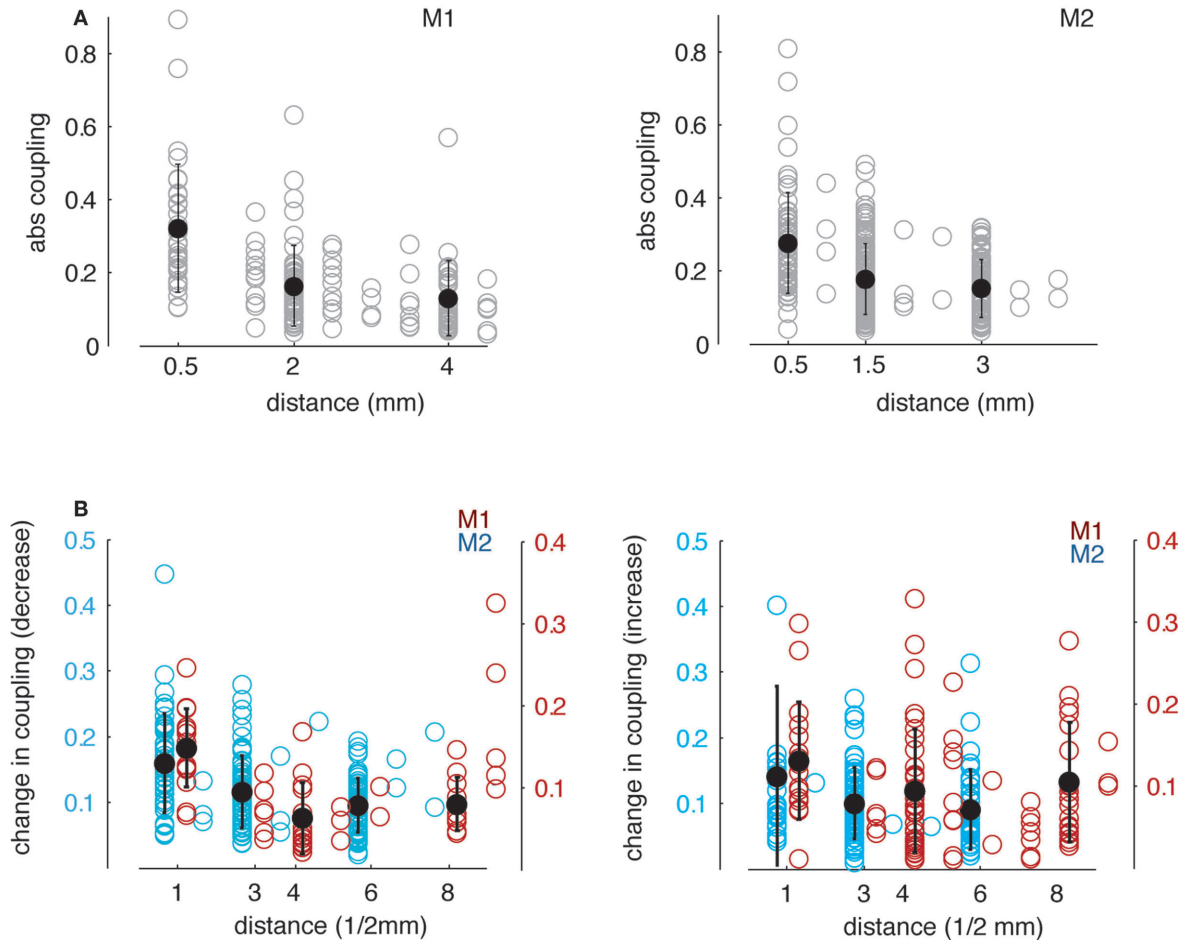


FIGURE 8 | Dependence of GPDC in theta band on distance between simultaneously recorded channels. (A) Absolute coupling as a function of Euclidian distance across all recorded pairs. Open symbols denote coupling values for single pairs, closed symbols represent the mean ± 1 SD across pairs within three bins. For both monkeys, absolute coupling decreases with increasing distance. **(B)** Dependence of increases and decreases in GPDC during delay relative to baseline in theta band on distance between simultaneously recorded channels. Absolute decrease (left) and increase (right) in coupling as a function of Euclidian distance across all recorded pairs showing decreases (left) or increases (right) in coupling. In both cases, changes in coupling during delay are higher for smaller distances. Note that we shifted the data points for M1 (blue) and M2 (red) slightly to the left and to the right, respectively for better visual discriminability.

(Schlögl and Supp, 2006; Bollimunta et al., 2009). Nevertheless, the application of coupling measures based on MVAR models has revealed important insights into neural interactions in many studies (Bressler et al., 1999; Brovelli et al., 2004; Supp et al., 2007; Kayser and Logothetis, 2009). For example, Brovelli et al. (2004) analyzed interaction patterns in monkey sensorimotor cortex and found unidirectional couplings from somatosensory areas to motor areas within the beta frequency range that might be used to control motor output. In a different study Kayser and Logothetis (2009) investigated interactions of monkey auditory and superior temporal cortices related to sensory integration and found that while interactions from auditory cortex to superior temporal regions prevail below 20 Hz, interactions in the other direction are more pronounced at frequencies above 20 Hz. A third example are the findings of the study by Supp et al. (2007) in which the authors demonstrated that visual processing of familiar and unfamiliar objects engages different cortical networks at different degrees of directionality via interactions in the gamma frequency range. In all these studies, MVAR models revealed insights into the directed spatio-temporal dynamics of multiple cortical areas during cognitive processing that went beyond the description of synchrony between these areas.

In our study we used MVAR models to provide a description of the directed coupling of theta oscillations during short-term memory. This oscillatory phenomenon has been described in a number of studies in relation to short-term memory processes both in humans and animals. There are mainly two lines of research that focus on the role of these oscillations for memory-related processes. A large set of studies provides strong evidence on a connection between theta oscillations and (especially spatial) memory for the rat hippocampus, revealing that the timing of spikes both within hippocampus and within regions like prefrontal cortex is strongly connected to the hippocampal theta rhythm (Okeefe, 1993; Mizuseki et al., 2000; Buzsaki and Draguhn, 2004; Siapas et al., 2005). A second line of research has concentrated on the importance of theta oscillations for memory performance in primates with the focus on EEG and LFP recordings in various cortical areas. Using

EEG in human subjects, multiple studies have shown that there are increases as well as decreases in theta power that can depend on the specific nature of the memory task demand (Klimesch, 1999; Raghavachari et al., 2001, 2006). For example, Raghavachari et al. (2001, 2006) showed an increase in theta power during the delay period of a memory task that co-varied with delay period length in multiple cortical regions, including frontal, temporal and occipital areas. Very recently, one study also incorporated MVAR modeling to reveal directional influences between different cortical regions (Anderson et al., 2009), providing evidence for memory-related theta-frequency interactions between prefrontal and medial temporal sites in the human brain. In contrast to human studies, only few studies have investigated the relation between theta oscillations and visual memory in the non-human primate. Here, research has focused on the extrastriate visual area V4, in which theta oscillations occur during the memory phase of the task, are modulated by task difficulty (Rainer et al., 2004) and are involved in the coding of visual stimuli during visual memory (Lee et al., 2005). In summary, research on theta synchrony during short-term memory has thus far provided evidence for the hypothesis that memory processing is accompanied by increased theta power and synchrony.

However, as previously mentioned, measures of directed coupling based on MVAR modeling have shown to be useful for investigating the complex interaction patterns in LFP during cognitive processing. Thus we applied these coupling measures to analyze neural interactions in the theta band during visual short-term memory within V4. Our analyses firstly confirmed earlier results, showing enhanced theta power during the delay period. Using the coupling measures based on MVAR models, we additionally found increases as well as decreases in coupling between recording sites in the delay period with respect to the baseline period that were most prominent in the theta band. This was evident in the coupling value estimates as well as in the proportion of site pairs showing significant changes in coupling. More importantly, however, we showed that these changes in coupling tend to be asymmetric between sites, i.e., they depend on the considered direction between site pairs. This finding suggests that not the mere occurrence of oscillatory activity or coherence in the theta band correlates with memory processing. Instead, the selective and direction-dependent change in theta coupling, which ultimately represents a change in functional connectivity within the neural circuit, plays an important role in this process. Our results on the asymmetrical nature of directed interaction during memory also favors the hypothesis that theta oscillations and therefore coupling arise locally within the V4 network. In contrast, if coupling would be a phenomenon due to common input, one would expect bidirectional coupling with similar strength in both directions if the data from the neural elements producing the common input is not incorporated by the model. When interpreting the results from MVAR models, one should keep in mind that the model is based on data from only a small subset of the whole neural system (Stevenson et al., 2008). In addition, we were able to confirm earlier work on the relations of interaction strength and spatial distance that showed decreasing coherence of signals with increasing distance between sites and extended their results by measuring direct causal interactions instead of coherence (Frien and Eckhorn, 2000; Raghavachari et al., 2006).

Taken together, these effects would not have been revealed using more traditional methods that incorporate only phase synchronization or coherence. Therefore, our work clearly shows the advantage of using directed coupling measures based on MVAR models for studying functional connectivity patterns within the brain and highlights the importance of direction-dependent modulations of local interactions between neural populations for studying sensory and cognitive processing.

Finally, we would like to point out that while the methods that we applied provide important insights into the functional connectivity patterns within the brain, their power is still limited because they can only assess linear interactions. While some extensions to nonlinear MVAR models (together with all the issues of nonlinear optimization) have been proposed (Pereda et al., 2005; Sun, 2008; Jachan et al., 2009), there is still work to be done to further improve these methods. In addition, our findings provide only the elementary description of the pattern of interaction between different oscillatory processes during visual memory. Further investigation will be needed to assess the specific role of directed coupling in relation to various cognitive parameters, for example task difficulty, performance or memory load. In addition, MVAR analysis can also be used to assess the interactions of LFP and spiking activity if the spike trains are properly preprocessed for this purpose. Here, it seems to be interesting to see how oscillations at the level of the LFP exert an influence on neuronal firing directly measured from the spiking activity of single neurons.

ACKNOWLEDGMENTS

Written under partial support by the Max Planck Society, the Austrian Science Fund FWF #S9102-N13 and projects #FP7-506778 (PASCAL2) and # FP7-231267 (ORGANIC) of the European Union. We would like to thank Jakob Macke for helpful discussions and useful comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/computationalneuroscience/paper/10.3389/fncom.2010.00014/

REFERENCES

Aertsen, A., and Arndt, M. (1989). Response synchronization in the visual cortex. *Curr. Opin. Neurobiol.* 3, 586–594.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723.

Anderson, K. L., Rajagovindan, R., Ghacibeh, G. A., Meador, K. J., and Ding, M. (2009). Theta oscillations mediate interaction between prefrontal cortex and medial temporal lobe in human memory. *Cereb. Cortex* (in press), doi: 10.1093/cercor/bhp223

Baccala, L., and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* 84, 463–474.

Baccala, L., Sameshima, K., and Takahashi, D. (2007). *Generalized Partial Directed Coherence*. 15th International Conference on Digital Signal Processing, 2007, Cardiff, 163–166.

Bollimunta, A., Chen, Y., Schroeder, C., and Ding, M. (2008). Neuronal mechanisms of cortical alpha oscillations in awake-behaving macaques. *J. Neurosci.* 28, 9976–9988.

Bollimunta, A., Chen, Y., Schroeder, C., and Ding, M. (2009). Characterizing oscillatory cortical networks with Granger causality. in *Coherent Behavior in Neuronal Networks*, eds K. Josić, J. Rubin, M. Matías, and R. Romo (New York: Springer), 169–189. doi: 10.1007/978-1-4419-0389-1_9.

Bressler, S. L., Ding, M., and Yang, W. (1999). Investigation of cooperative cortical dynamics by multivariate autoregressive modeling of

event-related local field potentials. *Neurocomputing* 26–27, 625–631.

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, 3875–3885.

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. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9849–9854.

Buzsaki, G. (2005). Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus* 15, 827–840.

Buzsaki, G. (2006). *Rhythms of the Brain*. New York: Oxford University Press.

Buzsaki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 204, 1926–1929.

Chen, Y., Bressler, S. L., and Ding, M. (2006). Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *J. Neurosci. Methods* 150, 228–237.

Cui, J., Xu, L., Bressler, S., M., D., and Liang, H. (2008). BSMART: a Matlab/C toolbox for analysis of multichannel neural time series. *Neural Netw.* 21, 1094–1104.

Ding, M., Bressler, S. L., Yang, W., and Liang, H. (2000). Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modelling: data preprocessing, model validation, and variability assessment. *Biol. Cybern.* 83, 35–45.

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Frien, A., and Eckhorn, R. (2000). Functional coupling shows stronger stimulus dependency for fast oscillations than for low frequency components in striate cortex of awake monkey. *Eur. J. Neurosci.* 12, 1466–1478.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480.

Fries, P., Reynolds, G., Rorie, A., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291, 1560–1563.

Gail, A., Brinksmeyer, H., and Eckhorn, R. (2004). Perception-related modulations of local field potential power and coherence in primary visual cortex of the awake monkey during binocular rivalry. *Cereb. Cortex* 14, 300–313.

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

Graimann, B., and Pfurtscheller, G. (2006). Quantification and visualization of event-related changes in oscillatory brain activity in the time-frequency domain. *Prog. Brain Res.* 159, 79–97.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.

Jachan, M., Henschel, K., Nawrath, J., Schad, A., Timmer, J., and Schelter, B. (2009). Inferring direct directed-information flow from multivariate nonlinear time series. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 80. 011138-1 – 011138-5. doi: 10.1103/PhysRevE.80.011138

Kaminski, M., and Blinowska, K. (1991). A new method of the description of the information flow in brain structures. *Biol. Cybern.* 65, 203–210.

Kayser, C., and Logothetis, N. K. (2009). Directed interactions between auditory and superior temporal cortices and their role in sensory integration. *Front. Integr. Neurosci.* 3:7. doi: 10.3389/neuro.07.007.2009.

Keil, A., Gruber, T., and Müller, M. (2001). Functional correlates of macroscopic high-frequency brain activity in the human visual system. *Neurosci. Biobehav. Rev.* 25, 527–534.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195.

Kus, R., Kaminski, M., and Blinowska, K. J. (2004). Determination of EEG activity propagation: pair-wise versus multichannel estimate. *IEEE Trans. Biomed. Eng.* 51, 1501–1510.

Lachaux, J.-P., Rodriguez, E., Le Van Quyen, M., Lutz, A., Martinerie, J., and Varela, F. J. (2000). Studying single-trials of phase synchronous activity in the brain. *Int. J. Bifurcat. Chaos* 10, 2429–2439.

Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208.

Lee, H., Simpson, G., Logothetis, N., and Rainer, G. (2005). Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron* 45, 147–156.

Liang, H., Bressler, S. L., Ding, M., Desimone, R., and Fries, P. (2003). Temporal dynamics of attention-modulated neuronal synchronization in macaque v4. *Neurocomputing* 52–54, 481–487.

Liang, H., Ding, M., and Bressler, S. L. (2000). On the tracking of dynamic functional relations in monkey cerebral cortex. *Neurocomputing* 32–33, 891–896.

Liang, H., Ding, M., and Bressler, S. L. (2001). Temporal dynamics of information flow in the cerebral cortex. *Neurocomputing* 38–40, 1429–1435.

Liebe, S., Fischer, E., Logothetis, N. K., and Rainer, G. (2009). Color and shape interactions in the recognition of natural scenes by human and monkey observers. *J. Vis.* 9, 1–16.

Marple, S. L. (1987). *Digital Spectral Analysis with Applications*. Englewood Cliffs: Prentice Hall.

Mizuseki, K., Sirota, A., Pastalkova, E., and Buzsaki, G. (2000). Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron* 64, 267–280.

Montemurro, M., Rasch, M., Murayama, Y., Logothetis, N., and Panzeri, S. (2008). Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Curr. Biol.* 18, 375–380.

Murthy, V., and Fetz, E. (1992). Coherent 25–25-hz oscillations in the sensorimotor cortex of awake behaving monkeys. *Proc. Natl. Acad. Sci. U.S.A.* 15, 5670–5674.

Nunez, P. L., Silberstein, R. B., Shi, Z., Carpenter, M. R., Srinivasan, R., Tucker, D. M., Doran, S. M., Cadusch, P. J., and Wijesinghe, R. S. (1999). EEG coherency II: experimental comparisons of multiple measures. *Clin. Neurophysiol.* 110, 469–486.

Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B., and Cadusch, P. J. (1997). EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.* 103, 499–515.

Okeefe, J. (1993). Hippocampus, theta and spatial memory. *Curr. Opin. Neurobiol.* 3, 917–924.

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

Pesaran, B., Pezaris, J., Sahani, M., Mitra, P., and Andersen, R. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* 5, 805–811.

Porcaro, C., Zappasodi, F., Rossini, P. M., and Tecchio, F. (2009). Choice of multivariate autoregressive model order affecting real network functional connectivity estimate. *Clin. Neurophysiol.* 120, 436–448.

Raghavachari, S., Kahana, M., Rizzuto, D., Caplan, J., Kirschen, M., Bourgeois, B., Madsen, J., and Lisman, J. (2001). Gating of human theta oscillations by a working memory task. *J. Neurosci.* 21, 3175–3183.

Raghavachari, S., Lisman, J., Tully, M., Madsen, J., Bromfield, E., and Kahana, M. (2006). Theta oscillations in human cortex during a working-memory task: evidence for local generators. *J. Neurophysiol.* 95, 1630–1638.

Rainer, G., Lee, H., Simpson, G., and Logothetis, N. (2004). Working-memory related theta (4–7 Hz) frequency oscillations observed in monkey extrastriate visual cortex. *Neurocomputing* 58–60. doi: 10.1016/j.neucom.2004.01.153

Salinas, E., and Sejnowski, T. (2001). Correlated neuronal activity and the flow of neuronal information. *Nat. Rev. Neurosci.* 2, 539–550.

Schelter, B. (2009). Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *J. Neurosci. Methods* 179, 121–130.

Schlögl, A. (2000). *The Electroencephalogram and the Adaptive Autoregressive Model: Theory and Applications*. Aachen: Shaker Verlag.

Schlögl, A. (2006). A comparison of multivariate autoregressive estimators. *Signal Processing* 86, 2426–2429. Special Section: Signal Processing in UWB Communications.

Schlögl, A., and Brunner, C. (2008). BioSig: a free and open source software library for BCI research. *Computer* 41, 44–50.

Schlögl, A., and Supp, G. (2006). Analyzing event-related EEG data with multivariate autoregressive parameters. *Prog. Brain Res.* 91, 135–147.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464.

Siapas, A., Lubenov, E., and Wilson, M. (2005). Prefrontal phaselocking to hippocampal theta oscillations. *Neuron* 46, 141–151.

Stevenson, I., Rebesco, J., Miller, L., and Körding, K. (2008). Inferring functional connections between neurons. *Curr. Opin. Neurobiol.* 18, 582–588.

Sun, X. (2008). "Assessing nonlinear Granger causality from multivariate time series," in *ECML PKDD '08: Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases* – Part II (Berlin/Heidelberg: Springer), 440–455.

Supp, G. G., Schlögl, A., Trujillo-Barreto, N., Müller, M. M., and Gruber, T.

(2007). Directed cortical information flow during human object recognition: analyzing induced EEG gamma-band responses in brain's source space. *PLoS ONE* 2, e684. doi: 10.1371/journal.pone.0000684.

Tallon-Baudry, C. (2009). The roles of gamma band oscillatory synchrony in human visual cognition. *Front. Biosci.* 14, 321–332.

Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci. (Regul. Ed.)* 3, 151–162.

Uhlhaas, P., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolic, D., and Singer, W. (2009). Neural synchrony in cortical networks: history, concept and current status. *Front. Integr. Neurosci.* 3:17. doi: 10.3389/neuro.07.017.2009.

Zetterberg, L. (1969). Estimation of parameters for a linear difference equation with application to EEG analysis. *Math. Biosci.* 5, 227–275.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 November 2009; paper pending published: 30 March 2010; accepted: 30 April 2010; published online: 26 May 2010.

Citation: Hoerzer GM, Liebe S, Schloegl A, Logothetis NK and Rainer G (2010) Directed coupling in local field potentials of macaque V4 during visual short-term memory revealed by multivariate autoregressive models. Front. Comput. Neurosci. ***4****:14. doi: 10.3389/fncom.2010.00014*

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 15 September 2010
doi: 10.3389/fncom.2010.00131

Analysis and modeling of ensemble recordings from respiratory pre-motor neurons indicate changes in functional network architecture after acute hypoxia

Roberto Fernández Galán[1]*, Thomas E. Dick[1,2] and David M. Baekey[2]

[1] *Department of Neurosciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA*
[2] *Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, School of Medicine, Case Western Reserve University, Cleveland, OH, USA*

Edited by:
Jakob H. Macke, University College London, UK

Reviewed by:
Jeffrey C. Smith, National Institute of Neurological Disorders and Stroke, USA
Robert J. Butera, Georgia Institute of Technology, USA

****Correspondence:***
Roberto Fernández Galán, Department of Neurosciences, School of Medicine, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4975, USA.
e-mail: rfgalan@case.edu

We have combined neurophysiologic recording, statistical analysis, and computational modeling to investigate the dynamics of the respiratory network in the brainstem. Using a multielectrode array, we recorded ensembles of respiratory neurons in perfused *in situ* rat preparations that produce spontaneous breathing patterns, focusing on inspiratory pre-motor neurons. We compared firing rates and neuronal synchronization among these neurons before and after a brief hypoxic stimulus. We observed a significant decrease in the number of spikes after stimulation, in part due to a transient slowing of the respiratory pattern. However, the median interspike interval did not change, suggesting that the firing threshold of the neurons was not affected but rather the synaptic input was. A bootstrap analysis of synchrony between spike trains revealed that both before and after brief hypoxia, up to 45% (but typically less than 5%) of coincident spikes across neuronal pairs was not explained by chance. Most likely, this synchrony resulted from common synaptic input to the pre-motor population, an example of stochastic synchronization. After brief hypoxia most pairs were less synchronized, although some were more, suggesting that the respiratory network was transiently "rewired" after the stimulus. To investigate this hypothesis, we created a simple computational model with feed-forward divergent connections along the inspiratory pathway. Assuming that (1) the number of divergent projections was not the same for all presynaptic cells, but rather spanned a wide range and (2) that the stimulus increased inhibition at the top of the network; this model reproduced the reduction in firing rate and bootstrap-corrected synchrony subsequent to hypoxic stimulation observed in our experimental data.

Keywords: neural control of respiration, working heart brainstem preparation, hypoxia, spike synchronization, bootstrap analysis, neural network simulation

INTRODUCTION

Respiration is controlled by neuronal circuits in the brainstem that have been studied extensively using various approaches and experimental techniques. The respiratory rhythm is generated in the ventrolateral brainstem and while the exact mechanism responsible for its genesis is debated (Del Negro et al., 2002; Rybak et al., 2008) the pathway by which the brainstem relays inspiratory signals to the diaphragm is well documented. As such, the brainstem, in particular the ventral respiratory column, represents an excellent model to study the neuronal control of motor responses that are self-regulated, adaptive and malleable.

Independent of the underlying interactions between intrinsic cellular properties and extrinsic network properties to generate inspiration, the formation of inspiratory motor activity is amenable to a hierarchical model with three layers. The first and second layers are in the rostral ventrolateral respiratory column. This group of neurons contains distinct neuronal populations. The first layer is in the pre-Bötzinger complex (preBötC) and is involved in the initiation of inspiration (pre-inspiratory, pre-I), whereas the second layer is less defined anatomically and contributes to the ramping output pattern (inspiratory-augmenting, I-Aug). The third layer is the rostral ventral respiratory group (rVRG), which is caudal to the preBötC and is a population of bulbo-spinal premotor neurons that transmit central inspiratory drive to spinal motoneurons, whose axons innervate the diaphragm. Thus, within this hierarchical organization the preBötC determines the timing of inspiration (rhythmicity), whereas each subsequent layer modulates the shape and amplitude of the efferent signals which directly control the mechanics of inspiration.

In this paper, we focus on the connectivity and coherent firing in a subpopulation of neurons that comprise a significant portion of the inspiratory motor output in the respiratory column (Feldman et al., 1984). This study utilizes multielectrode array recordings from an *in situ* rat preparation allowing us to monitor the neuronal activity of many neurons simultaneously as well as the phrenic motor output (Baekey et al., 2001). In addition, we performed a statistical analysis of these data using bootstrap techniques allowing us to identify temporal correlations (spike synchrony) between spike trains from different neurons that cannot be accounted for by chance. Synchronous activity may be derived from common input and as such, synchrony reflects the underlying connectivity from one layer to the next. In effect, recent experimental, theoretical and

computational studies have demonstrated that neurons receiving stochastic partially common inputs fire synchronous spikes (Galán et al., 2006b; Galán et al., 2007a; Ermentrout et al., 2008), a phenomenon known as *stochastic synchronization*. The fraction of synchronous spikes increases monotonically with increasing correlation (overlap) of the synaptic inputs (Galán et al., 2006b; Galán et al., 2007a; Ermentrout et al., 2008). Thus, the amount of synchrony between two spike trains that cannot be accounted for by chance reflects the amount of common input to the neuronal pair. In the final part of the study, we use this relationship to model the connections between presynaptic neurons with divergent projections onto postsynaptic neurons. In summary, we have recorded from pre-motor inspiratory neurons and from their pair-wise synchronization, we reverse-engineer the connectivity with an upstream layer of inspiratory-augmenting neurons (I-Aug), which in turn is driven by the neurons in the preBötC referred to above.

The respiratory rhythm is quite variable and can be modulated in amplitude and frequency by external stimuli, such as low oxygen (hypoxia). Whereas the frequency and amplitude of respiration increase during acute hypoxia (Powell et al., 1998), subsequently, the respiratory frequency decreases below baseline and gradually recovers over the next few minutes in what is known as post-hypoxic frequency decline, PHFD (Coles and Dick, 1996; Dick et al., 2004). We hypothesized this behavior is reflected in the neuronal activity of the ventral respiratory column. Thus, we expected changes not only in the firing rates of the neurons but also in their temporal cross-correlations, e.g., spike synchronization. In this paper, we report our analysis revealing significant changes in spike synchrony across pre-motor neurons after hypoxia.

MATERIALS AND METHODS

EXPERIMENTAL METHODS

All experiments were performed in accordance with the guidelines of the Institutional Animal Care and Use Committee (IACUC) of Case Western Reserve University.

General surgical methods

For these experiments we used the working heart brainstem preparation (WHBP) of the rat (Paton, 1996) (**Figure 1**). Male Sprague-Dawley rats (n = 6) (P21–P28, 60–100 g) were pretreated with heparin sodium (1000 units – IP), anesthetized with isoflurane (2–3%), then bisected below the diaphragm. The rostral half of the animal was submerged in cold artificial cerebrospinal fluid (aCSF) to decerebrate at the precollicular level. We removed the fur, skin and viscera, dissected the phrenic motor nerve and descending aorta and exposed the dorsal medullary surface.

After surgery, the preparation was moved to the recording chamber and mounted supine in a stereotaxic frame. The distal end of the descending aorta was cannulated with a #4 French, double-lumen catheter (Braintree Scientific) and perfused (21–28 ml/min – Marlow Watson 505S peristaltic pump) with an iso-osmotic aCSF saturated with 95% O_2/5% CO_2. Perfusion pressure was monitored through the other lumen (CWE TA-100 transducer-amplifier). The preparation was immobilized with vecuronium bromide (0.4 mg/200 ml perfusate). Perfusion pressure was maintained at 60–80 mmHg and corrected with 4 μM vasopressin (20 μl added to perfusate, as needed). Additionally, NaCN (0.1%, 50-μl bolus) was used to stimulate carotid chemoreceptors transiently initiating respiratory patterning in the preparation.

Electrode placement and neuronal sampling strategy

The distal end of the left phrenic nerve was drawn into a bipolar suction electrode. The signal was amplified (Grass P511), filtered (0.1–3 KHz) and digitized using a CED Power 1401 and Dell PC running Spike2 software. Pressure was adjusted to maintain an appropriate perfusion of the brainstem and pons indicated by a ramp patterned phrenic nerve output bursting at 15–30 breaths per minute and post inspiratory activity in the vagus nerve recording.

The 16-channel microelectrode array was secured on a stereotaxic frame aligning the tungsten electrodes (10–12 MΩ) perpendicular to the neural surface with eight electrodes on either side of the brainstem. Each set of eight was divided into two sagittally oriented linear rows of four electrodes separated by 250 μm while electrodes within each row are separated by 300 μm. Stereotaxic coordinates were used to position electrodes bilaterally among inspiratory pre-motor neurons in the rVRG. Each electrode was positioned in steps as small as a micron to isolate a single extracellular potential (**Figure 1**). In cases where more than one neuron was recorded on a single electrode, the principle component analysis (PCA) feature of the Spike2 software was used to discriminate individual spike trains (spike sorting). The independent depth adjustment of each electrode optimized the yield of parallel single neuron recordings.

Experimental protocol

With many single neurons monitored in the rVRG and a satisfactory nerve recording (signal-to-noise ratio >3), a 10-min baseline recording was made to characterize the recorded neurons, assess baseline synchrony, and for comparison with the poststimulus activity. After the baseline recording, the preparation was exposed to hypoxic perfusate (8% O_2/5% CO_2) for 15–25 s evoking a hypoxic ventilatory response followed by PHFD (Dick et al., 2001). Signals were recorded in Spike2 software for subsequent off-line analysis.

DATA ANALYSIS

Table 1 summarizes the source of the experimental data sets considered in this study. We recorded from brainstem respiratory neurons of six different rats. We recorded simultaneously from 8 to 23 neurons in a preparation and exposed each preparation to brief hypoxia (see Experimental Protocol) up to three times. While our recordings included inspiratory, expiratory, and non-modulated activity patterns, our analysis focused on the inspiratory activity. The total number of pairs of inspiratory neurons was 562. Some neuronal pairs were duplicates because each stimulation was considered as a different experiment, as the hypoxic stimuli were transient and reversible. If only one hypoxic exposure was analyzed for each animal, i.e., if each pair was considered once in our analysis, then the total number of pairs would be 203. However, these results were qualitatively the same as those for the 562 pairs.

Epochs (80 s) of the recordings were analyzed for baseline and PHFD. The recorded data included extracellular potentials from the microelectrode array and phrenic nerve activity (PNA). Multi-fiber PNA was "integrated" (low-pass filtered) with a linear integrator

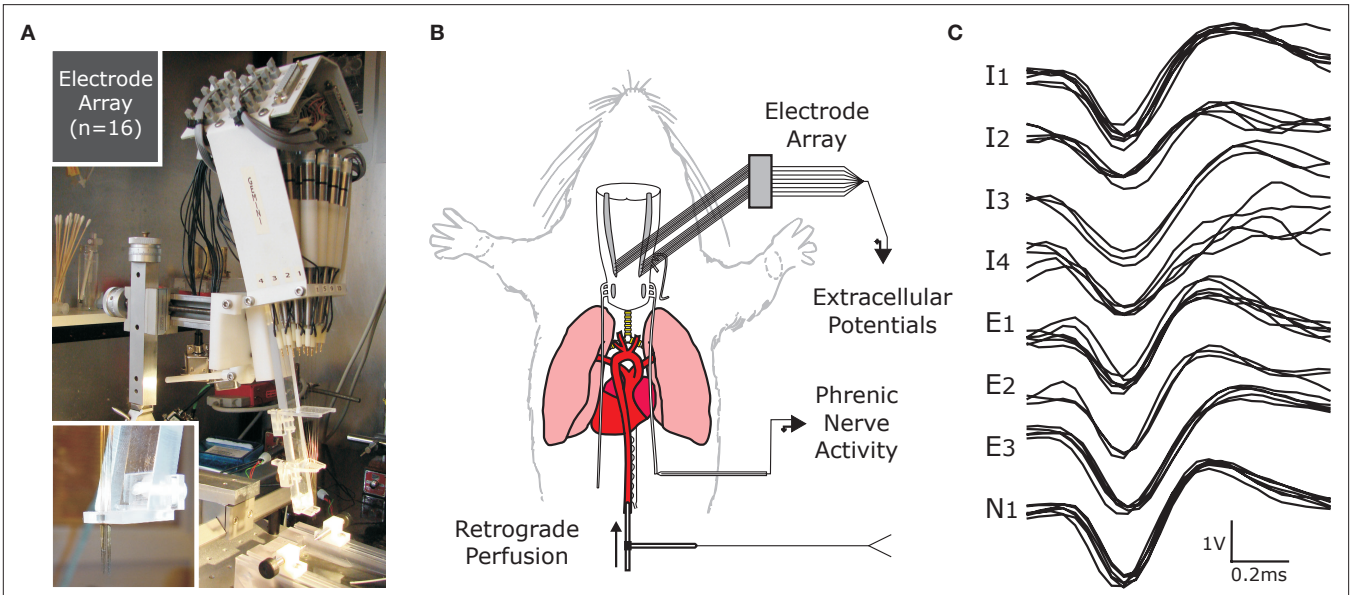


FIGURE 1 | Multielectrode array technology (A) was applied to an *in situ* perfused rat preparation **(B)** to obtain ensemble recordings **(C). (A)** The electrode array had 16 tungsten microelectrodes with impedance 10–12 MΩ. Each individual electrode was positioned with a micromotor in steps as small as 1 μm over a range of 1 cm. **(B)** Decerebrate P21–P28 rats were retrogradely perfused through the descending aorta with artificial cerebrospinal fluid bubbled with 95% O_2/5% CO_2. These preparations generate spontaneous respiratory pattern, which was monitored from phrenic nerve activity. **(C)** Extracellular action potentials were digitized and recorded using CED Spike2 software and a Power1401 data acquisition system. (The amplitude calibration as labeled is for seven of the recordings; the exception is I4 for which it is 0.2 V).

Table 1 | Origin of the data and number of cells and pairs investigated.

Preparation number	Stimulation number	Total number of cells	Number of I-cells	Number of I-pairs
1	1	12	5	10
2	1	8	4	6
3	1	23	17	136
	2	23	17	136
	3	23	17	136
4	1	9	6	15
	2	9	6	15
5	1	12	7	21
	2	12	7	21
	3	12	7	21
6	1	11	6	15
	2	11	6	15
	3	10	6	15
	Totals	175	111	562

using a 100-ms time constant to obtain a moving-time average of activity (**Figures 2A,B**) and using this integrated PNA, time stamps were added to indicate the onset of each inspiratory and expiratory phase. Action potentials of single neurons were converted to times of occurrence, i.e., spike trains (**Figures 2A,B**). These processed epochs were exported to Matlab (version R2008a) for further analysis.

The following measures were computed from a 10-min baseline period in order to ensure that there was only one neuron per channel and to characterize the cell-type: (1) *Autocorrelation histograms* were created for each spike train to ensure that it represents the activity of a single neuron (not shown). A spike train with potentials from two or more neurons would include short intervals not constrained by refractoriness. (2) *Cycle-triggered histograms* (**Figures 2C,D**) were used to classify activity patterns with significant respiratory modulation according to the phase (inspiratory – I, or expiratory – E) in which they are more active and by trends in their burst patterns (augmenting, decrementing or plateau). The cycle-triggered histograms were computed as the cross-correlation function of a spike train with respect to the phase of the phrenic nerve signal. The phase was calculated by linear interpolation of time between the beginning and the end of the respiratory cycle. Specifically, if t_k denotes the beginning of the k-th cycle, or equivalently, the end of the $(k-1)$-th cycle, the phase is defined as: $\varphi(t) = (t - t_k)/(t_{k+1} - t_k)$ for $t_k \leq t < t_{k+1}$. The beginning of the respiratory cycle, i.e., the trigger, was considered as the termination of inspiration.

Firing rates

The firing rate of a neuron is calculated as the inverse of the median interspike interval (ISI). This is a good approximation to the intraburst firing frequency. In other words, if two spike trains have different number of bursts but the firing frequency during the bursts is the same, then the firing rate will return very similar values.

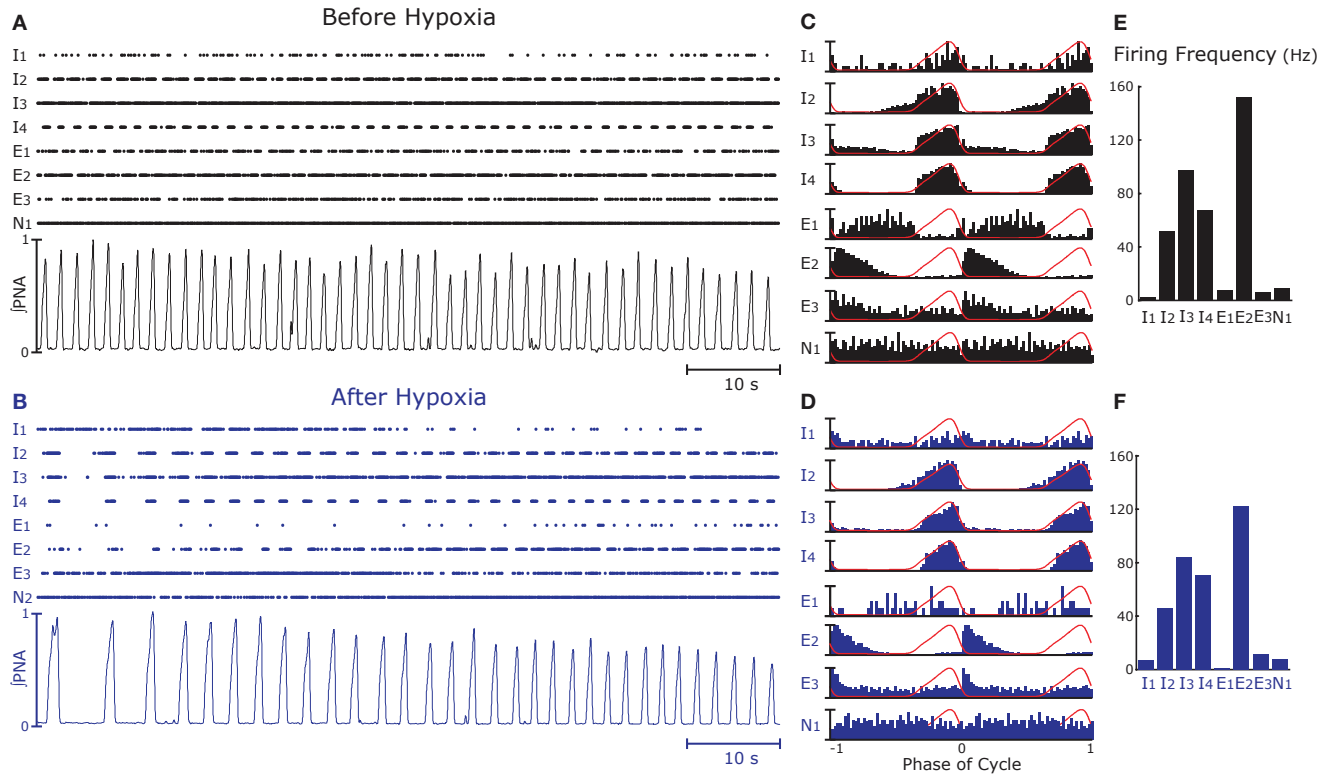


FIGURE 2 | Simultaneous recordings of neurons ($n = 8$) in the rostral ventral respiratory group (raster plots) and integrated phrenic nerve activity (∫PNA, see Data Analysis) before (A) and after (B) hypoxia. (A,B) The complete 80-s epochs analyzed; raster plots of eight simultaneously recorded neurons (those in **Figure 1**) and ∫PNA. In **(B)**, immediately after hypoxia (8% O_2, 5% CO_2, and 87% N_2 for 15 s), respiratory frequency is less than that before hypoxia **(A)** and amplitude of ∫PNA displays. **(C,D)** Cycle-triggered averages (CTA) of spike activity during the epochs. Recordings were from the caudal VRC and include neurons with inspiratory- (I1, I2, I3, I4), expiratory- (E1, E2, E3), and a non-respiratory (N1)-modulated activity. Except for I1 and E1, major features of the activity profiles in the CTA did not change after hypoxia. **(E,F)** Action potential firing frequency decreased after hypoxia.

Raw synchrony as a cross-correlation coefficient between spike trains

Our experiments are designed to quantify spike synchrony across neurons in the respiratory column, and their modulation across different states of the preparation. Spike synchronization between neuron X and Y is calculated in the following way. Let t_i^x be the time of the i-th spike in channel X and let t_j^y be the time of the j-th spike in neuron Y. For each spike time pair (t_i^x, t_j^y), the relative separation is compared to a tolerance, $\delta = 2$ ms. We then define R_{xy} as the number of pairs such that $|t_i^x - t_i^y| \leq \delta$. Analogously we define R_{xx} as the number of pairs such that $|t_i^x - t_j^x| \leq \delta$ and R_{yy} as the number of pairs such that $|t_i^y - t_j^y| \leq \delta$. A *raw* estimate of spike synchrony then reads:

$$\hat{S}_{xy}^{\text{raw}} = \frac{R_{xy}}{\sqrt{R_{xx}R_{yy}}}. \quad (1)$$

If δ is sufficiently small, i.e., much smaller than the typical ISI, and the two neurons fire at the same rate, expression (1) yields the fraction of spikes that occur at the same time in both neurons. This measure resembles the definitions of neuronal synchronization and spike time reliability reported in various publications (Hunter et al., 1998; Schreiber et al., 2004; Galán et al., 2006a, 2007b), which are equivalent to the cross-correlation coefficient of two spike trains convolved with a Gaussian or a step function of width 2δ. In fact, the algorithm described above is mathematically equivalent to the latter case but without calculating the convolution explicitly. This way, we only need to store spike times in the computer's memory and not the whole binary traces (1 = spike, 0 = no spike) representing the firing of each neuron, which would be required to calculate the convolution explicitly. Because the binary traces are sparse, i.e., they contain many more 0's than 1's, by keeping the spike times only we save a significant fraction of memory space, which in turn speeds the computational implementation of our synchrony algorithm considerably. For example, the analysis presented in **Figure 2** took just a few seconds to run fully in Matlab R2008a on a Dell PC with an Intel® Xeon® CPU (1.60 GHz with 2GB RAM). Furthermore, since this measure of synchrony is fundamentally a cross-correlation matrix, it allows us to run a clustering analysis to identify neuronal pairs that are more coherent among themselves than with respect to other neuronal groups (see **Figure 4**). This technique has demonstrated the existence of synchronized assemblies among inspiratory neurons of the central pattern generator of respiration (Baekey et al., 2009), which is upstream of the network investigated here.

Bootstrapping and bootstrap-corrected synchrony

Expression (1) quantifies the total amount of synchronization, including the fraction of synchronous spikes that would occur by chance in two uncoupled neurons receiving independent inputs. We therefore refer to it as *raw* synchrony. Since we are interested in the synchronous events that occur as a result of network interactions, we need to subtract the amount of synchrony expected by chance, S^0_{xy}, which is higher, the higher the firing rates of the neurons. To this end, we apply a standard bootstrap technique: we use surrogate data obtained by shuffling the spike times of each neuron independently.

By shuffling we mean that the ISI from the actual recordings are randomly permuted. For example: Let the times of four successive spikes be $\{t_1, t_2, t_3, t_4\}$. The ISI are $\{\Delta_1 = t_2 - t_1, \Delta_2 = t_3 - t_2, \Delta_3 = t_4 - t_3\}$. A randomly shuffled sequence of ISI $\{\Delta_2, \Delta_3, \Delta_1\}$ results in the shuffled spike train $\{t_1, t_1 + \Delta_2, t_1 + \Delta_2 + \Delta_3, t_1 + \Delta_2 + \Delta_3 + \Delta_1\} = \{t_1, t_1 - t_2 + t_3, t_1 - t_2 + t_4, t_4\}$, which has the same ISI distribution as the original spike train as well as the same mean firing rate. Note that shuffling preserves the times of the first and last spikes. As a result, all shuffles will have at least these two spikes fully synchronized. To correct for this artifactual "synchrony" the shuffled sequence was randomly shifted up to hundred milliseconds.

Because the firing pattern of each neuron is typically different during inspiration and expiration, we shuffle the spikes separately for the inspiratory and the expiratory phase. This way the ISI distribution of the surrogates during inspiration is identical with the ISI distribution of the experimental data during inspiration, and analogously, during expiration. Spike-shuffling, however, alters the timing of the spikes randomly and therefore, the auto- and cross-correlations of the actual data are not preserved in the surrogates. Since the spikes of each neuron are shuffled independently, the level of synchrony in the surrogate data represents the amount of synchrony that can be accounted for by chance. Obviously, this value depends on how the spikes were shuffled, i.e., on the random realization of the surrogate data set. Thus, in order to be more rigorous, we first generate $N = 300$ surrogate data sets for each neuronal pair XY and then calculate the distribution of synchrony values. The 99th percentile of this distribution is our estimate of synchrony by chance, $\hat{S}^0_{xy}$. This implies that if the synchrony level for that pair in the actual data, $\hat{S}^{\text{raw}}_{xy}$ is greater than $\hat{S}^0_{xy}$, then that synchrony level is significant with 99% confidence. Moreover, the difference $\hat{S}^{\text{raw}}_{xy} - \hat{S}^0_{xy}$ represents the amount of synchrony that cannot be accounted for by chance and is therefore due to temporal correlations emerging from network interactions. Our bootstrap-corrected synchrony measure, S_{xy} thus reads:

$$S_{xy} = \begin{cases} \hat{S}^{\text{raw}}_{xy} - \hat{S}^0_{xy}, & \text{if} \quad \hat{S}^{\text{raw}}_{xy} > \hat{S}^0_{xy} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Note that S_{xy} reports the level of non-trivial spike synchronization. Indeed, for two different neurons, S_{xy} is the amount of synchrony that cannot be accounted for by chance, and for the same neuron, S_{xx} is always zero, since a spike train is always perfectly synchronized with itself but this synchrony is trivial.

COMPUTATIONAL SIMULATIONS AND NETWORK MODEL

General considerations about the model

The purpose of the computer simulations is to investigate a minimalist model of the brainstem inspiratory network that replicates the following features of the experimental data: (1) during baseline conditions, significant spike synchrony in pairs of inspiratory premotor neurons reflects common synaptic inputs; (2) after hypoxia, the firing rate of inspiratory premotor neurons decreases significantly due to decreased excitation or increased inhibition; and (3) after hypoxia, there are significant changes in synchrony that cannot be explained by chance nor by the reduction in the firing rates. Most of these changes are negative (synchrony decreases) but some are positive.

Our network simulations, although original, borrow several elements from previously published models. We employ simplified but realistic models of the single neuron dynamics, as is the case with other models of the respiratory network (Rybak et al., 2004, 2008) and divide the network into a pre-inspiratory population, an inspiratory population and an inspiratory premotor population (see **Figure 6**). We record from the inspiratory premotor population in our experiments. These neurons project to motoneurons whose axons form the phrenic nerve, which are not included in our model.

The single-cell dynamics have been adapted from the neuronal model recently proposed by Izhikevich, which is basically a quadratic integrate-and-fire model with realistic phase-resetting properties. This model consists of two variables, the membrane potential and a recovery variable, both with a resetting threshold (Izhikevich, 2004).

Recent experimental work has provided evidence for functional SK channels (a subtype of calcium dependent potassium channels) in pre-motor neurons (Tonkovic-Capin et al., 2003). These channels, similar to other potassium channels, endow neurons with type II excitability, which can be modeled as neurons with a resonator-like phase-resetting curve (Galán et al., 2006a). Therefore we use the Izhikevich model of resonator neuron for the pre-motor population and for simplicity, for the inspiratory population as well. For the pre-inspiratory neurons in the preBötC we use a similar model but with a saw-tooth drive, $I(t)$ that mimics the intrinsic bursting properties of these neurons along the lines of models previously published by other groups (Butera et al., 1999a,b; Del Negro et al., 2001).

In order to produce population wide activity that is synchronized on the time scale of the inspiratory burst, all pre-inspiratory neurons were driven by a saw-tooth drive, $I(t)$, and noise was added to produce a temporal dispersion of spiking activity. The resulting pattern of pre-inspiratory activity mimicked the pattern described in previous models where biophysical mechanisms producing a slow wave, saw-tooth-like membrane potential trajectory were incorporated along with heterogeneity of cellular and synaptic properties (Butera et al., 1999a,b; Del Negro et al., 2001). We note that this form of coherent activity may also be facilitated by the presence of gap junctions in the system (Rekling et al., 2000; Bou-Flores and Berger, 2001; Solomon et al., 2001).

Network architecture

Stochastic synchronization is an efficient mechanism for generating coherent activity in neuronal networks (Galán et al., 2006b). This phenomenon emerges when uncoupled neurons receive

common fluctuating inputs, for example, synaptic barrages from divergent presynaptic terminals. Because this connectivity pattern is ubiquitous in the brain, stochastic synchronization can account for most temporal correlations observed in neural circuits. We therefore hypothesize that this is the phenomenon underlying spike synchronization in the brainstem inspiratory network. Since feed-forward divergent projections are sufficient to cause downstream synchrony, we modeled a pure feed-forward network (i.e., no connections within each layer). A simplified diagram is shown in **Figure 6** (left) to illustrate the fundamental features of the simulated network: pre-inspiratory neurons in the pre-Bötzinger complex (excitatory (+) and inhibitory (−) open circles); inspiratory-augmenting neurons in the area of the pre-Bötzinger complex (3 circles: red, yellow, and blue); and pre-motor neurons in the rVRG (9 circles: red, orange, purple, green, and blue). The similarity of the colors between any two premotor neurons indicates the proportion of common inputs, i.e., the blending of the primary colors: red, blue and yellow. While not all inputs to pre-motor neurons necessarily originate from the inspiratory-augmenting population in the pre-Bötzinger area, a significant portion do (Schwarzacher et al., 1995). For the purpose our model the exact anatomical location of the neurons is not essential. Note that in our model the concept of layer is topological, not anatomical: it refers to how neurons are connected, and not to where they are precisely located.

Recently published data demonstrate that a fraction of inspiratory pacemaker neurons in the pre-Bötzinger complex are inhibitory, as they express the glycine transporter 2 (GlyT2) gene (Morgado-Valle et al., 2010). We assume in our model that the majority of pre-inspiratory neurons are inhibitory. Although there is no direct evidence for this assumption, it is consistent with recent studies demonstrating that a major fraction of all inspiratory neurons in the rodent brainstem slice preparation are glycinergic (Winter et al., 2010).

In our model, both excitatory and inhibitory pre-Bötzinger complex neurons provide divergent connections to the inspiratory-augmenting population. Each of the pre-inspiratory "pacemaker" neurons has the same connection probability with any inspiratory-augmenting neuron. Modifying the ratio between inhibition and excitation while keeping the other parameters of the network unchanged, we determined that there must be at least 15% excitation (ratio inhibition/excitation = 85/15) in order for the premotor neurons to fire some spikes. However, the firing rate, the synchrony values and their changes after hypoxia were only comparable with the experimental results when the fraction of excitatory pre-inspiratory neurons was in the range between 25% and 45%. When that fraction was over 50%, the median bootstrap corrected synchrony between premotor neurons took large values (around 0.6 and higher) before and after hypoxia, indicating that most pairs were highly synchronized, contrary to what we observed experimentally. In the simulations producing the results shown in the figures, we took 70% of the pre-inspiratory neurons as inhibitory and 30% as excitatory (ratio 70/30).

The inspiratory-augmenting population in turn provides divergent excitatory connections to pre-motor neurons, but not with equal probability. The connection probability ranges from $P_{\min} = 10\%$ to $P_{\max} = 30\%$. Moreover, the number of projections from each inspiratory-augmenting neuron plotted versus its rank in the population follows a power-law distribution. The rank equals one for the neuron with the largest number of projections and equals the total number of neurons for the neuron with the smallest number of projections. The power-law distribution is not essential to reproduce the experimental data. In fact, a distribution that decays exponentially with the rank yields qualitatively the same results, and so will any distribution that decays sufficiently fast with the rank, as it guarantees that the inspiratory-augmenting neurons are not equally driving the premotor neurons. This implies that some are strong drivers (hubs) but most are weak. In particular, the power-law distribution as a function of the rank, $P(r)$, that we consider has the form: $P(r) = P_{\max} r^{-A}$, with $A = \log(P_{\max}/P_{\min})/\log(N)$. Although a uniform distribution can account for different levels of synchrony between pre-motor pairs, it can neither reproduce the range of synchrony nor the synchrony changes after brief hypoxia that we observed in the experiments.

In addition to the network topology, the fast synaptic kinetics ($\tau \sim 10$ ms) contribute to the overall level of stochastic synchronization in the pre-motor population, as recently shown in simulations and experiments in other parts of the brain (Galán et al., 2008).

Implementation details

The dynamics of the r-th neuron in each subpopulation is determined by the membrane potential, V and a recovery variable, U. The parameters of the model are chosen so that the membrane potential is in mV and time in ms. When the membrane potential exceeds 30 mV, the membrane potential is reset to $V = c$ and the recovery variable is reset to $U = d$ (see **Table 2**).The superscripts indicate the neuronal type of each subpopulation: p for pre-inspiratory (30 neurons), i for inspiratory-augmenting (90 neurons) and m for premotor (100 neurons). The dynamics of the synaptic conductances for each neuron, r, are denoted by G. The superscripts indicate the neuronal type and the nature of the synaptic conductance: I for inhibitory, E for excitatory. The synaptic connections are denoted by J and are generated randomly with the probabilities described in the previous paragraph every time the simulation program runs. The superscripts of J refer to the layers being connected: pi for pre-inspiratory to inspiratory-augmenting neurons, and im for inspiratory-augmenting to premotor neurons. The sign "+" as a superscript means that only the excitatory connections are considered and the inhibitory connections are ignored. Analogously, the

Table 2 | Parameters used in the simulations and their values.

Parameter	Value
a^p, a^i, a^m	0.20
b^p, b^i, b^m	0.21
c	−65
d	2
g^{iE}, g^{iI}	0.08
g^{mE}, g^{mI}	0.02
τ^{iE}, τ^{iI}, τ^{mE}, τ^{mI}	10
E_E	0
E_I	−75
σ	0.5

sign "–" as a superscript means that only the inhibitory connections are considered and the excitatory connections are ignored. The variable S tracks the firing of presynaptic neurons; $S = 0$, unless the presynaptic neuron fired at the previous time point, then $S = 1$. The dynamical equations were integrated in time with the Euler method ($dt = 0.5$ ms).

Dynamics of the pre-inspiratory population with an intrinsic saw-tooth drive, $I(t)$ and background Gaussian noise, $\eta(t)$ with standard deviation, σ:

$$\begin{cases} \dfrac{dV_r^p}{dt} = 0.08\left(V_r^p\right)^2 + 10V_r^p + 280 - 2U_r^p + I(t) + \sigma\eta(t) \\ \dfrac{dU_r^p}{dt} = a^p\left(b^pV_r^p - U_r^p\right) \end{cases}$$

$$I(t) = \frac{5\,\mathrm{mod}(t,2)}{2} \times \begin{cases} 1, \text{ if } \dfrac{\mathrm{mod}(t,6)}{2} < 1 \\ 0, \text{ otherwise} \end{cases}$$

Dynamics of the inspiratory population:

$$\begin{cases} \dfrac{dV_r^i}{dt} = 0.08\left(V_r^i\right)^2 + 10V_r^i + 280 - 2U_r^i - G^{iE}\left(V_r^i - E_E\right) - G^{iI}\left(V_r^i - E_I\right) \\ \dfrac{dU_r^i}{dt} = a^i\left(b^iV_r^i - U_r^i\right) \end{cases}$$

Dynamics of the pre-motor population:

$$\begin{cases} \dfrac{dV_r^m}{dt} = 0.08\left(V_r^m\right)^2 + 10V_r^m + 280 - 2U_r^m - G^{mE}\left(V_r^m - E_E\right) \\ \qquad - G^{mI}\left(V_r^m - E_I\right) \\ \dfrac{dU_r^m}{dt} = a^m\left(b^mV_r^m - U_r^m\right) \end{cases}$$

Temporal evolution of synaptic conductances for the inspiratory neurons:

$$\begin{cases} \dfrac{dG_r^{iE}}{dt} = -\dfrac{G_r^{iE}}{\tau^{iE}} + g^{iE}\sum_s J_{rs}^{pi+}S_s^p \\ \dfrac{dG_r^{iI}}{dt} = -\dfrac{G_r^{iI}}{\tau^{iI}} + g^{iI}\sum_s J_{rs}^{pi-}S_s^p \end{cases}$$

Temporal evolution of synaptic conductances for the premotor neurons:

$$\begin{cases} \dfrac{dG^{mE}}{dt} = -\dfrac{G^{mE}}{\tau^{mE}} + g^{mE}\sum_r J_{rs}^{im+}S_s^i \\ \dfrac{dG^{mI}}{dt} = -\dfrac{G^{mI}}{\tau^{mI}} + g^{mI}\sum_r J_{rs}^{im-}S_s^i \end{cases}$$

In the absence of stimulation the neurons of the model rest around −60 mV. Neither the resting potential nor the resetting parameters c and d play a significant role in the dynamics of this quadratic integrate-and-fire model. In this type of model, the time T to reach the resetting threshold from any starting value of the membrane potential V_0, is given by the arctangent function: $T(V_0) \sim \arctan\left(V_0/\sqrt{I}\right)$ where I is the driving current. Note, that the arctangent function is a sigmoid with a horizontal asymptote for large negative values of V_0. That means that for two different starting conditions V_0^1, and V_0^2 such that $V_0^1 \ll 0$ and $V_0^2 \ll 0$, the difference in time to reach threshold is negligible. Indeed, changing the value of c, d or the resting potential of the neurons has no effect on the frequency or synchrony in our model.

Modeling the dynamics after hypoxic stimulation

We hypothesize that hypoxic stimulation facilitates the recruitment of pre-inspiratory neurons. In our model, this implies that the inspiratory-augmenting neurons are being more inhibited, providing less excitation to the premotor neurons (**Figure 6,** right). As a result, the premotor neurons will have shorter and fewer bursts of activity, as observed in our experiments. In the simulations, the recruitment of pre-inspiratory cells is equivalent to adding more inhibitory neurons (10%), as compared to the network before hypoxia.

RESULTS

EXPERIMENTAL DATA

We recorded multielectrode data from six animals in the rVRG that included pre-motor inspiratory neurons, as reported in **Table 1**. A representative example of the recorded data for a given experiment is displayed in **Figures 1C and 2**. The raster plots show the firing of eight neurons over 80 s before and after hypoxia (**Figures 2A,B**, respectively). Integrated PNA is shown below the raster plots. The cross-correlation function between the phase of the phrenic nerve signal and the activity of each neuron, also known as cycle-triggered histogram (see Materials and Methods), is plotted in **Figures 2C,D**. These histograms allowed classification of the neurons according to the phase of the cycle when they were maximally active: inspiratory pre-motor neurons (I1–I4) were active during the burst of PNA (inspiration), whereas expiratory neurons (E1–E3) discharge when PNA was quiescent (expiration). One neuron was not modulated by respiration (N1). Neuronal firing frequency, calculated as the reciprocal of the median ISI (see Materials and Methods), is displayed in **Figures 2E,F**.

In the remaining analysis, we focused on the premotor neurons with inspiratory activity before and after hypoxia. After hypoxia there is an apparent decrease in the number and duration of bursts in these neurons. We then investigated whether the changes in spiking activity were consistent across neurons and whether, on a finer time scale, neurons are also tightly synchronized to each other.

For the 111 inspiratory pre-motor neurons, we observed a significant reduction of action potentials after hypoxia ($p < 0.05$, Wilcoxon sign rank test, **Figure 3A**). However, the intraburst firing frequency did not significantly change (**Figure 3B**). These results suggested that the inputs rather than the firing threshold of the neurons had changed, specifically an increase in inhibitory or decrease in excitatory inputs. We determined coincident (within a 2-ms time window) spikes during the 80-s epochs for each neuronal pair and found that the number of coincident events decreased significantly after hypoxia ($p < 0.05$ and Wilcoxon sign rank test, **Figure 3C**). This decrease could have resulted from the overall decrease in firing rate. Therefore, we calculated the harmonic mean of the action potentials for each neuron in the pair before and after hypoxia and found that it also decreased significantly (**Figure 3D**). However, we could not

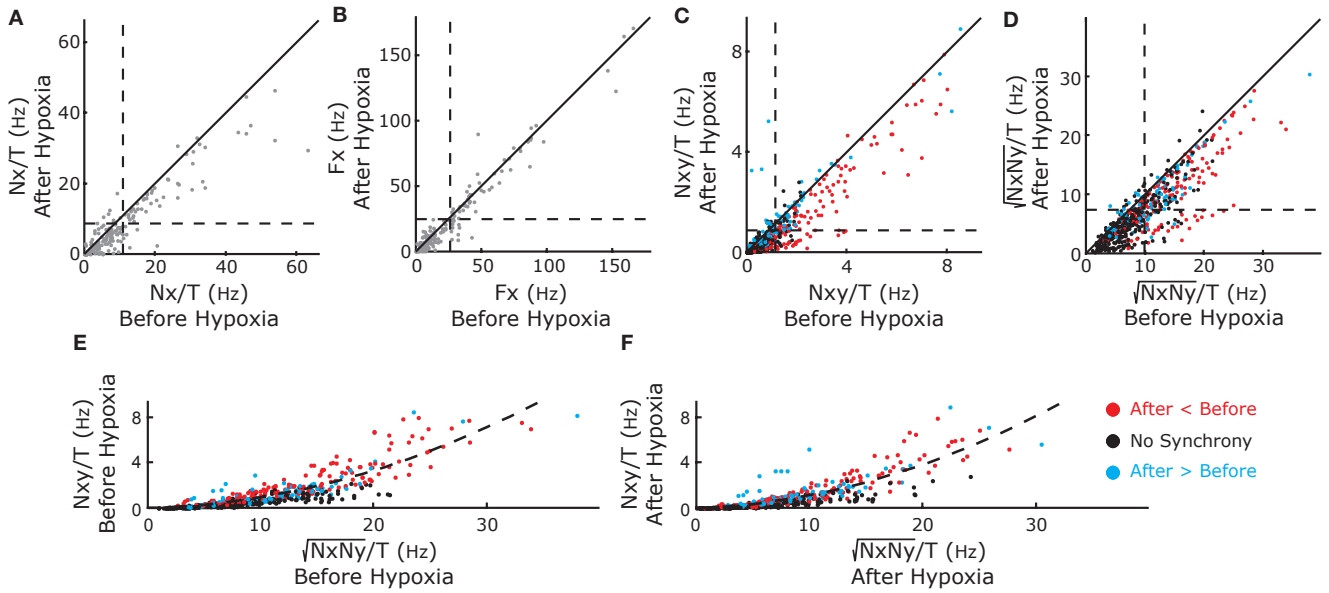


FIGURE 3 | Changes in activity of the recorded inspiratory premotor neurons (n=111) before and after hypoxia. (A) The total number of spikes discriminated divided by the duration of the analyzed epoch ($T = 80$ s) before (x-axis) versus after (y-axis) hypoxia. In this and **(B–D)**, the diagonal solid line is the line of identity. The majority of points, especially for the neurons that have the highest average frequency before hypoxia, are below the line of identity. The horizontal and vertical dashed lines plot the means before and after hypoxia, respectively [also for **(B–D)**]. The intersection of the dashed lines (11.0 ± 0.9 before versus 8.7 ±0.7 after hypoxia) is below the line of identity. **(B)** Firing frequency (Fx) of activity as the inverse of the mean ISI before versus after hypoxia. The intersection of the means is close to the line of identity (26.1 ± 2.3 before versus 24.3 ± 2.3 after hypoxia). **(C)** The frequency of coincident action potentials (Nxy, pairs of spikes occurring within 2 ms) decreased after hypoxia. The intersection of the dashed lines (1.2 ± 0.1 before versus 0.9 ± 0.1 after hypoxia) is below the line of identity. The color of the points is defined by where the points landed on the curve in **Figure 5** with values of synchrony corrected by the bootstrap method: red represents an increase in bootstrap-corrected synchrony; blue, a decrease; and black has no synchrony (these points are at the origin in **Figure 5**). This color coding is the same for all panels. **(D)** Harmonic mean of the number of fired spikes in each pair (√NxNy), before and after hypoxia. The intersection of the dashed lines (10.0 ± 0.2 before versus 7.4 ± 0.2 after hypoxia) is below the line of identity. **(E)** The relationship between coincident spikes (Nxy) and spike occurrence (√NxNy) before hypoxia. The dashed line is the relationship showing that as spike frequency increases, there is a greater probability of coincident spikes. **(F)** The relationship between coincident spikes (Nxy) and spike occurrence (√NxNy) after hypoxia. As before hypoxia, there is a greater probability of coincident spikes with increasing firing frequency. Means shown with errors as SEM.

conclude that the decrease in synchronous spikes resulted from the decrease in firing rate. Because neuronal synchrony was defined by the cross-correlation coefficient between two spike trains, which was the number of coincident spikes divided by the harmonic mean of the spikes fired by each neuron (see Materials and Methods), we compared the numerator of this expression with the denominator, before (**Figure 3E**) and after (**Figure 3F**) hypoxia. In both cases, the numerator grew faster than the denominator as the number of fired spikes increased. The quadratic fit had a slightly shallower slope after compared to before hypoxia, suggesting that hypoxia reduced the overall level of raw neuronal synchrony across pairs of inspiratory pre-motor neurons.

To exclude effects of firing rate on neuronal synchrony, we used a bootstrap technique as explained in Materials and Methods. Briefly, for each neuron we generated surrogate spike trains with the same number of spikes and the same ISI distribution as the real data and calculated the neuronal synchronization between each pair of surrogate spike trains. We repeated this process 300 times to obtain a distribution of synchrony values for the surrogate data sets. If the synchrony value of the real data was greater than the 99th percentile of that distribution, then synchrony was significant with a 99% confidence. Otherwise, the synchrony of the real data was not significant and mapped to zero. **Figure 4A**, shows the raw synchrony matrix, the 99th percentile of the synchrony distribution (described as synchrony confidence limit) and the bootstrap-corrected synchrony, before (top) and after (bottom) hypoxia for the experiment shown in **Figure 2**. Synchrony was significant among a subgroup of inspiratory pre-motor neurons but not among the rest, with the exception of two expiratory neurons that were weakly synchronized (E2–E3). In the bootstrap-corrected synchrony after hypoxia (**Figure 4B**), one pair of inspiratory neurons increased its synchrony (I3–I4) but two other pairs (I2–I3 and I2–I4) decreased their synchrony. The dendrograms derived from the bootstrap-corrected synchrony matrices reflected these changes and revealed a tight cluster of inspiratory neurons (I2, I3, and I4) that changed synchrony slightly after hypoxia (**Figure 4C**).

We observed the same trend for the bootstrap-corrected synchrony across the whole set of inspiratory pre-motor neurons (**Figure 5A**): 162 pairs (29%) decreased their synchrony after hypoxia; 86 pairs (15%) increased; and 314 pairs (56%) were not significantly synchronized either before or after hypoxia. Those who were synchronized before and/or after hypoxia had a significant trend to decrease their synchrony ($p < 0.05$, Wilcoxon sign rank test), as revealed by the histogram of difference in synchrony, which

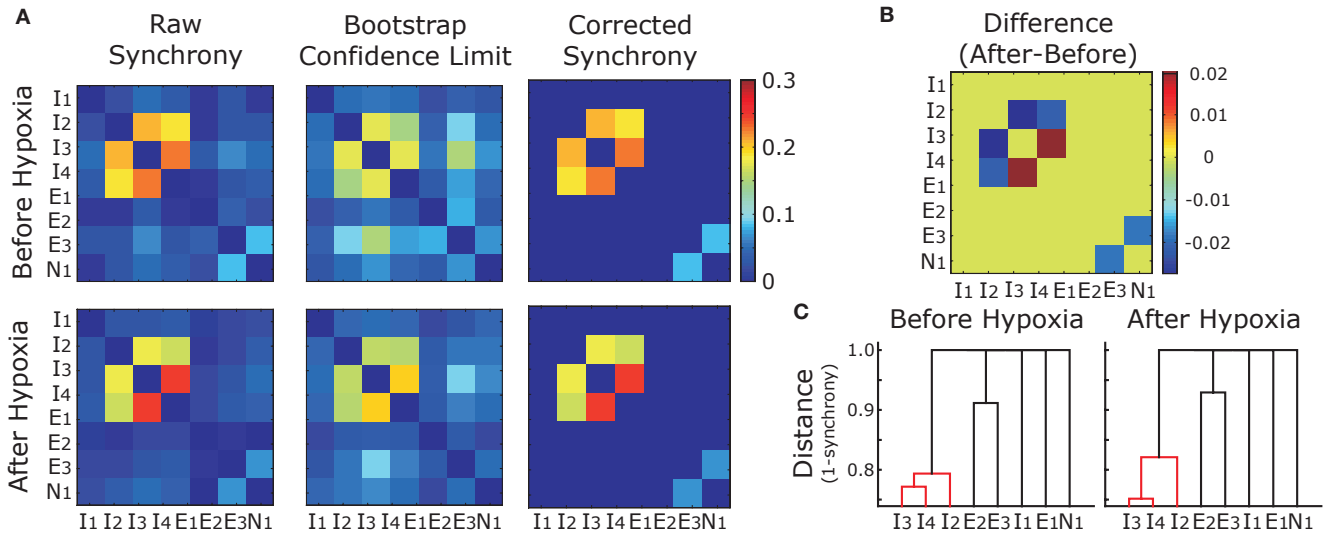


FIGURE 4 | Color-coded synchrony matrices (A, B) and dendrogram (C) display the synchrony among pairs of recorded neurons for the experiment in **Figures 1 and 2. (A)** Left-hand Column, the raw synchrony (not corrected for firing rate); Middle Column, upper confidence limit of synchrony in shuffled data, the bootstrap correction for the dependence of synchrony on the firing rate; and Right-hand Column, corrected synchrony (raw synchrony-synchrony of the shuffles). Top Row, synchrony in recordings before hypoxia; Bottom Row, after hypoxia. After bootstrap correction, the majority of pairs appear not to be synchronized, however, three of six pairs of premotor I neurons are synchronized. **(B)** The effect of brief hypoxia on synchrony (after-before hypoxia). One pair (red, I3 and I4) increases whereas the other three pairs decrease synchrony. **(C)** A dendrogram shows the clustering of neurons before (left) and after (right) hypoxia. In summary, these different graphics complement each other and consistently reveal significant changes in synchrony before and after hypoxia.

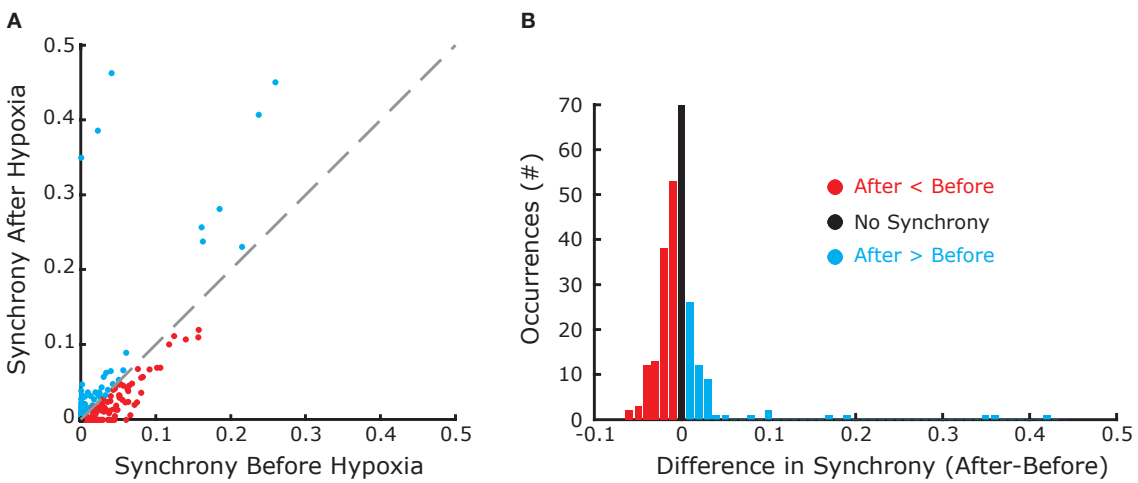


FIGURE 5 | Group data showing increases and decreases in synchrony before and after hypoxia. (A) The magnitude of bootstrap-corrected synchrony before hypoxia (*x*-axis) plotted against that after hypoxia (*y*-axis) for all pairs of neurons recorded. The diagonal dashed gray line is the line of identity. Blue dots above this line indicate pairs where synchrony increased after hypoxia; the red dots below indicate that synchrony decreased. Black dots, located at the origin, represent pairs that were not significantly synchronized either before or after hypoxia. **(B)** Histogram representing the synchrony difference (after–before). Blue, red, and black match with **(A)** Clearly, more pairs decreased than increased synchrony.

is clearly skewed to the left (**Figure 5B**). Obviously, since most pairs were not synchronized either before or after synchrony, considering all pairs in the analysis would not yield significant changes of synchrony at the population level.

Summarizing the experimental results, we observed significant synchrony across approximately half of the pairs of inspiratory pre-motor neurons before hypoxia. After transient hypoxic stimulation, synchrony decreased across approximately two-thirds and increased across one-third of those pairs. These synchrony changes were independent of the decrease in the number of spikes after hypoxia.

We propose that these effects can be accounted for by enhanced inhibitory input, as sketched in **Figure 6**. The inspiratory pre-motor neurons receive excitatory divergent input from

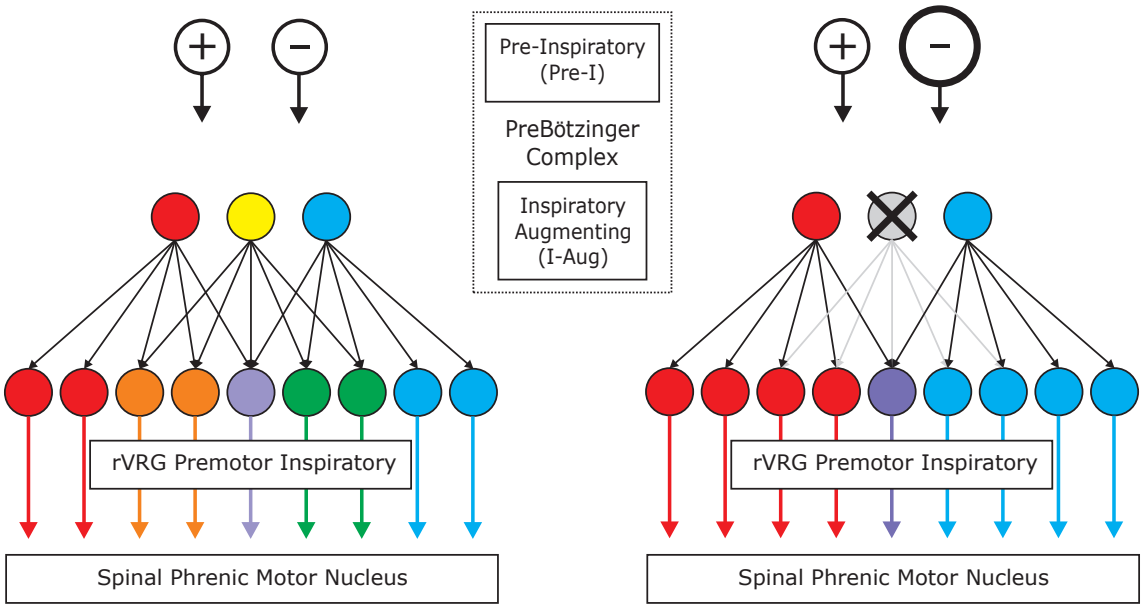


FIGURE 6 | Simplified diagram of the simulated network before (left) and after (right) hypoxia. The top layers [two Open Circles and three Filled Circles (Red, Yellow, and Blue)] contain two types of neurons in the preBötC; the excitatory and inhibitory Pre-I neurons as well as the excitatory I-Aug neurons. After hypoxia, there is a recruitment of inhibitory pre-I neurons in the preBötC which turns off some of the I-Aug activity in the preBötC [Yellow filled circle becomes inactive (gray with X)]. The third layer (nine Filled Circles) contains I-Aug neurons of the rVRG. The vast majority of these are bulbo-spinal premotor neurons and transmit central respiratory drive to spinal motor neurons. The blending of colors represents the degree of overlapping input and hence, the more similar the colors, the higher the synchrony. The elimination of a presynaptic input increases synchrony among some premotor pairs (red and blue or green and blue) but decreases synchrony among others (orange and green).

inspiratory-augmenting neurons in the preBötC, which in turn receive mostly inhibitory divergent input from preBötC pre-inspiratory neurons. A recruitment of more inhibitory pre-I neurons following hypoxia could account for the overall increase of inhibition in the hierarchical network. At the same time, the divergent projections between layers create temporal correlations across postsynaptic cells receiving overlapping inputs, which translates to significant spike synchronization. In **Figure 6**, neurons with similar colors are temporally correlated: the more similar the colors, the higher the synchrony. Note that after hypoxia, the elimination of a presynaptic input increases synchrony among some premotor pairs (red and blue or green and blue) but decreases synchrony among others (orange and green). To test our hypothesis, we implemented a computational model of the network dynamics with these ingredients (see details in Materials and Methods) and analyzed the simulated network dynamics in the same fashion as the experimental data.

COMPUTATIONAL SIMULATIONS

Figure 7 displays the raster plots of neuronal activity for the whole network during an inspiratory burst. The analysis of the traces for the pre-motor inspiratory population is presented in **Figures 8 and 9**. **Figure 8A** reveals a significant decrease in the number of fired spikes after hypoxia. However, the firing rate of the neurons (**Figure 8B**), calculated as the inverse of the mean ISI did not change significantly. This means that once the neurons started firing, they fired with the same frequency as before but they fired with interruptions (gaps) after hypoxia, as displayed in **Figure 7**. The number of coincident spikes over the simulated time interval also decreased after hypoxia (**Figure 8D**). As in the experiments, the number of coincident spikes versus the harmonic mean of the number of spikes in the pair fit a quadratic function whose slope at any point was slightly lower after hypoxia, which indicates an overall trend across the population to decrease synchrony after hypoxia. A bootstrap analysis of the simulated data was applied to disentangle the amount of synchrony and synchrony changes that can be explained by chance from the contribution due to the network architecture and its modification following hypoxia. The bootstrap-corrected synchrony is shown in **Figure 9**. As in the experimental data, neuronal synchronization can increase or decrease after hypoxia, but the decrease (58% of 4950 pairs) was much more pronounced across all pairs than the increase (38%). Also 4% of pairs did not show significant synchrony before or after hypoxia. As a result, the histogram of the synchrony change was significantly skewed to the left ($p < 0.05$, Wilcoxon sign rank test).

DISCUSSION

In our analysis of ensemble recordings of inspiratory premotor neurons in the rVRG, we applied state-of-the-art statistical tools and obtained several novel results. First, subpopulations of pre-motor neurons had synchronized spike activity indicating a common drive from upstream inspiratory neurons. Second, this synchrony was malleable. Physiological stimuli modulated synchrony: it decreased in most pairs but increased in others, suggesting alterations of the functional network connectivity. The synchrony changes are independent of changes in the firing rate. Third, whereas the intraburst firing frequency of the premotor

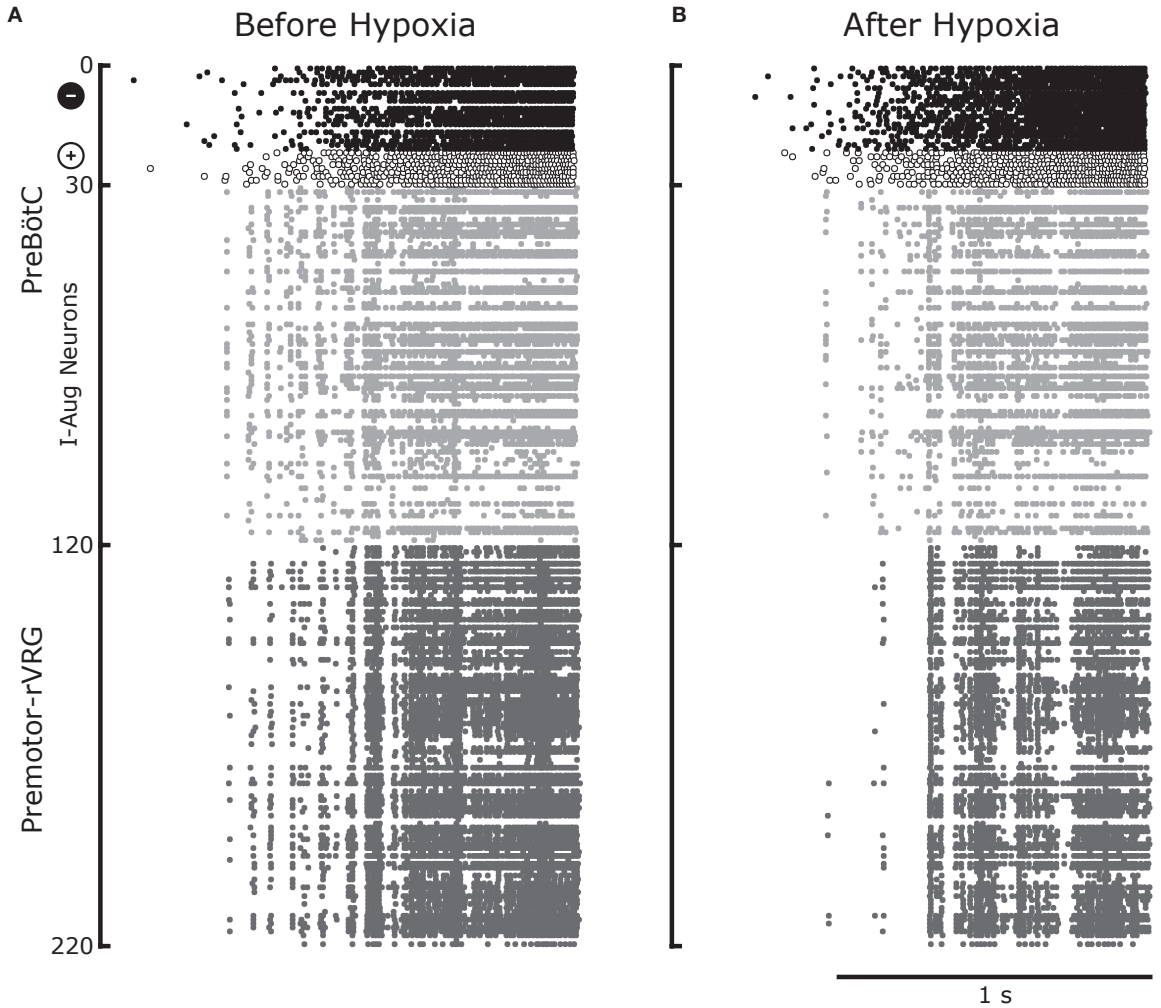


FIGURE 7 | Illustrative example of output from the computational model for an entire inspiratory burst. Before hypoxia **(A)** and after hypoxia **(B)** there was a recruitment of inhibitory pre-inspiratory preBötC neurons which resulted in enhanced inhibition of preBötC inspiratory-augmenting neurons, some of which were turned off. This led to a reduction of common inputs to the rVRG premotor neurons which in turn reduced the firing and altered spike synchronization among these neurons. The sudden termination of the burst is consistent with the experimental data of inspiratory neurons shown in **Figures 2C,D.**

inspiratory neurons was similar before and after hypoxia, the total number of spikes fired decreased significantly. Thus, the firing threshold of the pre-motor neurons may have been unaffected poststimulus; rather the neurons received less net excitation after hypoxia. Fourth, a simple computational model of inspiratory neurons in the ventral respiratory column reproduced efficiently the experimental observations. The model consisted of a feed-forward network with three layers: the pre-inspiratory cells, the inspiratory-augmenting neurons, and the pre-motor neurons. To qualitatively reproduce the experimental results, a key element of the model was that not all the inspiratory-augmenting neurons were equivalent: some projected onto many pre-motor neurons (hubs), while most projected onto just a few.

We should emphasize, however, that there is currently no direct evidence for the network architecture hypothesized in our studies. Consequently, besides common fluctuating inputs there may be other connectivity patterns contributing to the short timescale synchronization observed in our experiments. Moreover, there is no direct evidence for a selective increase in feed-forward inhibition as opposed to depression of excitatory inputs from the pre-inspiratory population after hypoxia. The assumptions used in our computational model should rather be considered as a proof of principle for plausible features of the respiratory column to be investigated in future.

How robust is our model? We have quantified how excitation and inhibition from the pre-inspiratory population affect the results of the computational model (data not shown). The firing rate, synchrony values and their changes after hypoxia were comparable with the experimental results when excitation was between 25% and 45%, keeping other neuron (threshold, etc.) and network parameters (synaptic strengths, connection probabilities, etc.) unchanged. This effective range may well vary as those parameters change. As a result, the validity of our model should not be limited when the actual ratio of excitation and inhibition is experimentally determined.

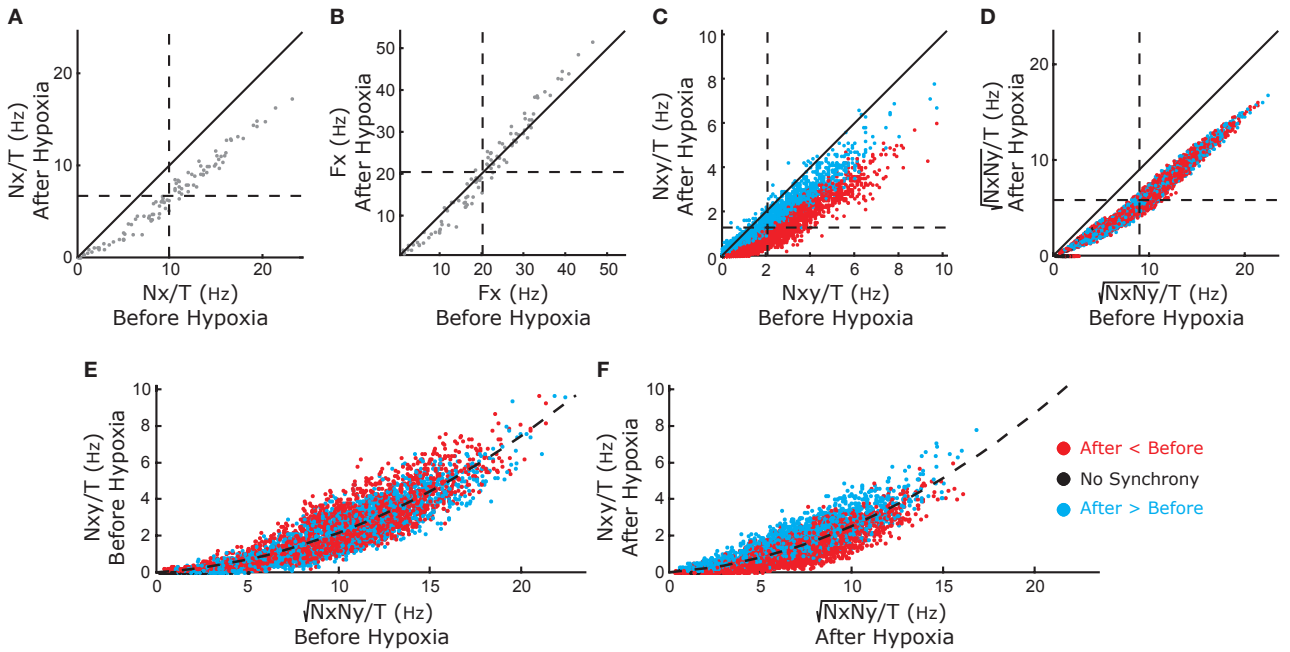


FIGURE 8 | Changes in activity produced by the model simulating inspiratory premotor neurons before and after hypoxia. These graphs show the relation of simulated activity before and after hypoxia as in **Figure 3** for the recorded data. In **(A–F)**, the synthetic data replicates the relationships in the original data. The color coding of the points is consistent across panels.

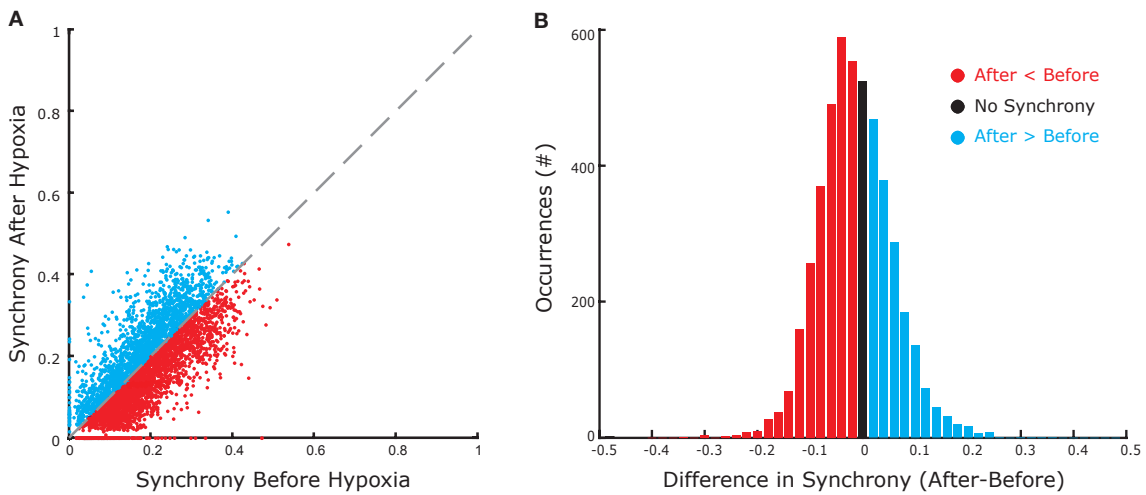


FIGURE 9 | *In silico* data showing increases and decreases in synchrony before and after hypoxia. (A,B) As in the recorded data, the magnitude of bootstrap-corrected synchrony decreased after hypoxia (number of red dots > blue dots).

Our analysis of multielectrode recordings has similarities and differences with respect to a previously published method, gravity analysis (Lindsey et al., 1997; Lindsey and Gerstein, 2006) which has been used to detect moments of synchrony in ensemble recordings. In gravity analysis, each neuron is represented as a particle in an *N*-dimensional space, *N* being the total number of neurons recorded. At the beginning of the recording, neurons are equidistant to each other. As time progresses, when two neurons fire within a time window δ, they attract each other, analogously to the gravitational force between two bodies, and move towards each other. In addition, there is an ongoing friction force quantified by a parameter γ that acts on every particle (neuron). The dissipative forces slow down the particles by opposing the attractive forces, which facilitates the agglomeration of particles representing neurons that tend to fire synchronously. At the end of the recording, clusters in the *N*-dimensional space identify assemblies of synchronized neurons. If δ and γ are properly chosen, our clustering analysis based on our synchrony measure (**Figure 4**) yields similar results

as the gravity analysis (data not shown). However, our analysis is computationally more efficient. In addition, our analysis does not require the parameter γ which has to be chosen heuristically in gravity analysis with an appropriate choice being crucial for the algorithm to work.

What is the physiological relevance of malleable synchronization in the respiratory column? Neuronal synchronization can transmit and process information at a low metabolic cost (Buzsaki, 2006). Action potentials are metabolically expensive, therefore, the implementation of firing rate codes for every purpose of brain function can be inefficient. In contrast, for a fixed and relatively low firing rate, neurons can synchronize their action potentials leading to a constructive summation of postsynaptic currents in downstream neurons or muscle fibers. The following two assumptions are implicit in this argument: (1) projections from neurons that are capable of synchronizing their spikes converge onto the same downstream target; (2) the postsynaptic currents are integrated sufficiently fast (within a few milliseconds), so that downstream targets can sum independent coincident spikes. Interestingly, both requirements are met across several areas of the brain including the olfactory bulb (Lagier et al., 2004; Galán et al., 2006b) hippocampus (Buzsaki, 2002; Vida et al., 2006), cerebellum (de Solages et al., 2008; Middleton et al., 2008) and neocortex (Hasenstaub et al., 2005). Thus synchrony across pre-motor and motor neurons in the respiratory control network may be an efficient way of regulating diaphragm contraction. Our data and simulations showing significant spike synchrony among inspiratory pre-motor neurons strongly support this idea. However, to validate this hypothesis further, the changes in neuronal synchronization across pre-motor neurons should co-vary with the shape and pattern of the phrenic nerve signal, i.e., the motor output of the network.

In our analysis, we have focused on the dynamics of inspiratory pre-motor neurons. There are two reasons why we focused specifically on these neurons and not on expiratory neurons. The first reason is a convenient anatomical feature: inspiratory pre-motor neurons are localized in a relatively well segregated area that facilitates the simultaneous recording from many of them. The second and most important reason, is that they are part of the effector output of the network that ultimately control the motoneuron activity and hence, the phrenic output and diaphragm. Whereas we have shown that spike synchronization within inspiratory pre-motor populations is significant and malleable, we surmise that the neuronal dynamics of expiratory populations along the ventral respiratory column are crucial to explain the natural variability of the respiratory rhythm. This will be investigated in future work with the techniques and analyses reported here.

ACKNOWLEDGMENTS

We are grateful to the reviewers for their valuable feedback. We thank Abigail Zaylor and Greg van Lunteren for their excellent technical assistance. This work has been supported by The Mount Sinai Health Care Foundation (RFG), The Alfred P. Sloan Foundation (RFG), the National Institutes of Health (HL-090554, TED) and the American Heart Association (SDG 0735037N, DMB).

REFERENCES

Baekey, D. M., Dick, T. E., and Galán, R. F. (2009). Spike synchronization is population specific in the respiratory pattern generator. *BMC Neurosci.* 10, P248. doi: 10.1186/1471-2202-10-S1-P248.

Baekey, D. M., Morris, K. F., Gestreau, C., Li, Z., Lindsey, B. G., and Shannon, R. (2001). Medullary respiratory neurones and control of laryngeal motoneurones during fictive eupnoea and cough in the cat. *J. Physiol. (Lond.)* 534, 565–581.

Bou-Flores, C., and Berger, A. J. (2001). Gap junctions and inhibitory synapses modulate inspiratory motoneuron synchronization. *J. Neurophysiol.* 85, 1543–1551.

Butera, R. J. Jr, Rinzel, J., and Smith, J. C. (1999a). Models of respiratory rhythm generation in the pre-Botzinger complex. II. Populations Of coupled pacemaker neurons. *J. Neurophysiol.* 82, 398–415.

Butera, R. J. Jr, Rinzel, J., and Smith, J. C. (1999b). Models of respiratory rhythm generation in the pre-Botzinger complex. I. Bursting pacemaker neurons. *J. Neurophysiol.* 82, 382–397.

Buzsaki, G. (2002). Theta oscillations in the hippocampus. *Neuron* 33, 325–340.

Buzsaki, G. (2006). *Rhythms of the Brain.* New York: Oxford University Press.

Coles, S. K., and Dick, T. E. (1996). Neurones in the ventrolateral pons are required for post-hypoxic frequency decline in rats. *J. Physiol. (Lond.)* 497(Pt 1), 79–94.

de Solages, C., Szapiro, G., Brunel, N., Hakim, V., Isope, P., Buisseret, P., Rousseau, C., Barbour, B., and Lena, C. (2008). High-frequency organization and synchrony of activity in the Purkinje cell layer of the cerebellum. *Neuron* 58, 775–788.

Del Negro, C. A., Johnson, S. M., Butera, R. J., and Smith, J. C. (2001). Models of respiratory rhythm generation in the pre-Botzinger complex. III. Experimental tests of model predictions. *J. Neurophysiol.* 86, 59–74.

Del Negro, C. A., Morgado-Valle, C., and Feldman, J. L. (2002). Respiratory rhythm: an emergent network property? *Neuron* 34, 821–830.

Dick, T. E., Dutschmann, M., and Paton, J. F. (2001). Post-hypoxic frequency decline characterized in the rat working heart brainstem preparation. *Adv. Exp. Med. Biol.* 499, 247–254.

Dick, T. E., Hsieh, Y. H., Morrison, S., Coles, S. K., and Prabhakar, N. (2004). Entrainment pattern between sympathetic and phrenic nerve activities in the Sprague-Dawley rat: hypoxia-evoked sympathetic activity during expiration. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 286, R1121–R1128.

Ermentrout, G. B., Galan, R. F., and Urban, N. N. (2008). Reliability, synchrony and noise. *Trends Neurosci.* 31, 428–434.

Feldman, J. L., McCrimmon, D. R., and Speck, D. F. (1984). Effect of synchronous activation of medullary inspiratory bulbo-spinal neurones on phrenic nerve discharge in cat. *J. Physiol. (Lond.)* 347, 241–254.

Galán, R. F., Ermentrout, G. B., and Urban, N. N. (2006a). Predicting synchronized neural assemblies from experimentally estimated phase-resetting curves. *Neurocomputing* 69, 1112–1115.

Galán, R. F., Fourcaud-Trocmé, N., Ermentrout, G. B., and Urban, N. N. (2006b). Correlation-induced synchronization of oscillations in olfactory bulb neurons. *J. Neurosci.* 26, 3646–3655.

Galán, R. F., Ermentrout, G. B., and Urban, N. N. (2007a). Stochastic dynamics of uncoupled neural oscillators: Fokker-Planck studies with the finite element method. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76, 056110.

Galán, R. F., Ermentrout, G. B., and Urban, N. N. (2007b). Reliability and stochastic synchronization in type I vs. type II neural oscillators. *Neurocomputing* 70, 2102–2106.

Galán, R. F., Ermentrout, G. B., and Urban, N. N. (2008). Optimal time scale for spike-time reliability: theory, simulations, and experiments. *J. Neurophysiol.* 99, 277–283.

Hasenstaub, A., Shu, Y., Haider, B., Kraushaar, U., Duque, A., and McCormick, D. A. (2005). Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron* 47, 423–435.

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

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Trans. Neural Netw.* 15, 1063–1070.

Lagier, S., Carleton, A., and Lledo, P. M. (2004). Interplay between local GABAergic interneurons and relay neurons generates gamma oscillations in the rat olfactory bulb. *J. Neurosci.* 24, 4382–4392.

Lindsey, B. G., and Gerstein, G. L. (2006). Two enhancements of the gravity

algorithm for multiple spike train analysis. *J. Neurosci. Methods* 150, 116–127.

Lindsey, B. G., Morris, K. F., Shannon, R., and Gerstein, G. L. (1997). Repeated patterns of distributed synchrony in neuronal assemblies. *J. Neurophysiol.* 78, 1714–1719.

Middleton, S. J., Racca, C., Cunningham, M. O., Traub, R. D., Monyer, H., Knopfel, T., Schofield, I. S., Jenkins, A., and Whittington, M. A. (2008). High-frequency network oscillations in cerebellar cortex. *Neuron* 58, 763–774.

Morgado-Valle, C., Baca, S. M., and Feldman, J. L. (2010). Glycinergic pacemaker neurons in preBotzinger complex of neonatal mouse. *J. Neurosci.* 30, 3634–3639.

Paton, J. F. (1996). A working heart-brainstem preparation of the mouse. *J. Neurosci. Methods* 65, 63–68.

Powell, F. L., Milsom, W. K., and Mitchell, G. S. (1998). Time domains of the hypoxic ventilatory response. *Respir. Physiol.* 112, 123–134.

Rekling, J. C., Shao, X. M., and Feldman, J. L. (2000). Electrical coupling and excitatory synaptic transmission between rhythmogenic respiratory neurons in the preBotzinger complex. *J. Neurosci.* 20, RC113.

Rybak, I. A., O'Connor, R., Ross, A., Shevtsova, N. A., Nuding, S. C., Segers, L. S., Shannon, R., Dick, T. E., Dunin-Barkowski, W. L., Orem, J. M., Solomon, I. C., Morris, K. F., and Lindsey, B. G. (2008). Reconfiguration of the pontomedullary respiratory network: a computational modeling study with coordinated in vivo experiments. *J. Neurophysiol.* 100, 1770–1799.

Rybak, I. A., Shevtsova, N. A., Paton, J. F., Dick, T. E., St-John, W. M., Morschel, M., and Dutschmann, M. (2004). Modeling the ponto-medullary respiratory network. *Respir. Physiol. Neurobiol.* 143, 307–319.

Schreiber, S., Fellous, J. M., Tiesinga, P., and Sejnowski, T. J. (2004). Influence of ionic conductances on spike timing reliability of cortical neurons for suprathreshold rhythmic inputs. *J. Neurophysiol.* 91, 194–205.

Schwarzacher, S. W., Smith, J. C., and Richter, D. W. (1995). Pre-Botzinger complex in the cat. *J. Neurophysiol.* 73, 1452–1461.

Solomon, I. C., Halat, T. J., El-Maghrabi, R., and O'Neal, M. H. 3rd (2001). Differential expression of connexin26 and connexin32 in the pre-Botzinger complex of neonatal and adult rat. *J. Comp. Neurol.* 440, 12–19.

Tonkovic-Capin, V., Stucke, A. G., Stuth, E. A., Tonkovic-Capin, M., Hopp, F. A., McCrimmon, D. R., and Zuperku, E. J. (2003). Differential processing of excitation by GABAergic gain modulation in canine caudal ventral respiratory group neurons. *J. Neurophysiol.* 89, 862–870.

Vida, I., Bartos, M., and Jonas, P. (2006). Shunting inhibition improves robustness of gamma oscillations in hippocampal interneuron networks by homogenizing firing rates. *Neuron* 49, 107–117.

Winter, S. M., Fresemann, J., Schnell, C., Oku, Y., Hirrlinger, J., and Hulsmann, S. (2010). Glycinergic interneurons in the respiratory network of the rhythmic slice preparation. *Adv. Exp. Med. Biol.* 669, 97–100.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

Received: 22 November 2009; paper pending published: 21 December 2009; accepted: 17 August 2010; published online: 15 September 2010.

Citation: Galán RF, Dick TE and Baekey DM (2010) Analysis and modeling of ensemble recordings from respiratory premotor neurons indicate changes in functional network architecture after acute hypoxia. Front. Comput. Neurosci. ***4****:131. doi: 10.3389/fncom.2010.00131*

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 31 March 2010
doi: 10.3389/fncom.2010.00002

A novel mechanism for switching a neural system from one state to another

Chethan Pandarinath[1], Illya Bomash[1], Jonathan D. Victor[2], Glen T. Prusky[1], Wayne W. Tschetter[1] and Sheila Nirenberg[1]*

[1] *Department of Physiology and Biophysics, Weill Cornell Medical College, Cornell University, New York, NY, USA,*
[2] *Department of Neurology and Neuroscience, Weill Cornell Medical College, Cornell University, New York, NY, USA*

Edited by:
Matthias Bethge, Max Planck Institute for Biological Cybernetics, Germany

Reviewed by:
Thomas Euler, University of Tübingen, Germany
Fred Rieke, University of Washington, USA
Guenther Zeck, Max Planck Institute of Neurobiology, Germany

****Correspondence:***
Sheila Nirenberg, Department of Physiology and Biophysics, Weill Cornell Medical College, Cornell University, 1300 York Avenue, New York, NY 10065, USA.
e-mail: shn2010@med.cornell.edu

An animal's ability to rapidly adjust to new conditions is essential to its survival. The nervous system, then, must be built with the flexibility to adjust, or shift, its processing capabilities on the fly. To understand how this flexibility comes about, we tracked a well-known behavioral shift, a visual integration shift, down to its underlying circuitry, and found that it is produced by a novel mechanism – a change in gap junction coupling that can turn a cell class on and off. The results showed that the turning on and off of a cell class shifted the circuit's behavior from one state to another, and, likewise, the animal's behavior. The widespread presence of similar gap junction-coupled networks in the brain suggests that this mechanism may underlie other behavioral shifts as well.

Keywords: gap junction, shunt, network shift, state change, adaptation, cable theory, horizontal cell, attention

INTRODUCTION

The nervous system has an impressive ability to self-adjust – that is, as it moves from one environment to another, it can adjust itself to accommodate the new conditions. For example, as it moves into an environment with new stimuli, it can shift its attention (Desimone and Duncan, 1995; Maunsell and Treue, 2006; Reynolds and Heeger, 2009); if the stimuli are low contrast, it can adjust its contrast sensitivity (Shapley and Victor, 1978; Ohzawa et al., 1982; Bonin et al., 2006); if the signal-to-noise ratio is low, it can change its spatial and temporal integration properties (Peskin et al., 1984; De Valois and De Valois, 1990). These shifts are well described at the behavioral level – and are clearly critical to our functioning – but how the nervous system is able to produce them is not clear. How is it that a network can change the way it processes information on the fly?

In this paper, we describe a case where it was possible to obtain an answer. It is a simple case, but one of the best-known examples of a behavioral shift – the shift in visual integration time that occurs as an animal switches from daylight to nightlight conditions (reviewed in De Valois and De Valois, 1990). In daylight conditions, when photons are abundant, and the signal-to-noise ratio is high, the visual system is shifted toward short integration times. In nightlight conditions, when photons are limited, and the signal-to-noise ratio is low, the system shifts toward long integration times. (See Appendix 1 for why the shift involves a network action, rather than a simple switch from cones to rods.)

Here we propose a hypothesis for how the shift takes place – it involves a change in gap-junction coupling among the horizontal cells of the retina. The idea is as follows: Horizontal cells are well-known to be coupled by gap junctions, and the coupling is light-dependent (Dong and McReynolds, 1991; Xin and Bloomfield, 1999; Weiler et al., 2000). When light levels are high, the gap junctions close, and there is little coupling. When light levels are low, the gap junctions open, and extensive coupling ensues. Since coupling shunts current, the idea is that the extensive coupling causes a shunting of horizontal cell current, effectively taking the horizontal cells out of the system. Since horizontal cells play a key role in shaping integration time – they provide feedback to photoreceptors that keeps integration time short (Baylor et al., 1971; Kleinschmidt and Dowling, 1975; Smith, 1995) – taking these cells out of the system makes integration time longer.

This hypothesis raises a new, and potentially generalizable idea – that a neural network can be shifted from one state to another by changing the gap-junction coupling of one of its cell classes. The coupling can act as a means to take a cell class out of a network, and by doing so, change the network's behavior. (For more on generalization, including the time scale of the coupling changes, see Discussion.)

We tested the hypothesis using transgenic mice that cannot undergo this coupling (Hombach et al., 2004; Shelley et al., 2006). They lack the horizontal cell gap-junction gene, and, as a result, their horizontal cells get locked into the uncoupled state (Hombach et al., 2004; Shelley et al., 2006). If the hypothesis is correct, these animals should not be able to undergo the shift to long integration times. Our results show that the hypothesis held: the shift was blocked completely at the behavioral level, and almost completely at the physiological (i.e., ganglion cell) level.

In sum, we tracked a behavioral change down to the neural machinery that implements it. This revealed a new, simple, and potentially generalizable, mechanism for how networks can rapidly adjust themselves to changing environmental demands.

RESULTS

Figure 1 gives the starting point for these experiments. It indicates that (a) the model system we are using, the mouse, shows the shift in visual integration time observed in other species (Kelly, 1961; van Nes et al., 1967; De Valois and De Valois, 1990; Umino et al., 2008) (**Figure 1A**), and (b) the part of the nervous system responsible for the shift, or at least a large part of it, is the retina, since the shift is readily detectable at the level of the retinal ganglion cells (**Figure 1B**). The shift at the behavioral level was measured using a standard optomotor task, where the stimuli were drifting sine wave gratings of different temporal frequencies. The shift at the ganglion cell level was measured using three different stimuli: drifting sine wave gratings of different temporal frequencies, a white noise stimulus, and a natural scene stimulus. As indicated in all the panels of the figure, there is a shift from short integration times to long, that is, from high temporal frequencies to low ($p < 10^{-3}$, *t*-test comparing the centers of mass of the frequency response curves for the night (scotopic) condition with those for the day (photopic) condition).

Figure 2 shows the proposed model for how the shift is generated. It builds on the well-established front-end circuit that shapes visual integration time (Baylor et al., 1971; Kleinschmidt and Dowling, 1975; reviewed in Dowling, 1987) (**Figure 2A**). The circuit contains three cell classes – photoreceptors, bipolar cells and horizontal cells – and operates, briefly, as follows: the photoreceptors send signals forward to both the bipolar and horizontal cells. The bipolar cells continue to send signals forward, while the horizontal cells send signals back onto the photoreceptors. The horizontal cell feedback shapes the photoreceptors' integration time[1] (Baylor et al., 1971; Kleinschmidt and Dowling, 1975).

Figure 2B shows how a change in the gap junction coupling of the horizontal cells can modulate the circuit's behavior – that is, how it can change it from one state to another. The scenario is the following: In daylight conditions the gap junctions close. This strengthens the signals of the horizontal cells, so they send strong feedback to the photoreceptors. Strong feedback cuts the photoreceptors' integration time short, producing the short integration times (high temporal frequency responses) observed experimentally (**Figure 2B**, left). In nightlight conditions, the gap junctions open. The opening produces a shunting of the horizontal cell current, which reduces or eliminates the horizontal cell signal. Without the feedback from the horizontal cells, there

[1]The integration time of the photoreceptor refers to the length of time over which it responds to light (i.e., the width of the impulse response).

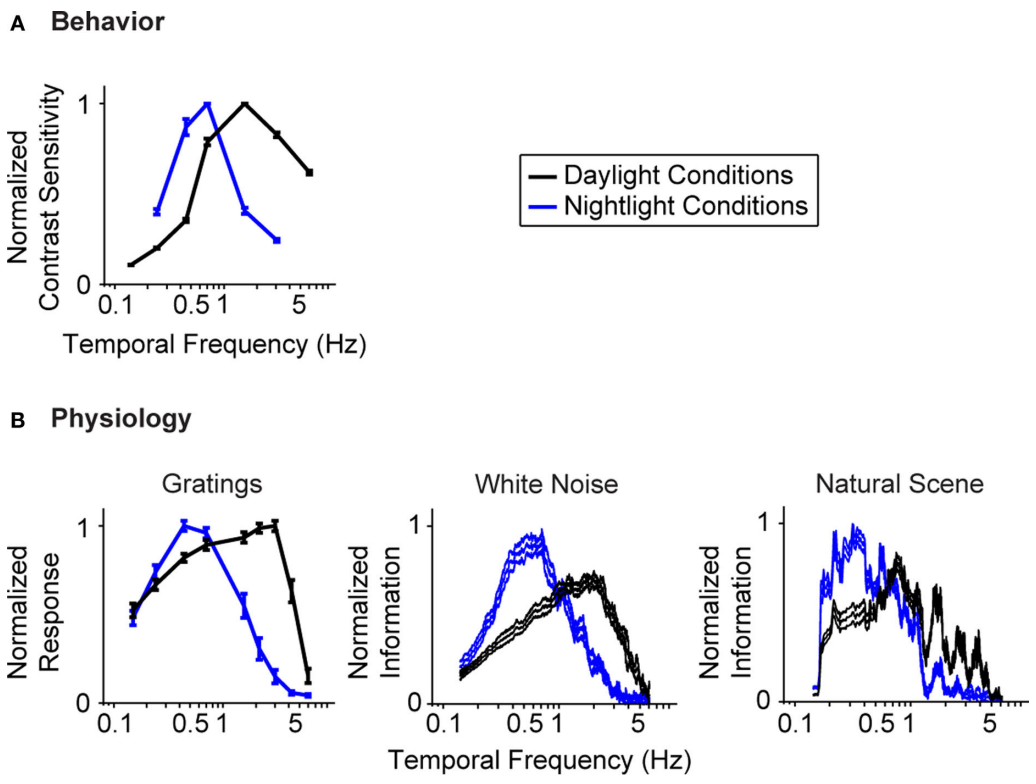


FIGURE 1 | The visual system undergoes a shift in integration time as it shifts from daylight to nightlight (photopic to scotopic) conditions. In daylight conditions, the system favors short integration times (high temporal frequencies); in nightlight conditions, it favors long integration times (low temporal frequencies). See Materials and Methods for light intensities for the two conditions. **(A)** The shift, measured at the behavioral level using drifting grating stimuli. **(B)** The shift, measured at the ganglion cell level, using three different kinds of stimuli: drifting gratings, white noise, and natural scenes. Behavioral performance was measured as contrast sensitivity, averaged across animals, and peak-normalized ($n = 5$, mean ± SEM). Ganglion cell performance in (**B**, *left*) was measured as first harmonic response, averaged across cells, and peak normalized; ganglion cell performance in (**B**, *middle and right*) was measured as information, normalized for equal area ($n = 20$, mean ± SEM).

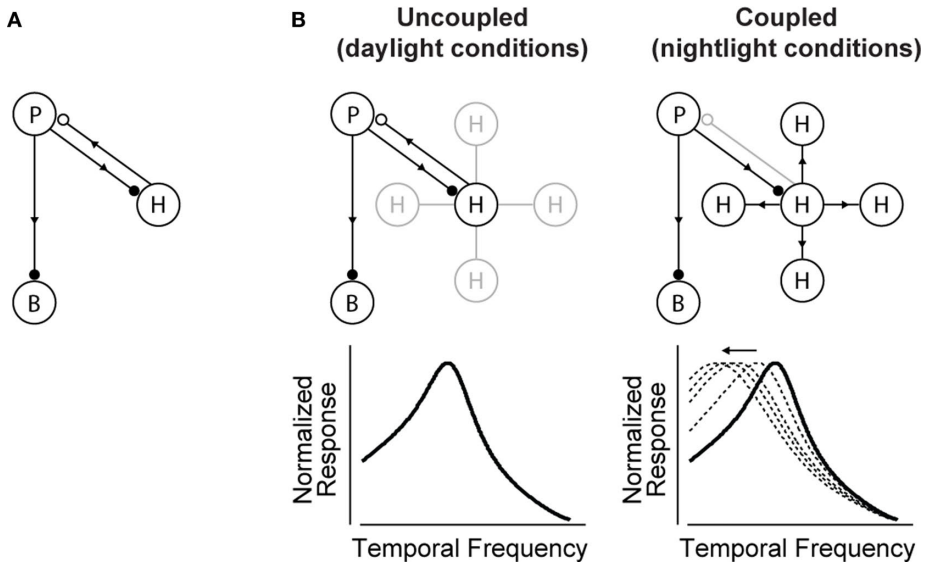


FIGURE 2 | The circuit that controls visual integration time can be shifted from one state to another by a change in the gap junction coupling of one of its cell classes. (A) Visual integration time is shaped, in large part, by a negative feedback loop in the outer retina: photoreceptors send signals forward to both bipolar cells and horizontal cells; the horizontal cells then, in turn, provide negative feedback to the photoreceptors (Baylor et al., 1971; Kleinschmidt and Dowling, 1975; Dowling, 1987). Note that the figure shows only one type of horizontal cell and a generic photoreceptor; this is consistent with our model system, the mouse retina, which has only one type of horizontal cell, and it acts on both rods and cones (Peichl and González-Soriano, 1994; Trumpler et al., 2008; Babai and Thoreson, 2009). (**B**, *left*) In daylight conditions, horizontal cell feedback is strong. This cuts photoreceptor integration time short, and the system shifts to high temporal frequency responses. At night (**B**, *right*), when the system needs longer integration times, a reduction in horizontal cell feedback is needed. The opening of the gap junctions provides a mechanism for achieving this. It produces a shunting of horizontal cell current that weakens or inactivates the horizontal cells. The photoreceptor integration time then becomes longer, and the system shifts to low temporal frequency responses. The change in the gap junction coupling acts, effectively, as a knob to regulate the strength of the negative feedback. (See Appendix 2 for a formalized version of the model.)

is no shortening of the photoreceptor integration time, and the system shifts to the observed long integration times (low temporal frequency responses) (**Figure 2B**, right).

The strength of the model is that it derives from well-established facts – specifically, that the integration time of photoreceptors (both rods and cones) changes (becomes extended) as an animal moves from day to night conditions (Kleinschmidt and Dowling, 1975; Daly and Normann, 1985; Schneeweis and Schnapf, 2000), that the strength of the horizontal cell signal changes (decreases) as the conditions move from day to night (Teranishi et al., 1983; Yang and Wu, 1989a), and, finally, that there is a change in the degree of horizontal cell coupling (an increase) with the change from day to night conditions (Dong and McReynolds, 1991; Xin and Bloomfield, 1999; Weiler et al., 2000). Put together, these facts lead to a mechanism for shifting the circuit's behavior. The novelty is the use of gap junction coupling as a shunting device (see Discussion) – the model makes use of the fact that coupling produces a shunt, and, therefore, has the capacity to weaken or inactivate a cell class. By casting the coupling as a shunting mechanism, the actions of the components of the circuit – the photoreceptors, the bipolar cells, the horizontal cells, and the light-dependent change in horizontal cell coupling – fall into place to explain how the system can shift from one state to another. A formalized version of the model is given in Appendix 2.

We test the proposal in **Figure 3**. To do this, we used a transgenic mouse line that cannot undergo horizontal cell coupling (Hombach et al., 2004; Shelley et al., 2006) (**Figure 3A**). These mice lack the gene for the gap junction specific to the horizontal cells, connexin 57 (Cx57), so their horizontal cells are locked into the uncoupled state. We emphasize that this particular gap junction gene is not expressed anywhere in the nervous system besides the horizontal cells (Hombach et al., 2004); thus, the elimination of this gene produces a very specific perturbation. **Figure 3B** shows the temporal integration curves from wild-type and knockout mice in the night condition, measured both at the behavioral level and at the ganglion cell level with the three stimuli used in **Figure 1**. In all cases, the shift to long integration times was impaired, that is, the normal increase in amplitude at low frequencies, and the normal decrease in amplitude at high frequencies did not occur (**Figure 3B**) or was significantly hindered (**Figure 3C**) [$p < 10^{-4}$ for the behavior, $p < 10^{-3}$ for the ganglion cell responses, *t*-test comparing the centers of mass of the frequency response curves for the night (scotopic) condition with those for the day (photopic) condition].

The robustness of the results is demonstrated in **Figure 3D**. Using data that allow a direct comparison to be made between behavioral and ganglion cell results, specifically, where the results were obtained using the same stimuli – the drifting sine wave gratings – we show the complete set of individual responses. The left side of **Figure 3D** shows the behavioral performance for all animals under day and night conditions, and the right side shows the performance for all ganglion cells under day and night conditions. As shown in the figure, by day, the performance of the knockout closely matches that of the wild-type, but at night, the two

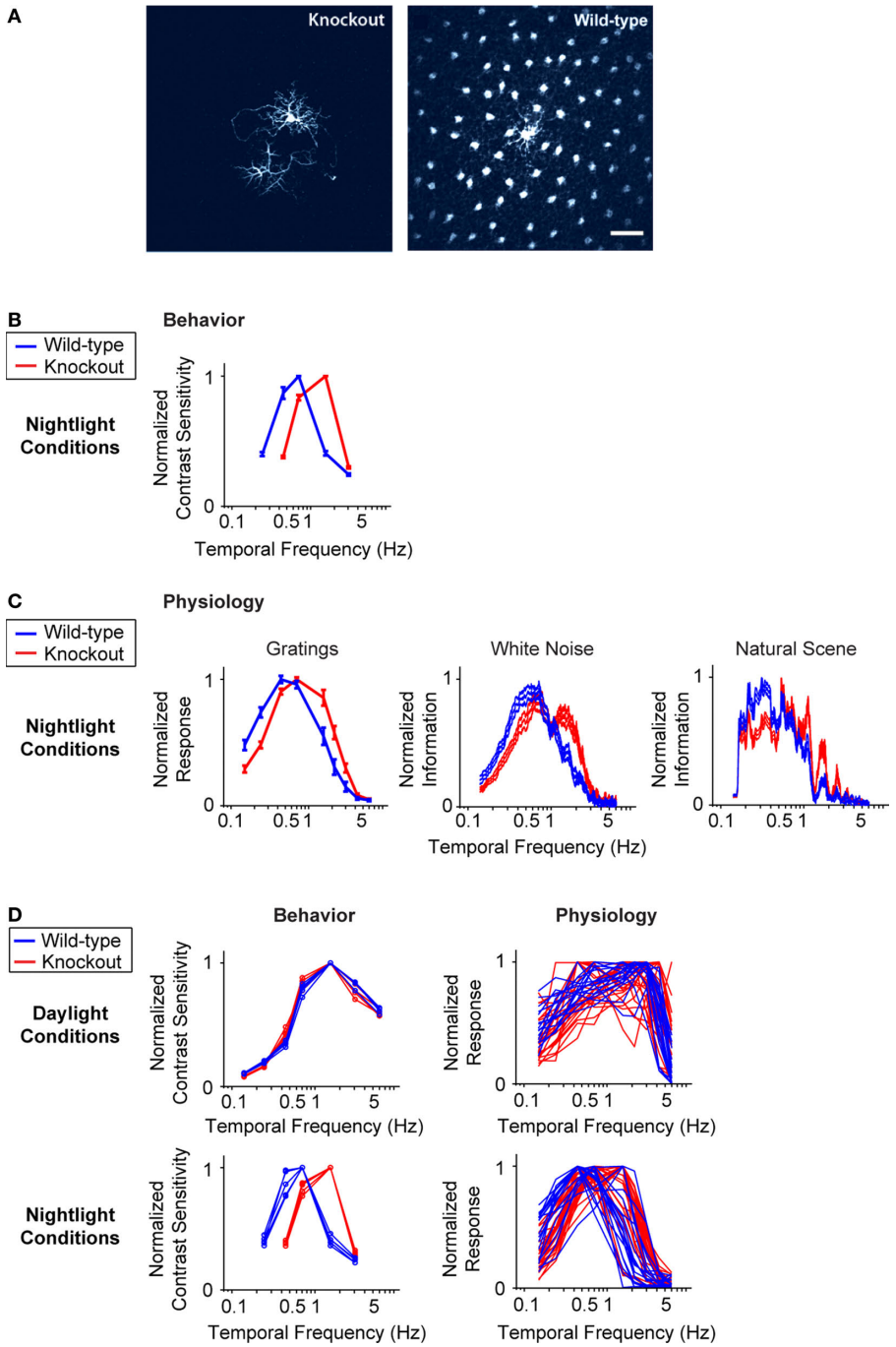


FIGURE 3 | When horizontal cell coupling is prevented, the shift to long integration times is impaired at both the behavioral level and the ganglion cell level. (A) Horizontal cell coupling in a retina from a Cx57 knockout versus horizontal cell coupling in a retina from a wild-type sibling control. In each retina, a single horizontal cell was injected with dye, and the extent of dye spread was measured for >1 h. Consistent with the results in (Hombach et al., 2004; Shelley et al., 2006; Dedek et al., 2008), coupling is abolished. Scale bar = 50 μm. **(B)** Behavioral performance curves measured from Cx57 knockouts and wild-type sibling controls under the night condition. The shift to long integration times (low temporal frequency responses) is significantly impaired ($p < 10^{-4}$). **(C)** Ganglion cell performance curves measured from Cx57 knockout animals and wild-type sibling controls under the night condition. As in **(B)**, the shift to long integration times is significantly impaired ($p < 10^{-3}$). All measurements were taken as in **Figure 1**; for the behavioral experiments, $n = 5$ wild-type mice, 5 knockout mice, and for the ganglion cell measurements, $n = 20$ cells from wild-type retinas, 24 cells from knockouts. **(D)** *Left*, performance for all animals shown individually. In daylight conditions, the performances of the knockouts are essentially identical to those of the wild-type animals. In night conditions, they diverge: the wild-type animals make the shift toward longer integration times, while the knockouts do not. *Right*, performance for all ganglion cells. Similar to plots on the left, the performances of the ganglion cells from the knockout and wild-type animals are the same in daytime conditions but diverge at night: the ganglion cells from the wild-type animals undergo the shift toward longer integration times, while those from the knockout are left behind.

performances diverge. At night, the wild-type makes the expected shift toward longer integration times, but the knockout – which lacks horizontal cell coupling – does not.

DISCUSSION

The nervous system faces a shifting problem. It has to shift its mode of operation from one state to another as it faces new demands (i.e., it has to shift its attention, its contrast sensitivity, its temporal integration time, etc.). How it achieves this isn't clear. Here we examined a case where it was possible to obtain an answer, and the answer was intriguingly simple: the system produced the shift by changing the gap junction coupling of one of its cell classes. The coupling acted as a way to inactivate the cell class, and, by doing so, change the system's behavior.

The findings are both surprising and exciting: surprising, because a seemingly complicated problem was solved with a simple mechanism, and exciting, because the mechanism is present not just in the retina, but throughout the brain, suggesting it might generalize to other network shifts. To be specific, gap junction coupled networks are present in visual cortex, motor cortex, frontal cortex, hippocampus, cerebellum, hypothalamus, and striatum, among many other places (Galarreta and Hestrin, 1999, 2001; Bennett and Zukin, 2004).

Furthermore, a regulator is also in place. In the retina, the regulator is a neuromodulator, dopamine: Light triggers the release of dopamine, which closes gap junctions via second messengers (McMahon et al., 1989; Dong and McReynolds, 1991; Weiler et al., 2000). Dopamine, as well as noradrenaline and histamine, have been found to open and close gap junctions in several of these brain areas (Cepeda et al., 1989; Yang and Hatton, 2002; Onn et al., 2008; Zsiros and Maccaferri, 2008).

The possibility for generalization to other networks is substantial and straightforward to see:

(1) While the results in this paper show the mechanism in non-spiking neurons, it readily applies to spiking cells as well and thus to networks in the brain. This is because the mechanism involves only basic biophysics – a change in cells' input resistance. Briefly, if a cell class is coupled by gap junctions, it has the potential to have its input resistance turned up and down. When the junctions are closed, the input resistance of the cells is high. This makes the cells more responsive to incoming signals and allows them to send strong signals out. When the junctions are opened, the input resistance drops. This makes the cells *less* responsive to incoming signals and allows them to send out only weak signals. In the case of spiking neurons, the signals can become so weak that the probability of firing can be reduced essentially to 0; i.e., the cells can be effectively turned off.

(2) The mechanism has the potential to affect many types of network operations. While the one presented in this paper was a negative feedback loop – the gap junction coupling provided a way to turn the feedback on or off (or up or down) – one can readily imagine many other operations that could be altered by turning the activity of a pivotal cell class in a network on or off, such as alterations in feedforward signaling, lateral signaling, recurrent signaling (e.g., the stabilization of attractors), to name a few.

(3) The timescale over which the mechanism operates, that is, the timescale over which the change in coupling occurs – a scale of seconds (McMahon et al., 1989; McMahon and Mattson, 1996) – is consistent with many state changes, such as changes in arousal, changes in attentional set, shifts in decision-making strategies, e.g., shifts in the weighting of priors, shifts to speed versus accuracy (Standage and Paré, 2009), allowing it to mediate many behavioral processes.

(4) Since the cellular machinery for regulation of gap junction conductances is in place, the mechanism can evolve via a change in a single gene, a gene for a gap junction protein. This makes it an easy gain from an evolutionary standpoint. A powerful selective advantage – the ability to shift a network from one state to another – could be rapidly acquired, and, in addition, acquired independently in multiple networks. (For a review of gap junction proteins, see Bennett and Zukin, 2004.)

Figure 4 emphasizes this latter point, that this gap junction coupling mechanism offers a single gene solution to a seemingly complicated set of problems, network state changes. To address this, we used, again, the horizontal cells, as an example. Specifically, we took the behavioral results from the wild-type and Cx57 knockout animals and imposed them on a predator-detection scenario. We filmed an approaching predator, restricting the movies to the temporal frequencies available to each genotype, as indicated in **Figure 3D** left. The results are shown in **Figure 4**. In day conditions the movies for the two genotypes are essentially the same; the predator can be seen when it is moving, i.e., when the movie is dominated by high temporal frequencies, and when it is still, i.e., when the movie is dominated by low temporal frequencies. In contrast, in night conditions, the movies diverge. In the movie filtered through the frequencies visible to the wild-type animal, the predator remains visible even when it is still; this is consistent with the wild-type's ability to shift to low temporal frequencies. In the movie filtered though the frequencies visible to the knockout, the predator disappears. Only a ghost is present (see Supplementary Material for the complete movies). The wild-type's maintenance of visual contact with the predator gives it an obvious selective advantage.

ESTIMATING THE EXTENT TO WHICH INPUT RESISTANCE CAN BE REDUCED BY COUPLING

As discussed above, changes in coupling can act as a dial to turn the input resistance of a cell up or down. We can estimate the range of the dial as follows: The standard experimental measure of coupling is the length constant (Xin and Bloomfield, 1999; Shelley et al., 2006). Xin and Bloomfield measured the length constant of horizontal cells under several scotopic and photopic light levels and found the maximal difference to be a factor of ~3. The maximal difference occurred when the scotopic light level was 1–1.5 log units above rod threshold and the photopic light level was >3 log units above rod threshold, levels that we matched for this paper. Since, for 2-D coupling (Lamb, 1976), input resistance is inversely proportional to the square of the length constant (detailed in Materials and Methods and Appendix 2), the input resistance of the horizontal cells at the scotopic light level is estimated to be about a factor of 9 less than that at the photopic light level.

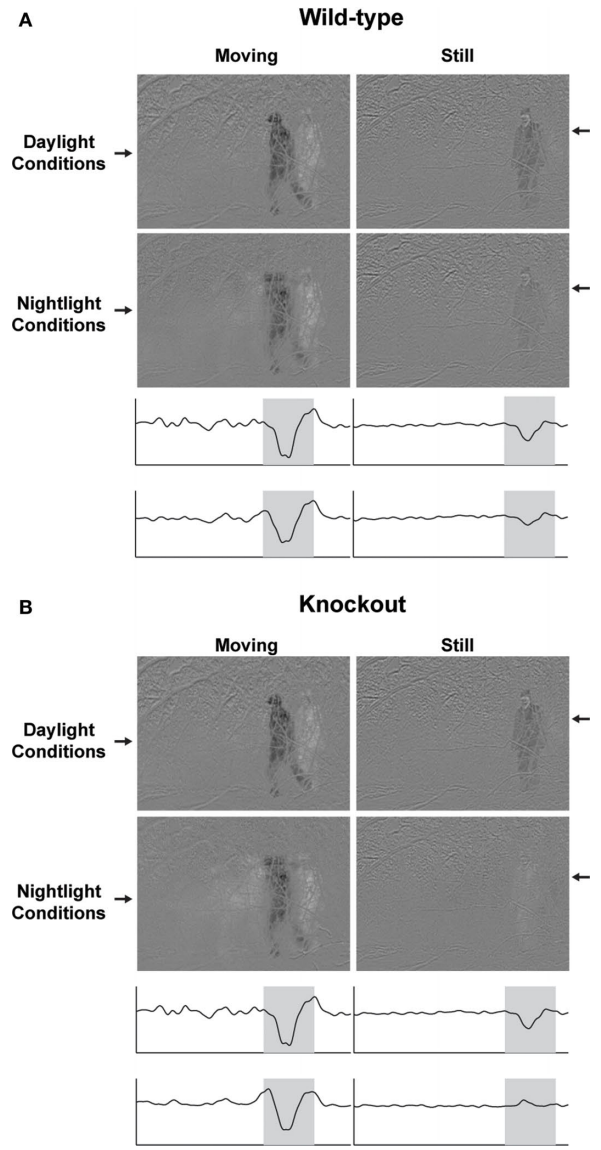


FIGURE 4 |The selective disadvantage of a Cx57 gene loss. (A) Movie of an approaching predator, filtered through the frequencies available to the wild-type animal, as provided by **Figure 3D** left. In day conditions, the predator can be seen both when it is moving (when the movie is dominated by high temporal frequencies), and when it is still (when the movie is dominated by low temporal frequencies). In night conditions, the signal is weaker, but the predator can still be seen both when moving or still. The visibility in the still condition is possible because of the shift to low temporal frequencies that occurs in the dark. The traces below the figures provide the intensity of each pixel in a horizontal slice through the image; the location of the slice is indicated by the arrow. **(B)** Same movie, filtered through the frequencies available to the knockout animal. In night conditions, the predator vanishes, see trace below figures. The wild-type's continued visual detection of the predator gives it an obvious selective advantage. (For the frequencies available to each genotype, see **Figure 3D** left: specifically, the range of frequencies seen by the knockout at night (red curves in 3D bottom left) is a subset of the range seen by the wild-type (blue curves); the lack of low frequency sensitivity in the knockout (below ~0.3 Hz) causes the predator, when it is still, to disappear. Note that the temporal filtering was applied to the entire movie; only a representative frame from each filtered version is shown here. For the complete filtered versions, see Supplementary Material).

In the general case, as with horizontal cells, the extent to which gap junction coupling can shunt a cell is the ratio of the total conductances of the gap junctions that can be modulated, to the cell's baseline ("leak") conductances. Many factors – including the cell's geometry and the complement and distribution of channels and gap junctions – combine to determine this ratio. The example of horizontal cells shows that this can be as much as an order of magnitude.

LINKING A BEHAVIOR TO A NEURAL MECHANISM

Following a behavioral change down to the mechanism that underlies it is often not possible experimentally. It was possible here because of a confluence of factors: the relevant network could be identified and its component cell classes are known (as shown in **Figures 1 and 2**), and the protein around which the mechanism revolves, the particular gap junction protein, Cx57, is present only in one cell class (the horizontal cells) and not elsewhere in the brain (Hombach et al., 2004), allowing the circuit to be selectively disrupted. The significance of the latter is that it allowed a direct connection to be made between the disruption in the circuit and the disruption in the behavior, since no other circuits were perturbed.

POTENTIAL ALTERNATIVE MODELS FOR THE SHIFT TOWARD LOW TEMPORAL FREQUENCIES

As an animal moves from a light-adapted to a dark-adapted state, several changes occur in the retina other than the change in horizontal cell coupling via the Cx57 gap junctions. How can we be sure that our result – the shift toward low temporal frequencies – is not produced by these other changes? Here we systematically go through them.

The most well known change is the shift from cone to rod photoreceptors. This can't account for our results, because the knockout undergoes the same cone-to-rod shift, and it doesn't undergo the shift to low frequencies (**Figure 3**). In addition, it's well known that the cone-to-rod shift affects high frequencies, not low. We show this in Appendix 1, **Figure 5**, specifically for our species, the mouse. As shown in the figure, the frequency response curves for the rod and cone are both flat below 0.5 Hz, meaning there is no frequency-dependent change in this region. In contrast, our results show a selective boost at frequencies below 0.5 Hz; that is, the system shifts to favor low frequencies. The shift from cones to rods can't account for this.

Another change that occurs during dark adaptation is rod–cone coupling (see Ribelayga et al., 2008, for rod–cone coupling as a result of circadian rhythms; also Yang and Wu, 1989b; Wang and Mangel, 1996; Trumpler et al., 2008). Rod–cone coupling, though, is mediated by gap junctions formed by Cx36, Cx35, and Cx34.7 (reviewed in Li et al. (2009)), not Cx57 (Janssen-Bienhold et al., 2009). Cx57 is not present in rods and cones (Hombach et al., 2004; Janssen-Bienhold et al., 2009) and thus the knockout is not perturbing these couplings.

Similarly, gap junction coupling in the inner retina likely plays a role in dark adaptation, since the AII amacrine cells of the rod pathway are coupled by gap junctions (Bloomfield et al., 1997). However, Cx57 is not a gap junction in these cells (Janssen-Bienhold et al., 2009), so changes in inner retinal coupling can not account for our results.

Recent reports have indicated that some gap junctions act as hemichannels (Kamermans et al., 2001; Shields et al., 2007). If Cx57 acted in this fashion, it could provide for ephaptic transmission of a feedback signal. However, the possibility that Cx57 is a hemichannel has been examined at the ultrastructural level, and ruled out (Janssen-Bienhold et al., 2009). Furthermore, feedback to photoreceptors has been shown to be intact in the Cx57 knockout by two groups (Shelley et al., 2006; Dedek et al., 2008).

Finally, a standard concern with most or all knockout experiments is that knocking out a gene could lead to secondary developmental effects. While we can't completely rule this out, there is no evidence for altered development in the Cx57 knockout: retinal anatomy appears unperturbed (Hombach et al., 2004; Shelley et al., 2006), temporal tuning by day, as measured at the ganglion cell and behavioral level, remains intact, i.e., is the same as in wild-type (**Figure 3D**), and spatial processing, also measured at the ganglion cell and behavioral level, remains intact as well (Dedek et al., 2008). While compensatory effects are possible, the likelihood that they would lead to such close matches along all these axes is very low.

Thus, while cone-to-rod shifts, photoreceptor coupling, and other factors contribute to dark adaptation, they can't account for the results presented here, and the probability that the results could be accounted for by developmental effects, as mentioned above, is very low.

One issue that we can't completely rule out, though, is the following: even though horizontal cell feedback to photoreceptors is known to be present and can account for our results, we can't completely rule out the possibility that the shunting of horizontal cell current causes the shift in tuning through some other action. For example, if horizontal cells were to act as a mediator between multiple circuits with different kinetics (e.g., different photoreceptor readout circuits), then the shunting of the horizontal cell current could shift tuning by causing a switch from one circuit to another. But note that any alternative model must be consistent with the known constraints: (a) the difference between wild-type and knockout is present under scotopic conditions (**Figure 3**), where all responses are rod-driven, (b) the tuning shift involves low frequencies, (c) the mouse retina has only one kind of horizontal cell, and it serves both kinds of photoreceptors, and (d) connexin-57 is only involved in horizontal cell-to-horizontal cell coupling. We chose the horizontal cell feedback model shown in **Figure 2** because it is a parsimonious model that satisfies these constraints and is consistent with current known actions of horizontal cells.

We conclude by mentioning that in one species (the rabbit), when light levels are much lower, more than an order of magnitude below the scotopic level used in this study, gap junctions close (Xin and Bloomfield, 1999) with no corresponding reversal of the shift in integration times (Nakatani et al., 1991). This suggests that in this extreme range, other mechanisms must take over, mechanisms likely intrinsic to the photoreceptors, as described in Tamura et al. (1989).

RELATION OF CX57 TO SPATIAL PROCESSING IN THE DARK- AND LIGHT-ADAPTED CONDITIONS

Horizontal cells provide negative feedback to photoreceptors (Werblin and Dowling, 1969) and antagonistic feedforward to bipolar cells (Yang and Wu, 1991), and it has long been thought that they contribute to the receptive field surround. One might expect, therefore, that eliminating coupling in these cells would alter spatial processing as well as temporal processing as the retina shifts from day to night vision. A previous study, though, shows that spatial tuning remains normal in the Cx57 knockout (Dedek et al., 2008). The likely basis for this is the fact that the surround is generated by circuits in more than one layer – specifically, by amacrine cell circuits in the inner retina, as well as by horizontal cells in the outer retina (Cook and McReynolds, 1998; Taylor, 1999; Roska et al., 2000; Flores-Herr et al., 2001; McMahon et al., 2004; Sinclair et al., 2004). As mentioned in Dedek et al. (2008), the lack of a change in spatial tuning in the knockout implies that inner retinal mechanisms dominate for the problem of adjusting spatial tuning to different light-adaptation levels.

COUPLING AS A MECHANISM TO PRODUCE SYNCHRONY

We conclude by mentioning that gap junction coupling has also been proposed as a mechanism to create synchronous firing among neurons, e.g., for creating oscillations (for review, see Bennett and Zukin, 2004). The idea presented in this paper – that changes in coupling serve as a way to inactivate a cell class or reduce its impact – is not mutually exclusive with this proposal. This is because the effect of coupling depends on the state of the cell. As mentioned above, when a cell becomes coupled to other cells, its input resistance drops. For spiking neurons, this means the probability of reaching threshold and firing is reduced. If, however, the cell receives strong enough input to allow it to cross threshold, its firing can produce synchronous spikes in coupled cells. Thus, gap junction coupling can potentially mediate more than one network operation.

MATERIALS AND METHODS

ANIMALS

Generation of the Cx57-deficient mouse line was previously reported (Hombach et al., 2004). Briefly, part of the coding region of the *Cx57* gene was deleted and replaced with the *lacZ* reporter gene (Hombach et al., 2004). Cx57-deficient mice ($Cx57^{lacZ/lacZ}$) and wild-type (littermate) controls aged 2–4 months were used for all experiments. After each behavioral test or recording, the genotype of the retina was confirmed with staining for β-galactosidase activity and PCR as described (Hombach et al., 2004). All experiments were conducted in accordance with the institutional guidelines for animal welfare.

THE DEGREE OF HORIZONTAL CELL COUPLING AND LIGHT INTENSITY

Light intensities (photopic and scotopic) were chosen to span the range where changes in horizontal cell coupling are at, or are close to, their largest. Xin and Bloomfield (1999) showed that coupling reaches its maximum between 1 and 1.5 log units above rod threshold and its minimum at or above rod saturation (estimated at 3 log units above rod threshold). For the behavior experiments, scotopic intensity was 1.4×10^{-4} cd/m^2, which is between 0.9 and 2.1 log units above rod threshold, with mouse rod threshold estimated at 1×10^{-6} to 1.8×10^{-5} cd/m^2 (Umino et al., 2008; G.T. Prusky, Personal communication). Photopic intensity, 142 cd/m^2, was more than 3 log units above rod saturation (Xin and Bloomfield, 1999). The light source was Dell, 2007FPb, Phoenix, AZ, USA; neutral density filters were used to attenuate the monitor's output to the desired photopic and scotopic levels.

For the electrophysiology experiments, which were carried out with a different light source (Sony, Multiscan CPD-15SX1, New York, NY, USA), the intensities were, for the scotopic, 4×10^{-4} cd/m^2, which is between 1.3 and 2.6 log units above rod threshold, and, for the photopic, 23 cd/m^2, which is still >3 log units above rod saturation. As above, neutral density filters were used to attenuate the monitor's output to the desired photopic and scotopic levels.

THE RELATION OF HORIZONTAL CELL INPUT RESISTANCE TO COUPLING FOR SCOTOPIC VERSUS PHOTOPIC CONDITIONS AND FOR WILD-TYPE VERSUS KNOCKOUT ANIMALS

As mentioned in the Discussion, the standard experimental measure of horizontal cell coupling is the length constant (Xin and Bloomfield, 1999; Shelley et al., 2006). Xin and Bloomfield measured length constants physiologically in the rabbit (via the dependence of the voltage response on distance from a light stimulus) under different scotopic and photopic conditions and found the maximal scotopic-to-photopic ratio to be ~3. (As indicated in the previous section, the conditions used in this paper were matched to those that produce the maximal ratio.) Given this length constant ratio and the relations below, we can find the quantity we need, the input resistance ratio due to gap junction coupling. As given in Xin and Bloomfield (1999),

$$\lambda = \sqrt{\frac{R_m}{R_s}}, \tag{1}$$

where λ is the length constant, R_m is the membrane resistance, and R_s is the junctional resistance (also referred to as the sheet resistance). Rearranging in terms of R_s gives

$$R_s = \frac{R_m}{\lambda^2}. \tag{2}$$

For a 2-D cable and a point source, the input resistance, Z, is proportional to R_s. This follows from Eq. 2 of Lamb (1976) (see Appendix 2 Eqs 14–19 for details). Thus, it follows from Eq. 2 that

$$Z \propto \frac{R_m}{\lambda^2}. \tag{3}$$

This indicates that a 3-fold greater value of λ, as was measured by Xin and Bloomfield, corresponds to a 9-fold smaller value of Z, assuming that R_m remains the same in the scotopic and photopic conditions. Bloomfield notes that R_m may actually be higher in the photopic, indicating that a factor of 9 may be an underestimate.

The same analysis can be used to determine the input resistance ratio for the knockout and wild-type mouse using the measurements of Shelley et al. (2006), which were taken in these animals. These measurements, however, were taken only at one light level, and thus can provide only a lower bound on the ratio. Shelley et al. report a 2.3-fold greater value for λ in wild-type as compared to knockout, which, following Eq. 3, corresponds to a $2.3^2 = 5.29$-fold lower value for Z. It should be noted that R_m, as measured in isolated horizontal cells, is 27% lower in the knockout than the wild-type. When this is taken into account in Eq. 3, the wild-type-to-knockout ratio for Z is $(1 - 0.27)/(1/2.3^2) = 3.86$. We emphasize again that this is a lower bound on the input resistance ratio, since, as mentioned above, Shelley et al. measured length constants in knockout and wild-type only at a single light level.

Note that the 27% decrease in R_m has an additional implication: the observed change in temporal tuning that results from the change in coupling constitutes a lower bound, as the decrease in R_m would have the effect of reducing the difference between knockout and wild-type.

BEHAVIORAL TESTING USING A VIRTUAL OPTOKINETIC SYSTEM

Behavioral responses were measured using the Prusky/Douglas virtual optokinetic system (Prusky et al., 2004; Douglas et al., 2005). Briefly, the freely moving animal was placed in a virtual reality chamber. A video camera, situated above the animal, provided live video feedback of the testing arena. A pattern was projected onto the walls of the chamber in a manner that produced a drifting sine wave grating of fixed spatial frequency when viewed from the animal's position (0.128 cycles/degree, following the stimulus protocol of Umino et al., 2008). A drifting grating of a pre-selected temporal frequency at 100% contrast appeared, and the mouse was assessed for tracking behavior, as in Prusky et al. (2004). Grating contrast was systematically reduced until no tracking response was observed. The reciprocal of this threshold contrast was taken as the contrast sensitivity.

STIMULATING AND RECORDING GANGLION CELL RESPONSES

Three stimuli were used: drifting sine wave gratings, a binary random checkerboard (white noise), and a spatially uniform stimulus with natural temporal statistics (natural scene). The sine wave gratings were presented at eight temporal frequencies, ranging from 0.15 to 6 Hz, all with a spatial frequency of 0.039 cycles/degree. This spatial frequency was lower than the one used in the behavioral experiments, to ensure robust responses at the scotopic intensity. Each temporal frequency was presented for 2 min. The white noise stimulus was a random checkerboard at a contrast of 1, in which the intensity of each square (9×9 in mouse) was either white or black, randomly chosen every 0.067 s (large checkers were chosen to ensure stimulation of the large ganglion cell receptive fields at scotopic intensities, as indicated in Dedek et al., 2008). The natural scene stimulus was a spatially uniform movie whose intensities were taken from a time series of natural intensities (van Hateren, 1997), resampled for presentation at a 0.100-s frame period. This movie was 2 min long and presented 10 times, interleaved with a 2-s gray (mean intensity) screen. Measurements always started at the scotopic intensity. After all three stimuli were presented, the light intensity was increased. After 20 min of adaptation to the photopic intensity, the stimuli were presented as above.

Extracellular recordings made from central retina using a multielectrode array, as described previously (Nirenberg et al., 2001; Sinclair et al., 2004). Retina pieces were approximately 1.5–2 mm across, which corresponds to 4.5–6 horizontal cell length constants under scotopic conditions and 15–20 under photopic (as indicated above, there is an estimated factor of 3 difference in length constant between the scotopic and photopic conditions used here, with the photopic condition taken from Shelley et al. (2006) **Figure 7B**, which gives a wild-type light-adapted length constant). Spike

trains were recorded and sorted into units (cells) using a Plexon Instruments Multichannel Neuronal Acquisition Processor (Dallas, TX, USA), as described previously (Nirenberg et al., 2001).

Only ON ganglion cells were used, since the optomotor response in rodents is driven exclusively by the ON pathway (Dann and Buhl, 1987; Giolli et al., 2005). With respect to cell selection, only cells with readily detectable (by eye) spike triggered averages (STAs) were included in the data set; this corresponds to cells whose STA in the center checker of the receptive field was approximately 1.5 times above background.

DATA ANALYSIS

Temporal tuning curves were created from ganglion cell responses to drifting sine wave gratings using standard methods (Enroth-Cugell and Robson, 1966; Purpura et al., 1990; Croner and Kaplan, 1995). Briefly, for each grating, the first harmonic of the cell's response, $R(\mathbf{f})$, was calculated as follows:

$$R(\mathrm{f}) = \left| \frac{1}{L} \sum_j \exp\left[-i 2\pi \mathrm{f} t_j \right] \right|, \quad (4)$$

where f is the temporal frequency of the drifting sine wave grating (cycles/s), L is the duration of the stimulus (s), which was always an integer multiple of 1/f, and t_j is the time of the jth spike of the cell's response to the given grating. For averaging across cells, responses were weighted by the reciprocal of the peak sensitivity, so that each cell's tuning curve contributed approximately equally to the average, independent of its absolute sensitivity.

Mutual information was estimated between the input and responses (for the white noise, the input was the stimulus intensity of the checkerboard square that produced the largest response for a given cell; for the natural scene, the input was the full-field intensity). Information was estimated at each frequency using the coherence rate, following van Hateren and Snippe (2001):

$$I(\mathrm{S},R) = -\log_2\left(1 - \gamma^2(\mathrm{f})\right), \quad (5)$$

where $\gamma(\mathrm{f})$ is the coherence between stimulus and response at temporal frequency f. Coherence was estimated using the multi-taper method [Chronux library for Matlab (Mitra and Bokil, 2007), available at http://chronux.org], using effective bandwidths of 0.27 Hz (white noise) and 0.33 Hz (natural scene). For averaging across cells, information curves were weighted by the reciprocal of their areas, so that each cell's information curve contributed approximately equally to the average. Note that the above estimation of information is only rigorously correct for a Gaussian linear channel, and is necessarily an underestimate of the true information. However, our focus is not on the amount of information *per se*, but on its frequency-dependence.

FILTERED PREDATOR MOVIES

The "predator" movie, taken with a handheld digital camera (Casio, Exilim EX-Z750, Dover, NJ, USA), was filmed at 33 frames/s. The complete movie was filtered for each genotype, according to the behavioral data in **Figure3D** left: 0.1–6 Hz for the wild-type photopic, the same for knockout photopic, 0.16–3.2 Hz for wild-type scotopic, and 0.38–3.13 Hz for knockout scotopic. Representative frames from each filtered version are shown in **Figure 4**; the complete filtered versions are shown in Supplementary Material.

APPENDIX 1:

THE FREQUENCY RESPONSE DIFFERENCE BETWEEN THE RODS AND CONES LIES IN THE HIGH FREQUENCIES, NOT THE LOW

The shift to low temporal frequencies cannot be accounted for by the shift from cones to rods, as the cone-to-rod shift affects the high frequencies, not the low; see cone and rod impulse responses in Luo and Yau (2005), Nikonov et al. (2006). Here we show this explicitly in the model system we are using, the mouse. **Figure 5A** shows the impulse responses of the two photoreceptors, and **Figure 5B** shows the frequency responses, the latter generated by the Fourier transformation of the impulse responses. As shown in the figure, the frequency response difference lies in the high frequencies.

APPENDIX 2:

FORMAL TREATMENT OF THE MODEL IN FIGURE 2: THE EFFECT OF GAP JUNCTION COUPLING ON HORIZONTAL CELL FEEDBACK TO THE PHOTORECEPTOR

Section A formalizes the model of the photoreceptor–horizontal cell circuit to show how changing the strength of the horizontal cell feedback shapes the photoreceptor's temporal tuning, and, ultimately, the ganglion cell's temporal tuning. Section B then shows how a change in gap junction coupling modulates the strength of the horizontal cell feedback. Section C describes how these considerations apply to spatial configurations of the stimulus, and Section D briefly discusses how these considerations apply to other network geometries.

Section A

We start by briefly reiterating the model shown in **Figure 2**. As mentioned in the main text, it builds on the well-known negative feedback between the horizontal cell and the photoreceptor, whereby the horizontal cell sends a signal to the photoreceptor that shortens the latter's integration time (Baylor et al., 1971; Kleinschmidt and Dowling, 1975; see also Smith, 1995).

To understand how the photoreceptor is able to shift its integration time from short to long as the retina is shifted from a light-adapted to a dark-adapted state, we proposed the following: In the

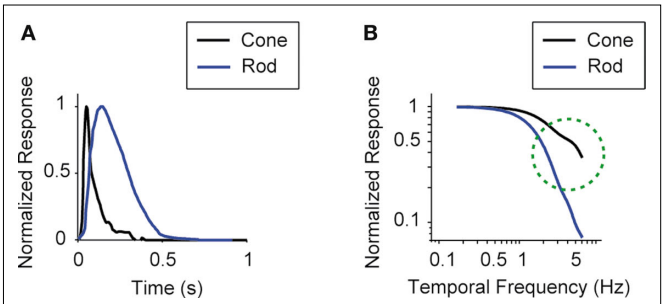


FIGURE 5 | The frequency response difference between the rods and cones lies in the high frequencies, not the low. (A) Impulse responses of the two photoreceptors, reproduced from Nikonov et al. (2006) for cone and Luo and Yau (2005) for rod. **(B)** Frequency responses of the two photoreceptors, generated by Fourier transforming the impulse responses.

light-adapted condition, the gap junctions of the horizontal cells close. This makes the horizontal cell feedback signal strong and keeps the photoreceptor integration time short. In the dark, the gap junctions open. This causes a shunting of horizontal cell current, which reduces horizontal cell feedback and shifts the photoreceptors to long integration times.

The proposal is based on three established facts – that the integration time of photoreceptors increases as the retina moves from light-adapted to dark-adapted conditions (Kleinschmidt and Dowling, 1975; Daly and Normann, 1985; Schnapf et al., 1990), that the strength of the horizontal cell feedback signal decreases as the retina moves from the light-adapted to the dark-adapted condition (Teranishi et al., 1983; Yang and Wu, 1989a) and that the degree of horizontal cell coupling increases as the retina moves from the light adapted to the dark-adapted condition (Dong and McReynolds, 1991; Xin and Bloomfield, 1999; Weiler et al., 2000). Taken together, these facts led to a proposal for how the circuit shifts its behavior. The novelty was the view of gap junction coupling as a shunting device, that is, a mechanism that can turn up or down the activity of a cell class, in this case, the horizontal cells. With this view, the three facts can account for the shift from one state to another.

In the main text, we proposed this schematically. Here we formalize it and use the formalized model to determine the feedback strength required to produce the observed state change.

We start with the well-known data of Schneeweis and Schnapf (2000). The data are measurements of photoreceptor responses across a range of light-adaptation levels and show the shift in integration time that occurs as the retina moves from the dark-adapted state to states of increasing levels of light-adaptation. We use the model to determine the change in feedback strength needed to produce the changes in photoreceptor integration time in Schneeweis and Schnapf (2000) and, ultimately, to produce the changes in ganglion cell integration time shown in this paper. (In Section B we show that the changes in feedback strength can be accounted for by the differences in horizontal cell coupling that occur in the dark- and light-adapted states.)

With these goals in mind, we use a linear systems approach. We do this for simplicity and generality, and because it allows us to focus on the essential features that lead to the shifts.

To construct the linear model, we denote the transfer function between light and the photoreceptor response in the absence of the feedback by $\tilde{P}(\omega)$, the feedback transfer function (photoreceptor output to horizontal cell, and back to photoreceptor) by $\tilde{F}(\omega)$, and the strength of the feedback by g. With this setup, the photoreceptor's output, $\tilde{L}(\omega,g)$, is given by the standard feedback formula (Oppenheim et al., 1997)

$$\tilde{L}(\omega,g)=\frac{\tilde{P}(\omega)}{1+g\tilde{F}(\omega)}. \tag{6}$$

To assign physiological values to the quantities in Eq. 6, we use, as mentioned above, the measurements of Schneeweis and Schnapf (2000), who present photoreceptor responses in the dark-adapted state (i.e., the no-feedback or essentially-no-feedback state, $g = 0$) through several light-adapted states (i.e., various levels of feedback up to $g = 1$) (**Figure 6**).

We determine the photoreceptor transformation P directly from Schneeweis and Schnapf's dark-adapted data, since when $g = 0$, $P = L$ (see Eq. 6). Specifically, we use their fit for $P(t)$, which is a phenomenological fit, given by:

$$P(t)=\big(1-w(t)\big)At^{n}e^{-t/\tau_1}+w(t)Be^{-t/\tau_2}\text{, where}$$

$$w(t)=\frac{1}{1+\left(\tau_3/t\right)^{m}}, \tag{7}$$

and $A = 3999, B = 1.68, \tau_1 = 0.063$ s, $\tau_2 = 0.646$ s, $\tau_3 = 0.200$ s, $n = 3$, and $m = 4$. The corresponding transfer function $\tilde{P}(\omega)$ is then determined from the impulse response $P(t)$ by Fourier transformation. Both $P(t)$ and $\tilde{P}(\omega)$ are shown in **Figure 7A**.

We then determine the feedback transformation F from the light-adapted measurements of Schneeweis and Schnapf. Since F was not measured directly, we proceed as follows. As mentioned above, F is the net result of two synapses in series: photoreceptor to horizontal cell, and horizontal cell back to photoreceptor. For simplicity, we use the same impulse response $f(t)$ for each synapse, and we use a difference of exponentials, a standard synaptic impulse response (Destexhe et al., 1995) for its functional form:

$$f(t)=e^{-(t-\delta)/\tau_a}-e^{-(t-\delta)/\tau_b}\text{ for }t\geq\delta. \tag{8}$$

Since the two synapses act in series, the feedback transfer function $\tilde{F}(\omega)$ is proportional to the product of the transfer functions at each synapse. We also include an overall scale factor F_0 in $\tilde{F}(\omega)$, so that we can pin the modeled response at $g = 1$ to the measured response at the highest level of light adaptation. Since we use the same transfer function $\tilde{f}(\omega)$ for the two synaptic components of F, the transfer function of the feedback transformation is given by

$$\tilde{F}(\omega)=F_0\tilde{f}(\omega)^2. \tag{9}$$

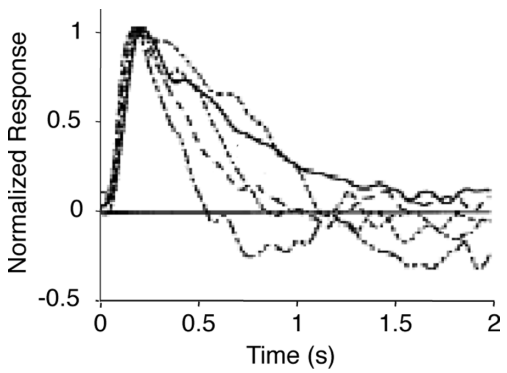


FIGURE 6 | Measured photoreceptor responses at increasing levels of light adaptation. Photoreceptor (macaque rod) responses under dark-adapted conditions (solid curve) and at increasing levels of light adaptation (dashed curves). The dark-adapted curve corresponds to the no-feedback or essentially-no-feedback condition; the light-adapted curves correspond to increasing levels of feedback. Adapted from Schneeweis and Schnapf (2000) Noise and light adaptation in rods of the macaque monkey. Visual Neuroscience 17, pp. 659–666, with permission of the publisher, Cambridge University Press. Curves are peak-normalized and inverted so that light responses are plotted up.

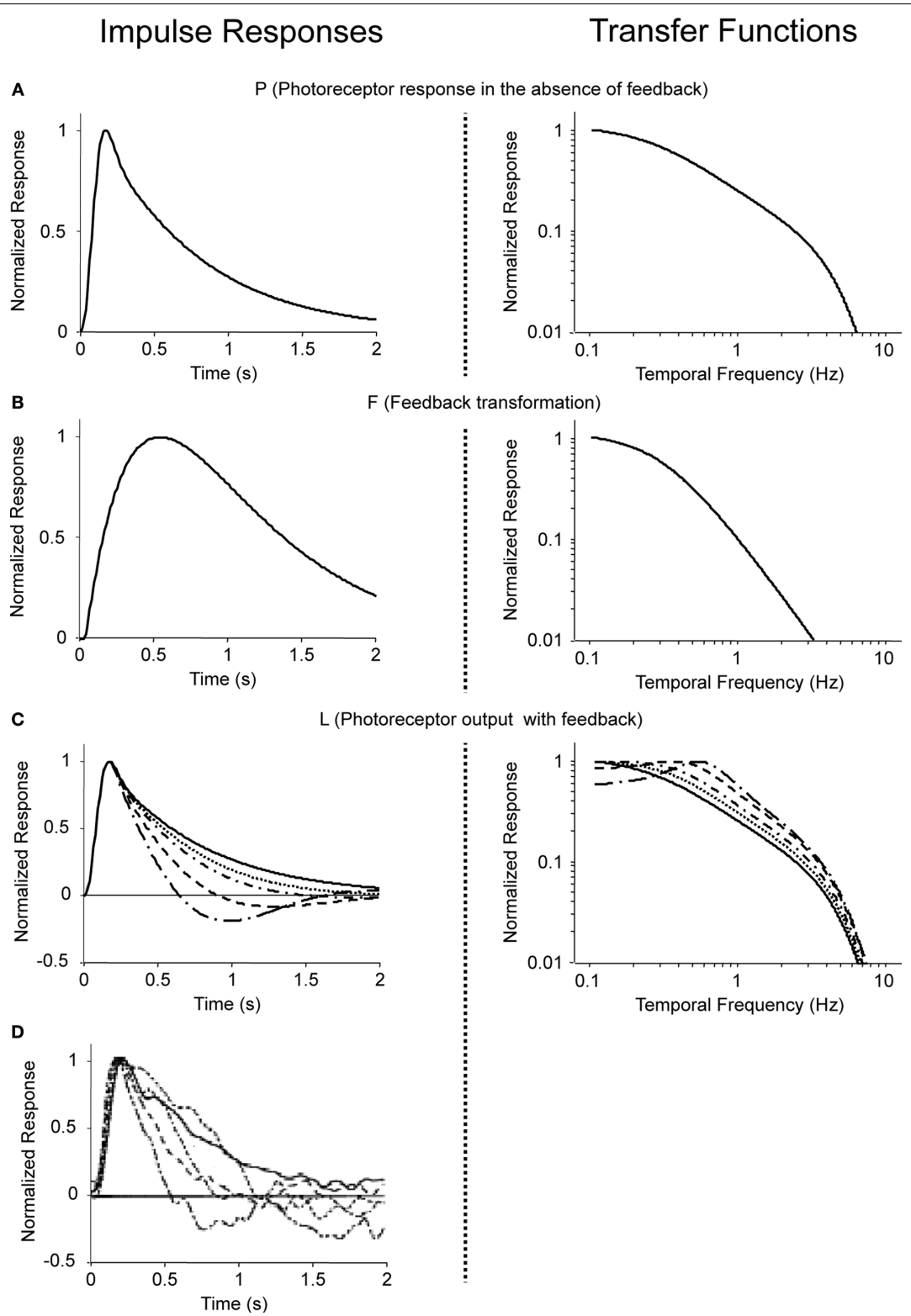


FIGURE 7 | Modeled photoreceptor responses at increasing levels of light adaptation. Impulse responses (left) and transfer functions (right) for the components of a simple feedback model of photoreceptor responses at increasing levels of light adaptation. **(A)** *P*, the response of the photoreceptor in the absence of feedback, corresponding to the dark-adapted state. **(B)** The feedback transformation *F*. **(C)** The resulting photoreceptor output, *L* (Eq. 6). The $g = 0$-curve (solid) is the same as **(A)**; the dashed curves correspond to $g = 0.1$, 0.2, 0.5, and 1. **(D)** The photoreceptor responses reported by Schneeweis and Schnapf (2000), as in **Figure 6**. All curves are shown peak-normalized; transfer functions are plotted as a function of frequency, $f = \omega/2\pi$.

The parameters ($\tau_a = 0.5$ s, $\tau_b = 0.01$ s, $\delta = 0.01$ s, and $F_0 = 10$) are chosen so that for a maximal feedback strength of $g = 1$, the photoreceptor output L given by Eq. 6 matches the most light-adapted response obtained by Schneeweis and Schnapf. The feedback impulse response $F(t)$ is the inverse Fourier transform of $\tilde{F}(\omega)$; both are shown in **Figure 7B**. As seen in **Figure 7C**, without changing this feedback transformation – just changing its strength g – the feedback model accounts for Schneeweis and Schnapf's responses at intermediate light levels.

To summarize, then, the modeled photoreceptor responses (**Figure 7C**) closely match the observed photoreceptor responses of Schneeweis and Schnapf (**Figure 6**) (also reproduced in **Figure 7D** for the reader's convenience). This enables us to obtain an estimate of the horizontal cell feedback strength needed to produce the range of changes in photoreceptor tuning. As shown in the figure, an approximate 10-fold change is needed: since $g = 0$ and 0.1 give nearly identical responses, we take $g = 0.1$ as the lower end of the range.

We now relate the photoreceptor output to the ganglion cell output. Specifically, we take into account the transformations that occur in the second processing layer of the retina (the inner plexiform layer). While these transformations have many details (Werblin and Dowling, 1969; Victor, 1987; Sakai and Naka, 1988), the common denominator is that signals become more transient, i.e., high-pass filtering occurs. We represent this with a standard RC filter in feedback configuration,

$$\tilde{X}(\omega) = \frac{1 + i\omega\tau_I}{1 + k_I + i\omega\tau_I}, \tag{10}$$

choosing the parameter values ($k_I = 4$ and $\tau_I = 6$ s) to match the dark-adapted ganglion cell response, as in **Figure 3C** (wild-type). Thus, the ganglion cell response is determined by the output of the photoreceptor–horizontal cell feedback circuit (Eq. 6), followed by the schematic inner plexiform layer filter (Eq. 10):

$$\tilde{R}(\omega, g) = \tilde{X}(\omega)\tilde{L}(\omega, g) = \tilde{X}(\omega)\frac{\tilde{P}(\omega)}{1 + g\tilde{F}(\omega)}. \tag{11}$$

Figure 8 shows the results. **Figure 8A** recapitulates the photoreceptor output from **Figure 7C**, and **Figure 8B** shows the corresponding ganglion cell output after applying Eq. 11. (**Figure 8C** shows the same result on a semilog plot, to be consistent with the main text.) As shown in **Figure 8C**, as horizontal cell feedback strength decreases, the temporal tuning of the ganglion cell response shifts to lower frequencies. The shift in the peak frequency is approximately 3-fold, from 0.6 to 0.2 Hz, and can be accounted for by a factor of 10 reduction in horizontal cell feedback strength. Since the shift we observe in **Figure 3C** is a subset of this, a 10-fold change in feedback strength more than suffices to account for the shift in tuning we observe at the ganglion cell output.

To summarize: Using the data of Schneeweis and Schnapf (2000) as the starting point, we showed that, as horizontal cell feedback strength increases, the tuning of the photoreceptor and, ultimately, the ganglion cell, shifts to higher frequencies. As shown in **Figure 8C**, the peak frequency shift is approximately 3-fold and can be accounted for by a 10-fold change in horizontal cell feedback strength. Since the shift we present in the main text (**Figure 3C**) is a subset of this, a 10-fold change in feedback strength is more than sufficient to account for it.

In the next section, we show how the measured changes in gap junction coupling are sufficient to produce the changes in feedback strength (an expansion of the analysis presented in Materials and Methods).

We conclude the section by mentioning that the analysis done here focused on rod conditions, that is, rod responses were shown with various levels of horizontal cell feedback. We focused on rod conditions, since these are directly compared in the main figure of the paper, **Figure 3C**. Specifically, **Figure 3C** compares the rod condition in the high feedback state (the state in the knockout in the dark, where horizontal cells are forced to remain uncoupled) with the low feedback state (the state in the wild-type in the dark, where horizontal cells are maximally coupled).

Section B

In this section we detail the relationship between changes in gap junction coupling and horizontal cell feedback strength, an expansion of the description in the Materials and Methods section "The

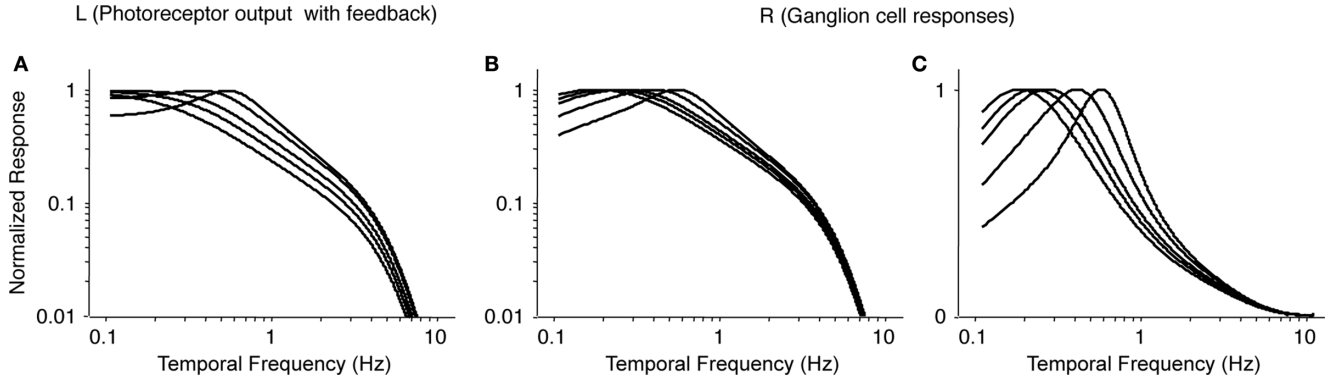


FIGURE 8 | Modeled ganglion cell responses at increasing levels of light adaptation. (A) Transfer functions for photoreceptor output (L, Eq. 6) at increasing levels of light adaptation (i.e., increasing levels of horizontal cell feedback), taken from **Figure 7C**. **(B)** Transfer functions for ganglion cell responses (R, Eq. 11), obtained by high-pass filtering the curves in **(A)**. **(C)** Same as **(B)**, but plotted on semilog coordinates. These curves are reproduced in **Figure 2** of the main text. From left to right within each panel, values of feedback strength are: $g = 0$, 0.1, 0.2, 0.5, and 1.0.

Relation of Horizontal Cell Input Resistance to Coupling for Scotopic Versus Photopic Conditions and for Wild-type Versus Knockout Animals". We show that the measured changes in coupling are sufficient to produce a 10-fold change in feedback strength and thus are sufficient to account for our results and also for the larger range of shifts shown in **Figure 8C**.

As mentioned in "Materials and Methods," the standard measure of horizontal cell coupling is the length constant. The strength of the horizontal cell signal, on the other hand, is determined by the cell's input resistance, since the cell's voltage response is the input resistance multiplied by the input current (Ohm's law). Thus, to determine how much the horizontal signal changes, we need to determine how much of a change in input resistance is produced by a measured change in length constant.

This is readily accomplished with a well-known model of the horizontal cell network, the two-dimensional cable (Naka and Rushton, 1967; Lamb, 1976; Xin and Bloomfield, 1999; Packer and Dacey, 2005; Shelley et al., 2006). We use the two-dimensional cable model to link horizontal cell coupling and length constant, and then to link length constant and input resistance. As we will show, input resistance is inversely proportional to the square of the length constant (for a point source of current, but see also Section C). Xin and Bloomfield (1999) measured length constants under different degrees of coupling. Their results showed that length constant increases by a factor of 3 between the minimally- and maximally-coupled states. A 3-fold increase in length constant corresponds to a 9-fold decrease in feedback strength, nearly the 10-fold change needed to account for the complete range of shifts in **Figure 8C**.

The following details the link between horizontal cell coupling and length constant, and then the link between length constant and input resistance. We focus on the regime in which capacitative effects can be neglected, since the phenomena of interest occur below 2 Hz. At the end of Section D, we comment on how the analysis can be extended to include capacitative effects.

As mentioned above, we start by modeling the horizontal cells as a two-dimensional sheet, as is standard (Naka and Rushton, 1967; Lamb, 1976; Xin and Bloomfield, 1999; Packer and Dacey, 2005; Shelley et al., 2006). Within this sheet, horizontal cell coupling determines resistance to current flow, and we denote the sheet resistance by R_s. Thus, our immediate goal is to link R_s to length constant, denoted by λ.

This linkage is well-known, and is given by the classic work of Lamb (1976). As Lamb showed (his Eq. 2) the length constant of a two-dimensional sheet is given by

$$\lambda = \sqrt{R_m / R_s}, \tag{12}$$

corresponding to Eq. 1 in the main text. Rearranging this yields

$$R_s = R_m / \lambda^2, \tag{13}$$

corresponding to Eq. 2. Equation 13 demonstrates the relationship between length constant λ and horizontal cell coupling, as measured by the sheet resistance R_s.

The next step is to link input resistance to length constant. We start with a point source current, and consider other geometries in Sections C and D. For a point source current, we begin with Lamb (1976) (his Eq. 8), which provides the voltage response of the sheet. At a distance r from the injection of a current i_0, the resulting voltage $V(r)$ is

$$V(r) = i_0 \frac{R_s}{2\pi} K_0(r/\lambda), \tag{14}$$

where K_0 is a modified Bessel function of the second kind.

Input resistance is the ratio of the voltage response to the injected current. At a distance r from the point source, the ratio $Z_r = V(r)/i_0$ is

$$Z_r = \frac{V(r)}{i_0} = \frac{R_s}{2\pi} K_0(r/\lambda), \tag{15}$$

which follows from Eq. 14.

We would like to use Eq. 15 to determine Z_r at $r = 0$ (the point of injection), and how it depends on the horizontal cell parameters. Since the Bessel function in Eq. 15 diverges at the origin, Z_0 is formally undefined. However, real measurements correspond to values of r that are small but not 0. Therefore, instead of focusing on Z_0, we focus on the limiting behavior of Z_r when r is small[2].

To determine the behavior in the small-r limit, we approximate the Bessel function in Eq. 15, whose argument is $u = r/\lambda$. When this argument is small (i.e., when $r << \lambda$), the Bessel function has an asymptotic expansion, $K_0(u) = -(\ln u)\,[1 + o(u)]$ (Abramowitz and Stegun, 1965, Eq. 9.6.54). Therefore,

$$Z_r = -\frac{R_s}{2\pi}\ln(r/\lambda)\big(1+o(r/\lambda)\big) = \frac{R_s}{2\pi}\big(\ln(\lambda)-\ln(r)\big)\big(1+o(r)\big). \tag{16}$$

In the small-r limit, the $-\ln(r)$-term grows, eventually dominating the $\ln(\lambda)$-term, Thus, Z_r has an asymptotic expansion

$$Z_r = -\frac{R_s}{2\pi}\ln r\big(1+o(r)\big). \tag{17}$$

Equation 17 shows that in the limit of a point current injection, input resistance and sheet resistance are proportional (corresponding to the comment following text Eq. 2). Finally, we use the relationship between sheet resistance and length constant (Eq. 13) to rewrite Eq. 17 as

$$Z_r = -\frac{R_m}{2\pi\lambda^2}\ln r\big(1+o(r)\big). \tag{18}$$

Thus, in the small-r limit, the input resistance is proportional to R_m and inversely proportional to λ^2, as in text Eq. 3:

$$Z_r \propto \frac{R_m}{\lambda^2}. \tag{19}$$

To summarize: horizontal cell coupling (sheet resistance) determines the length constant via Eq. 12, and these are linked to input resistance via Eqs 17 and 18.

[2]For an alternative derivation that relies only on a dimensional analysis, see Section D.

Section C

Above, we considered the input resistance for a point input source; we now turn to consider other spatial patterns. To do this systematically, we determine the input resistance for spatial grating pattern of spatial frequency k, which we denote $Z(k)$. That is, $Z(k)$ is the ratio of the voltage response to an applied grating-shaped current. We determine this voltage response by first determining the response to a current injected along a narrow line. Then we superimpose a continuum of line sources to form the grating.

In the scenario of a current injected along a narrow line (say, along the y-axis) into a sheet in the (x, y)-plane, there is translational symmetry along the y-axis. Along the x-axis, the problem reduces to that of a one-dimensional cable. (This is the geometry considered by Xin and Bloomfield, 1999). Thus, we can use standard one-dimensional cable theory to determine the resulting voltage distribution: at a distance x from a line of injected current I_0, the resulting voltage distribution is:

$$V_{\text{line}}(x) = I_0 Z_0 e^{-|x|/\lambda}, \tag{20}$$

where

$$Z_0 = \frac{1}{2}\sqrt{R_m R_s} \tag{21}$$

is the input resistance of the equivalent one-dimensional cable (Koch and Segev, 1998).

Next, we create a grating from these line sources. At each location x_0 along the x-axis, we place a source with strength $I(x_0;k) = I_0\cos(kx_0)$; the net result of these sources is a spatial grating of current. Each of these sources yields a voltage response according to Eq. 20, and they superimpose to yield the voltage response to the grating. Specifically, the contribution of the line source at position x_0 to the voltage at position x is $V_{\text{line}}(x - x_0)\cos(kx_0)$, and superimposing them yields the grating response:

$$V_{\text{grating}}(x;k) = \int_{-\infty}^{\infty} V_{\text{line}}(x - x_0)\cos(kx_0)dx_0. \tag{22}$$

Carrying out this Fourier integral yields

$$V_{\text{grating}}(x;k) = \cos(kx)\int_{-\infty}^{\infty} V_{\text{line}}(u)e^{iku}du = \cos(kx)\frac{I_0 Z_0}{\lambda}\frac{2}{1/\lambda^2 + k^2}. \tag{23}$$

Thus, $Z(k)$, the input resistance for a current injection patterned as a sinusoid of spatial frequency k, is the ratio of the voltage response to the applied current:

$$Z(k) = \frac{V_{\text{grating}}(0;k)}{I_0} = \frac{Z_0}{\lambda}\frac{2}{1/\lambda^2 + k^2} = R_m\frac{1}{1 + \lambda^2 k^2}, \tag{24}$$

where we have used Eqs 12 and 21 in the last step.

Equation 24 shows how length constant and spatial frequency interact to determine the input resistance. At sufficiently low spatial frequencies, the shunt current has nowhere to go, so the input resistance is R_m, independent of the length constant. At sufficiently high frequencies, the shunt is very effective: input resistance is inversely proportional to λ^2, just as in the point source. For example, at $k = 3/\lambda$, $Z(k) = R_m/10$, indicating that 90% of the input resistance can be shunted away, while at $k = 1/\lambda$, $Z(k) = R_m/2$, indicating that half of the input resistance can be shunted away. Since spatial frequency k is measured in radians, the latter corresponds to a spatial wavelength of $2\pi\lambda$. Thus, perhaps counterintuitively, Eq. 24 shows that the shunt retains effectiveness even for a grating pattern whose period is a fairly large multiple (2π) of the length constant.

To summarize: the reduction in input resistance due to gap junction coupling diminishes at low spatial frequencies, but the falloff is gentle, as shown in Eq. 24. For gratings whose period is small in comparison to $2\pi\lambda$, the shunt remains large. This was the case in the present experiments under scotopic conditions. We used gratings of 0.039 cycles/degree, corresponding to a spatial period of 795 μm [in the mouse retina, 1 = 31 μm (Remtulla and Hallett, 1985)], and a spatial frequency k of $2\pi/795 = 0.0079$ μm^{-1}. Given the estimated scotopic length constant of $\lambda = 300$ μm (see Stimulating and Recording Ganglion Cell Responses), Eq. 24 yields $Z(k) = 0.15R_m$, indicating that 85% of the signal can be shunted away.

We conclude by mentioning that while the interaction of spatial pattern and gap junction coupling is a potentially interesting topic, the paper focused on temporal processing and, thus, was not set up to explore this: this is because of a limitation in the size of the retinal pieces used for the multi-electrode array recording. To test the predictions in Eq. 24, retinal pieces of greater than twice the size would be needed to avoid edge effects (shunting through contact with the edge of the retinal piece) and to allow sampling of sufficiently low spatial frequencies. We included the above discussion of the theoretical effects of spatial pattern in any case, because it makes predictions for future work, both in retina and other brain areas where gap junction coupled networks are present.

Section D

Because the gap-junction switch has the potential to operate in a wide range of neural networks, here we briefly note how the above considerations generalize to geometries not directly related to the horizontal cell network of the retina.

First, we mention that the notion that gap junction conductance modulates input resistance is not limited to situations in which the gap-junction-coupled cells form part of a feedback loop. That is, opening the gap junctions of a group of neurons is simply a general way to reduce their gain and thus remove them functionally from a network, whatever their role.

For networks within the brain parenchyma, a three-dimensional space-filling network may be a more appropriate caricature than a two-dimensional syncytial sheet. (We have in mind a scenario in which each neuron is connected to its neighbors in all three spatial dimensions, but that only a part of the volume is occupied by these neurons.) In this case, the dependence of input resistance on gap junction coupling is $Z \propto R_s^{3/2}$, an even stronger dependence than the proportionality which holds in two-dimensional case, Eq. 17.

To see this, we apply a dimensional analysis. In three dimensions, the resistance R_m to the bath (i.e., extracellular space) has units of ohm-cm^3, and the internal resistance, R_s, has units of ohm-cm. Thus, the input resistance for a point source must be proportional to $\sqrt{R_s^3/R_m}$, since this is the only parameter combination that has units of ohms. The length constant λ is still $\sqrt{R_m/R_s}$, so the input resistance is also proportional to R_m/λ^3.

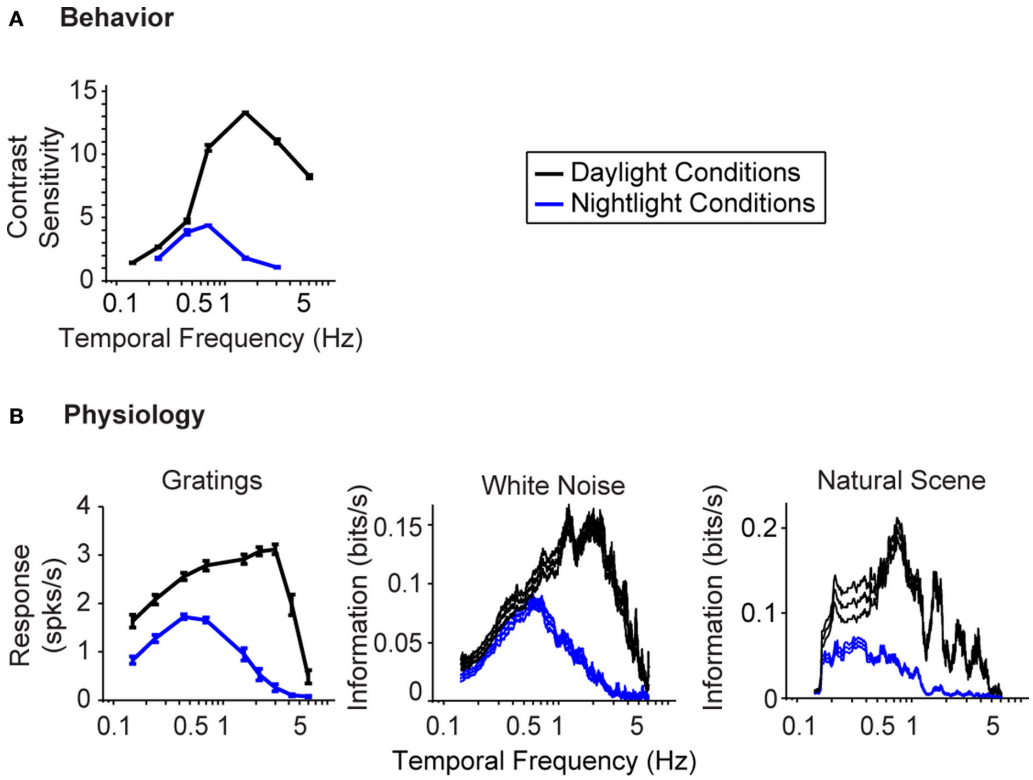


FIGURE 9 | The visual system undergoes a shift in integration time as it shifts from day to night (photopic to scotopic) conditions. This figure reproduces the data in **Figure 1** of the main text, but un-normalized. Note that in the un-normalized plots, the shift in tuning to low temporal frequencies is superimposed on an overall decrease in sensitivity, as is well-known at the behavioral level [Kelly, 1961 (human); Umino et al., 2008 (mouse)] and at the ganglion cell level (Purpura et al., 1990).

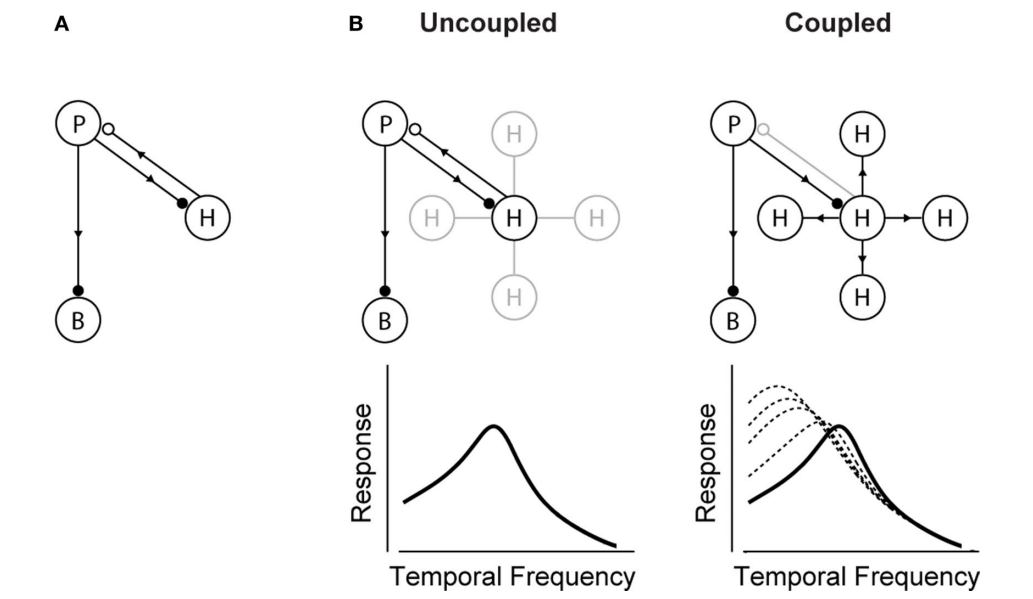


FIGURE 10 | The circuit that controls visual integration time can be shifted from one state to another by a change in the gap junction coupling of one of its cell classes. This figure reproduces the model shown in the main text, but with the response curves un-normalized (see **Figure 2** of the main text or **Figure 8C** of Appendix 2).

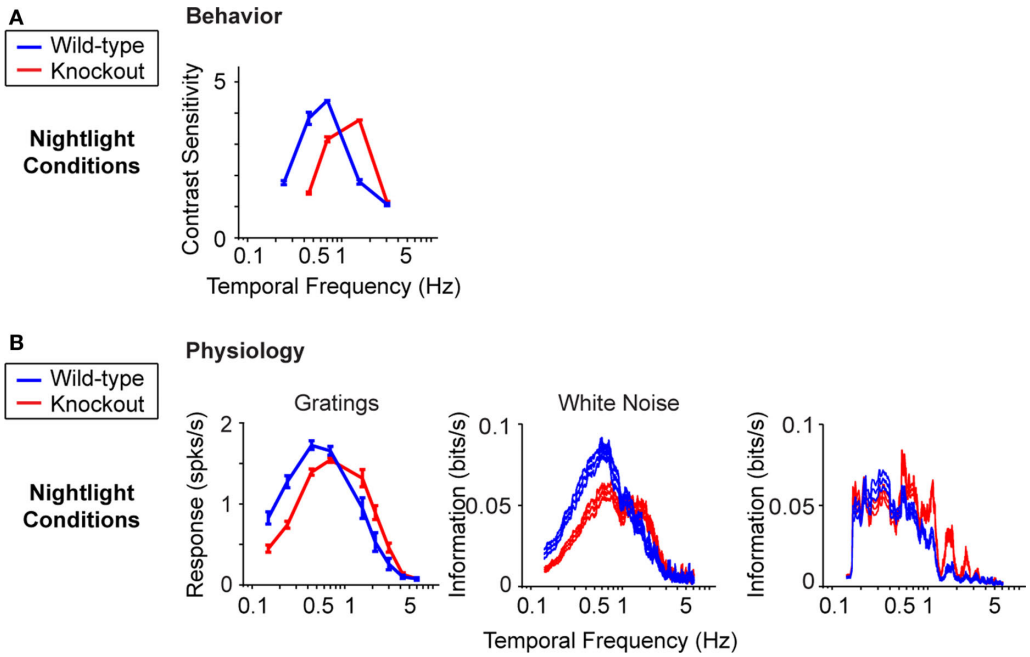


FIGURE 11 | When coupling is prevented, the shift to long integration times is impaired at both the behavioral level and the ganglion cell level. This figure reproduces the data in **Figure 3** of the main text, but un-normalized. As in the main text, the knockout response fails to make the normal shift in tuning to low temporal frequencies, because the feedback signal is not reduced by the shunt. Note that the un-normalized plots show that at low temporal frequencies, the wild-type response is higher, while at high temporal frequencies, the knockout response is higher. This is predicted by the model (**Figure 10**, which shows the un-normalized model predictions; lower right of figure). Note also that this crossover (the higher response in the no-feedback state at low frequencies, and the higher response in the high-feedback state at high frequencies) is a well-known phenomenon in light adaptation (Purpura et al., 1990).

There is a simple intuition behind this result and the corresponding ones results in the earlier sections: for a point source, the input resistance decreases in proportion to the number of neurons to which an input current spreads. In a "cable" of effective dimension D and length constant λ, this number is proportional to λ^D.

Finally, we mention that in all of the above analyses, we have considered the gap-junction-coupled network to be purely resistive. This is a reasonable approximation for the experiments considered here: the phenomena of interest occur below 2 Hz. These frequencies are much slower than the estimated RC time constant for the horizontal cell, which is 20 ms, based on membrane resistance and capacitance values provided by Smith (1995). Nevertheless, our treatment immediately generalizes to scenarios in which capacitive effects become relevant, by replacing the resistance parameters R_m, R_s, and Z by corresponding frequency-dependent impedances (Koch and Poggio, 1985). The cable formalism still applies, but now, the effective length constant will be frequency-dependent, and the shunt may be associated with a phase shift.

APPENDIX 3:

FIGURES UN-NORMALIZED

Figures 1, 2, and 3 are reproduced in un-normalized form as Figures 9, 10, and 11.

ACKNOWLEDGMENTS

We thank A. Molnar for helpful discussion, Y. Roudi, and K. Purpura for comments on the manuscript, and K. Willecke and colleagues for the use of the $Cx57^{lacZ/lacZ}$ mouse line. **Figure 3A** was adapted from a previous paper by our group, Dedek et al. (2008). This work was supported by funds from National Eye Institute R01 EY12978 to S. Nirenberg; C. Pandarinath was supported in part by T32-EY07138; J. Victor is supported in part by funds from National Eye Institute RO1 EY7977 and EY9314".

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/computationalneuroscience/paper/10.3389/fncom.2010.00002

FILTERED MOVIES

The selective disadvantage of a Cx57 gene loss, demonstrated using a natural movie

As indicated in the main text, we filmed an approaching predator and restricted the movies to the temporal frequencies available to each genotype, using the data from **Figure 3D** left. In **Figure 4** we showed single frames from the movies; here we show the movies in total. As indicated in the main text, in daytime conditions, the movies for the two genotypes are essentially the same – see Video 1, Wild-type by Day, and Video 2, Knockout by Day. In nighttime conditions, though, the two movies diverge. In the movie filtered through the frequencies visible to the wild-type animal, the predator is visible both when it is moving, i.e., when the movie is dominated by high frequencies, and when it is still, i.e., when the movie is dominated by low frequencies. In the movie filtered though the frequencies visible to the knockout, the predator disappears in the still condition. Only a ghost remains – see Video 3, Wild-type at Night, and Video 4, Knockout at Night.

REFERENCES

Abramowitz, M., and Stegun, I. A. (1965). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables.* New York, Courier Dover Publications.

Babai, N., and Thoreson, W. B. (2009). Horizontal cell feedback regulates calcium currents and intracellular calcium levels in rod photoreceptors of salamander and mouse retina. *J. Physiol.* 587, 2353–2364.

Baylor, D. A., Fuortes, M. G., and O'Bryan, P. M. (1971). Receptive fields of cones in the retina of the turtle. *J. Physiol.* 214, 265–294.

Bennett, M. V. L., and Zukin, R. S. (2004). Electrical coupling and neuronal synchronization in the Mammalian brain. *Neuron* 41, 495–511.

Bloomfield, S. A., Xin, D., and Osborne, T. (1997). Light-induced modulation of coupling between AII amacrine cells in the rabbit retina. *Vis. Neurosci.* 14, 565–576.

Bonin, V., Mante, V., and Carandini, M. (2006). The statistical computation underlying contrast gain control. *J. Neurosci.* 26, 6346–6353.

Cepeda, C., Walsh, J. P., Hull, C. D., Howard, S. G., Buchwald, N. A., and Levine, M. S. (1989). Dye-coupling in the neostriatum of the rat: I. Modulation by dopamine-depleting lesions. *Synapse* 4, 229–237.

Cook, P. B., and McReynolds, J. S. (1998). Lateral inhibition in the inner retina is important for spatial tuning of ganglion cells. *Nat. Neurosci.* 1, 714–719.

Croner, L. J., and Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Res.* 35, 7–24.

Daly, S. J., and Normann, R. A. (1985). Temporal information processing in cones: effects of light adaptation on temporal summation and modulation. *Vision Res.* 25, 1197–1206.

Dann, J. F., and Buhl, E. H. (1987). Retinal ganglion cells projecting to the accessory optic system in the rat. *J. Comp. Neurol.* 262, 141–158.

Dedek, K., Pandarinath, C., Alam, N. M., Wellershaus, K., Schubert, T., Willecke, K., Prusky, G. T., Weiler, R., and Nirenberg, S. (2008). Ganglion cell adaptability: does the coupling of horizontal cells play a role? *PLoS ONE* 3, e1714. doi:10.1371/journal.pone.0001714.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Destexhe, A., Mainen, Z. F., and Sejnowski, T. J. (1995). Synaptic currents, neuromodulation, and kinetic models. In *The Handbook of Brain Theory and Neural Networks,* M. A. Arbib, ed. (Cambridge, MA, MIT Press), pp. 956–959.

De Valois, R. L., and De Valois, K. K. (1990). *Spatial Vision.* New York, NY, Oxford University Press.

Dong, C. J., and McReynolds, J. S. (1991). The relationship between light, dopamine release and horizontal cell coupling in the mudpuppy retina. *J. Physiol.* 440, 291–309.

Douglas, R. M., Alam, N. M., Silver, B. D., McGill, T. J., Tschetter, W. W., and Prusky, G. T. (2005). Independent visual threshold measurements in the two eyes of freely moving rats and mice using a virtual-reality optokinetic system. *Vis. Neurosci.* 22, 677–684.

Dowling, J. E. (1987). *The Retina: An Approachable Part of the Brain.* Cambridge, MA, Belknap Press.

Enroth-Cugell, C., and Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol.* 187, 517–552.

Flores-Herr, N., Protti, D. A., and Wässle, H. (2001). Synaptic currents generating the inhibitory surround of ganglion cells in the mammalian retina. *J. Neurosci.* 21, 4852–4863.

Galarreta, M., and Hestrin, S. (1999). A network of fast-spiking cells in the neocortex connected by electrical synapses. *Nature* 402, 72–75.

Galarreta, M., and Hestrin, S. (2001). Electrical synapses between GABA-releasing interneurons. *Nat. Rev. Neurosci.* 2, 425–433.

Giolli, R. A., Blanks, R. H. I., and Lui, F. (2005). The accessory optic system: basic organization with an update on connectivity, neurochemistry, and function. *Prog. Brain Res.* 151, 407–440.

Hombach, S., Janssen-Bienhold, U., Sohl, G., Schubert, T., Bussow, H., Ott, T., Weiler, R., and Willecke, K. (2004). Functional expression of connexin57 in horizontal cells of the mouse retina. *Eur. J. Neurosci.* 19, 2633–2640.

Janssen-Bienhold, U., Trumpler, J., Hilgen, G., Schultz, K., De Sevilla Muller, L., Sonntag, S., Dedek, K., Dirks, P., Willecke, K., and Weiler, R. (2009). Connexin57 is expressed in dendro-dendritic and axo-axonal gap junctions of mouse horizontal cells and its distribution is modulated by light. *J. Comp. Neurol.* 513, 363–374.

Kamermans, M., Fahrenfort, I., Schultz, K., Janssen-Bienhold, U., Sjoerdsma, T., and Weiler, R. (2001). Hemichannel-mediated inhibition in the outer retina. *Science* 292, 1178–1180.

Kelly, D. H. (1961). Visual response to time-dependent stimuli. I. Amplitude sensitivity measurements. *J. Opt. Soc. Am.* 51, 422–429.

Kleinschmidt, J., and Dowling, J. E. (1975). Intracellular recordings from gecko photoreceptors during light and dark adaptation. *J. Gen. Physiol.* 66, 617–648.

Koch, C., and Poggio, T. (1985). A simple algorithm for solving the cable equation in dendritic trees of arbitrary geometry. *J. Neurosci. Methods* 12, 303–315.

Koch, C., and Segev, I. (1998). *Methods in Neuronal Modeling: From Ions to Networks.* Cambridge, MA, MIT Press.

Lamb, T. D. (1976). Spatial properties of horizontal cell responses in the turtle retina. *J. Physiol.* 263, 239–255.

Li, H., Chuang, A. Z., and O'Brien, J. (2009). Photoreceptor coupling is controlled by connexin 35 phosphorylation in zebrafish retina. *J. Neurosci.* 29, 15178–15186.

Luo, D.-G., and Yau, K.-W. (2005). Rod sensitivity of neonatal mouse and rat. *J. Gen. Physiol.* 126, 263–269.

Maunsell, J. H. R., and Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci.* 29, 317–322.

McMahon, D. G., Knapp, A. G., and Dowling, J. E. (1989). Horizontal cell gap junctions: single-channel conductance and modulation by dopamine. *Proc. Natl. Acad. Sci. U.S.A.* 86, 7639–7643.

McMahon, D. G., and Mattson, M. P. (1996). Horizontal cell electrical coupling in the giant danio: synaptic modulation by dopamine and synaptic maintenance by calcium. *Brain Res.* 718, 89–96.

McMahon, M. J., Packer, O. S., and Dacey, D. M. (2004). The classical receptive field surround of primate parasol ganglion cells is mediated primarily by a non-GABAergic pathway. *J. Neurosci.* 24, 3736–3745.

Mitra, P., and Bokil, H. (2007). *Observed Brain Dynamics.* New York, NY, Oxford University Press.

Naka, K., and Rushton, W. (1967). The generation and spread of S-potentials in fish (Cyprinidae). *J. Physiol.* 192, 437–461.

Nakatani, K., Tamura, T., and Yau, K. W. (1991). Light adaptation in retinal rods of the rabbit and two other nonprimate mammals. *J. Gen. Physiol.* 97, 413–435.

Nikonov, S. S., Kholodenko, R., Lem, J., and Pugh, E. N. J. (2006). Physiological features of the S- and M-cone photoreceptors of wild-type mice from single-cell recordings. *J. Gen. Physiol.* 127, 359–374.

Nirenberg, S., Carcieri, S. M., Jacobs, A. L., and Latham, P. E. (2001). Retinal ganglion cells act largely as independent encoders. *Nature* 411, 698–701.

Ohzawa, I., Sclar, G., and Freeman, R. D. (1982). Contrast gain control in the cat visual cortex. *Nature* 298, 266–268.

Onn, S.-P., Lin, M., Liu, J.-J., and Grace, A. A. (2008). Dopamine and cyclic-AMP regulated phosphoprotein-32-dependent modulation of prefrontal cortical input and intercellular coupling in mouse accumbens spiny and aspiny neurons. *Neuroscience* 151, 802–816.

Oppenheim, A., Willsky, A., and Nawab, S. (1997). *Signals and Systems.* Englewood Cliffs, Prentice Hall.

Packer, O. S., and Dacey, D. M. (2005). Synergistic center-surround receptive field model of monkey H1 horizontal cells. *J. Vision* 5, 1038–1054.

Peichl, L., and González-Soriano, J. (1994). Morphological types of horizontal cell in rodent retinae: a comparison of rat, mouse, gerbil, and guinea pig. *Vis. Neurosci.* 11, 501–517.

Peskin, C. S., Tranchina, D., and Hull, D. M. (1984). How to see in the dark: photon noise in vision and nuclear medicine. *Ann. N. Y. Acad. Sci.* 435, 48–72.

Prusky, G. T., Alam, N. M., Beekman, S., and Douglas, R. M. (2004). Rapid quantification of adult and developing mouse spatial vision using a virtual optomotor system. *Invest. Ophthalmol. Vis. Sci.* 45, 4611–4616.

Purpura, K., Tranchina, D., Kaplan, E., and Shapley, R. M. (1990). Light adaptation in the primate retina: analysis of changes in gain and dynamics of monkey retinal ganglion cells. *Vis. Neurosci.* 4, 75–93.

Remtulla, S., and Hallett, P. E. (1985). A schematic eye for the mouse, and comparisons with the rat. *Vis. Res.* 25, 21–31.

Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185.

Ribelayga, C., Cao, Y., and Mangel, S. C. (2008). The circadian clock in the retina controls rod–cone coupling. *Neuron* 59, 790–801.

Roska, B., Nemeth, E., Orzo, L., and Werblin, F. S. (2000). Three levels of lateral inhibition: a space-time study of the retina of the tiger salamander. *J. Neurosci.* 20, 1941–1951.

Sakai, H., and Naka, K. (1988). Neuron network in catfish retina: 1968–1987. *Prog. Retin. Res.* 7, 149–208.

Schnapf, J. L., Nunn, B. J., Meister, M., and Baylor, D. A. (1990). Visual transduction in cones of the monkey Macaca fascicularis. *J. Physiol.* 427, 681–713.

Schneeweis, D. M., and Schnapf, J. L. (2000). Noise and light adaptation in rods of the macaque monkey. *Vis. Neurosci.* 17, 659–666.

Shapley, R. M., and Victor, J. D. (1978). The effect of contrast on the transfer properties of cat retinal ganglion cells. *J. Physiol.* 285, 275–298.

Shelley, J., Dedek, K., Schubert, T., Feigenspan, A., Schultz, K., Hombach, S., Willecke, K., and Weiler, R. (2006). Horizontal cell receptive fields are reduced in connexin57-deficient mice. *Eur. J. Neurosci.* 23, 3176–3186.

Shields, C. R., Klooster, J., Claassen, Y., Ul-Hussain, M., Zoidl, G., Dermietzel, R., and Kamermans, M. (2007). Retinal horizontal cell-specific promoter activity and protein expression of zebrafish connexin 52.6 and connexin 55.5. *J. Comp. Neurol.* 501, 765–779.

Sinclair, J. R., Jacobs, A. L., and Nirenberg, S. (2004). Selective ablation of a class of amacrine cells alters spatial processing in the retina. *J. Neurosci.* 24, 1459–1467.

Smith, R. G. (1995). Simulation of an anatomically defined local circuit: the cone-horizontal cell network in cat retina. *Vis. Neurosci.* 12, 545–561.

Standage, D., and Paré, M. (2009). Flexible control of speeded and accurate decisions afforded by temporal gain modulation of decisional processes. In 39th Annual Meeting of Society for Neuroscience, Chicago, IL.

Tamura, T., Nakatani, K., and Yau, K. W. (1989). Light adaptation in cat retinal rods. *Science* 245, 755–758.

Taylor, W. R. (1999). TTX attenuates surround inhibition in rabbit retinal ganglion cells. *Vis. Neurosci.* 16, 285–290.

Teranishi, T., Negishi, K., and Kato, S. (1983). Dopamine modulates S-potential amplitude and dye-coupling between external horizontal cells in carp retina. *Nature* 301, 243–246.

Trumpler, J., Dedek, K., Schubert, T., de Sevilla Muller, L. P., Seeliger, M., Humphries, P., Biel, M., and Weiler, R. (2008). Rod and cone contributions to horizontal cell light responses in the mouse retina. *J. Neurosci.* 28, 6818–6825.

Umino, Y., Solessio, E., and Barlow, R. B. (2008). Speed, spatial, and temporal tuning of rod and cone vision in mouse. *J. Neurosci.* 28, 189–198.

van Hateren, J. H. (1997). Processing of natural time series of intensities by the visual system of the blowfly. *Vision Res.* 37, 3407–3416.

van Hateren, J. H., and Snippe, H. P. (2001). Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells. *Vision Res.* 41, 1851–1865.

van Nes, F. L., Koenderink, J. J., Nas, H., and Bouman, M. A. (1967). Spatiotemporal modulation transfer in the human eye. *J. Opt. Soc. Am.* 57, 1082–1088.

Victor, J. D. (1987). The dynamics of the cat retinal X cell centre. *J. Physiol.* 386, 219.

Wang, Y., and Mangel, S. C. (1996). A circadian clock regulates rod and cone input to fish retinal cone horizontal cells. *Proc. Natl. Acad. Sci. U.S.A.* 93, 4655–4660.

Weiler, R., Pottek, M., He, S., and Vaney, D. I. (2000). Modulation of coupling between retinal horizontal cells by retinoic acid and endogenous dopamine. *Brain Res. Brain Res. Rev.* 32, 121–129.

Werblin, F. S., and Dowling, J. E. (1969). Organization of the retina of the mudpuppy, *Necturus maculosus*. II. Intracellular recording. *J. Neurophysiol.* 32, 339–355.

Xin, D., and Bloomfield, S. A. (1999). Dark- and light-induced changes in coupling between horizontal cells in mammalian retina. *J. Comp. Neurol.* 405, 75–87.

Yang, Q. Z., and Hatton, G. I. (2002). Histamine H1-receptor modulation of inter-neuronal coupling among vasopressinergic neurons depends on nitric oxide synthase activation. *Brain Res.* 955, 115–122.

Yang, X. L., and Wu, S. M. (1989a). Effects of background illumination on the horizontal cell responses in the tiger salamander retina. *J. Neurosci.* 9, 815–827.

Yang, X. L., and Wu, S. M. (1989b). Modulation of rod–cone coupling by light. *Science* 244, 352–354.

Yang, X. L., and Wu, S. M. (1991). Feedforward lateral inhibition in retinal bipolar cells: input-output relation of the horizontal cell-depolarizing bipolar cell synapse. *Proc. Natl. Acad. Sci. U.S.A.* 88, 3310.

Zsiros, V., and Maccaferri, G. (2008). Noradrenergic modulation of electrical coupling in GABAergic networks of the hippocampus. *J. Neurosci.* 28, 1804–1815.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 October 2009; paper pending published: 10 December 2009; accepted: 27 February 2010; published online: 31 March 2010.

Citation: Pandarinath C, Bomash I, Victor J, Prusky G, Tschetter WW and Nirenberg S (2010) A novel mechanism for switching a neural system from one state to another. Front. Comput. Neurosci. ***4****:2. doi: 10.3389/fncom.2010.00002*

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 09 September 2010
doi: 10.3389/fncom.2010.00025

Wrestling model of the repertoire of activity propagation modes in quadruple neural networks

Hanan Shteingart[1,2]*, Nadav Raichman[2], Itay Baruchi[2] and Eshel Ben-Jacob[2]

[1] *Interdisciplinary Center for Neural Computation, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel*
[2] *School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel*

Edited by:
Philipp Berens,
Baylor College of Medicine, USA;
MaxPlanck Institute for Biological Cybernetics, Germany

Reviewed by:
Michal Zochowski,
University of Michigan, USA
Anna Levina, Max Planck Institute for Dynamics and Self-Organization, Germany
Philipp Berens, Baylor College of Medicine, USA; Max Planck Institute for Biological Cybernetics, Germany

***Correspondence:**
Hanan Shteingart, Silberman Building Room 3-342, The Hebrew University, The Edmond J. Safra Campus at Givat Ram, Jerusalem 91904, Israel.
e-mail: hanan.shteingart@mail.huji.ac.il

The spontaneous activity of engineered quadruple cultured neural networks (of four-coupled sub-networks) exhibits a repertoire of different types of mutual synchronization events. Each event corresponds to a specific activity propagation mode (APM) defined by the order of activity propagation between the sub-networks. We statistically characterized the frequency of spontaneous appearance of the different types of APMs. The relative frequencies of the APMs were then examined for their power-law properties. We found that the frequencies of appearance of the leading (most frequent) APMs have close to constant algebraic ratio reminiscent of Zipf's scaling of words. We show that the observations are consistent with a simplified "wrestling" model. This model represents an extension of the "boxing arena" model which was previously proposed to describe the ratio between the two activity modes in two coupled sub-networks. The additional new element in the "wrestling" model presented here is that the firing within each network is modeled by a time interval generator with similar intra-network Lévy distribution. We modeled the different burst-initiation zones' interaction by competition between the stochastic generators with Gaussian inter-network variability. Estimation of the model parameters revealed similarity across different cultures while the inter-burst-interval of the cultures was similar across different APMs as numerical simulation of the model predicts.

Keywords: microelectrode array, synchronous-bursting-event, burst-initiation zones, activity propagation, engineered neural networks, mutual synchronization, power-law scaling, Zipf-law

INTRODUCTION

MULTIELECTRODE ARRAYS AND SBEs

The human brain is considered to be one of the most complex systems and thus understanding the principles which underlie its activity requires simpler models (Koch and Laurent, 1999). Cultured neural networks with engineered geometry provided simple model systems for studying important motives of mutual synchronization and activity propagation (Baruchi et al., 2008; Raichman and Ben-Jacob, 2008; Raichman et al., 2009). Multielectrode arrays (MEA) have provided simple, tractable and efficient model systems for studying important motives of cultured networks and also provide a useful framework to study general information processing properties and specific basic learning mechanisms in the nervous system (Potter, 2001; Baruchi and Ben-Jacob, 2007; Chiappalone et al., 2007).

The spontaneous activity of many types of cultured networks is characterized by rapid collective neuronal firings called synchronized bursting events (SBEs) or "network bursts." These bursts last hundreds of milliseconds and are followed by longer (seconds) inter-burst-intervals (IBI) of sporadic firings (Segev et al., 2002; Raichman and Ben-Jacob, 2008) (**Figures 1A1, 1A2**). It was found that SBEs are important for the development of the nervous system, in the initiation of epileptic seizures, and in cortical integration of sensory information (Chiappalone et al., 2007).

There are a few suggested mechanisms for SBE activity, one of which is based on the hypothesized presence of localized initiation zones. These are characterized by high neuronal density and by recurrent and inhibitory network connections. This hypothesis is anchored on the fact that spontaneous activity is often observed to emanate from localized sources or burst-initiation zones (BIZ), propagating from them to excite large populations of neurons (Raichman and Ben-Jacob, 2008, reviews possible mecahnsimis).

Most of the firing activity is observed within a very short time window at the beginning of the SBE which is then followed by decay over longer period of time (Raichman et al., 2006). Moreover, each neuron in a SBE has its own temporal firing pattern which can greatly vary between different neurons but is usually consistent over days (Raichman and Ben-Jacob, 2008).

The capability of cultured neural networks to spontaneously generate repeating motifs on long time scales (hours) is highly significant for various applications. For example it affords neuronal networks *in vitro* to maintain long-term memory (Raichman and Ben-Jacob, 2008; Raichman et al., 2009). It was shown that printed (by local chemical stimulation) new activity motifs (activity propagation patterns) can also be maintained by the cultured networks for long times (Baruchi and Ben-Jacob, 2007). The number of motifs and the statistics of their appearance are connected with the architecture (topology, geometry and strengths of synaptic connections) of the network (Volman et al., 2005).

Large networks can generate few different SBEs, each with its own characteristic spatial-temporal pattern of activity propagation across the network (Hulata et al., 2004; Segev et al., 2004). Engineered coupled network, such as the quadruple networks

studied here, exhibit different types of mutual SBE, each with its own order of activity propagation between the sub-networks (Baruchi et al., 2008; Raichman and Ben-Jacob, 2008).

SBE SORTING

Dimensionality reduction clustering algorithms (e.g., principle component analysis) are used to identify and sort the different SBE motifs (sometimes referred to as *network repertoire*). These algorithms enable to simplify the representation of the network activity. In evoked activity experiments where the states of the system are expected due to the controlled stimuli, supervised algorithms can be used (Marom and Shahaf, 2002). However, in the case of spontaneous activity, only un-supervised methods are applicable.

In previous studies, identifications of the distinct SBEs were based on a measure of burst similarity (correlation) metric space. This similarity was defined either by (i) the firing intensity of individual neurons, with disregard of the temporal delays between neurons (Mukai et al., 2003; Madhavan et al., 2006) or (ii) by the time-space correlation between neuronal spike-trains (Segev et al., 2004). The latter approach enables to distinguish between bursts in which the firing profiles of the individual neurons are conserved but with different time delays between the activity of the different neurons. More recently, a delay similarity method was proposed (Baruchi et al., 2008; Raichman and Ben-Jacob, 2008). The method identifies repeating motifs that strictly depend on the delays between initiations of neuronal activity, while disregard the burst intensity and burst duration.

Despite the importance of timing, it has been shown that the information about evoked stimulus position can be retrievable just from the recruitment order, regardless of precise timing (Shahaf et al., 2008). Motivated by these observations we characterize here the different activity propagation modes (APM) of the mutual SBEs in terms of the order of activity propagation between the sub-networks (**Figure 1B1**).

It is believed that a central property of a complex system is the possible occurrence of coherent large scale collective behaviors with a very rich structure, resulting from the repeated non-linear interactions among its constituents.

Given such a complex system as neuronal network, a first standard attempt in order to quantify and classify the characteristics and the possible different dynamics consists in (i) identifying discrete events, (ii) measuring their features, and (iii) constructing their probability distribution (Sornette, 2007).

In our analysis, these discrete events are the SBE timings and their measured feature is the different APM assigned to each SBE.

POWER-LAW SCALING

Once identified, we investigated the network repertoire – the statistics of the frequency of appearance of the different APMs. The idea is that similar to the case of other complex systems, the statistics of system level events can provide important clues about the underlying mechanisms that regulate the network activity (Sornette, 2007).

Identification and understanding of such underlying mechanisms that regulate the activity of coupled neural networks can provide important clues on how to regulate, control and change the repertoire of such networks.

We found that half of the networks had algebraic scaling between the frequency of appearance of the leading (more frequent) APMs reminiscence of the Zipf's power-law scaling of words in natural language. During the 1930s Zipf showed that a power-law distribution described word counts in the English language (Zipf, 1932, 1935). A modern demonstration of this concept on Wikipedia's corpus has also been shown (Grishchenko, 2006). Research on the origins of the power-law and efforts to observe and validate them in the real world is extremely active in many fields of modern science, and seems to be a ubiquitous statistical feature of complex systems (Bak, 1996; Sornette, 2007).

There is a body of experimental (Beggs and Plenz, 2003, 2004; Petermann et al., 2009) and theoretical work (De Arcangelis et al., 2006; Kinouchi and Copelli, 2006; Levina et al., 2007) on occurrence of power-laws with cutoff in cultured neural networks. In these references the power-law scaling was of the time intervals between events (neuron firing and network bursts). Here, we investigated the statistics of the frequency of appearance of the different APM regardless of their timing.

THE WRESTLING MODEL

Toward the interpretation of the observed repertoire, we modeled the interplay between the intrinsic potential to fire of the different BIZ in terms of interacting "clocks" with variable rates. Once one BIZ fires, it stimulates the other BIZs and resets their "clock," thus disabling their initiation of spontaneous activity. This variable-clock game is an extension of the "boxing arena" model proposed for two coupled networks (Feinerman et al., 2007). Here we extended this work to multiple BIZs and used a Lévy distribution for the clock internal variability and Gaussian distribution for the inter-variability between the clocks. Using maximum likelihood we estimated the model's parameters and we observed similarity between parameters across cultures with different typical inter-burst-time intervals.

MATERIALS AND METHODS

CULTURE AND PREPROCESSING

The experimental protocol of the recordings of the coupled networks' activity which were analyzed here has been previously presented in details (Raichman and Ben-Jacob, 2008). We used six recorded cultures which are summarized in **Table 1** along with their characteristics.

Table 1 | Time characteristics of the recorded cultures.

#	T	N	B	D	A	IBI
A	3.5	49	1053	1.3 ± 0.5	0.8 ± 0.1	11 ± 5
B	49.7	19	748	1.7 ± 0.7	0.6 ± 0.1	250 ± 120
C	2.3	14	82	0.3 ± 0.4	0.3 ± 0.1	100 ± 100
D	23.0	49	5904	1.2 ± 0.6	0.7 ± 0.1	16 ± 6
E	47.2	37	9620	0.5 ± 0.3	0.6 ± 0.2	9 ± 7
F	31.0	16	4969	0.1 ± 0.1	0.3 ± 0.1	14 ± 10

#, Culture label; T, duration of recording (h); N, number of sorted neurons in the recording; B, total number of SBEs during the recorded time; D, SBE duration (s); A, SBE activity (fraction of participating neurons); and IBI, inter-burst-interval (s). All columns show mean ± standard deviation where applicable.

The networks were grown on MEA consisting of 60 round spot recording sites (each with diameter of 30 μm). The spatial organization was specially designed. The electrode array was consisted out of four clusters in the corners of a 1.8 mm × 1.4 mm rectangle. Each cluster was consisted of 13 equally spaced electrodes (250 μm). Other 7 electrodes were located in the regimes between the clusters.

Spike sorting of the extra-cellular recordings was based on wavelet packet decomposition (Hulata et al., 2002). This resulted in a (binary) time series of spike timings with a resolution of milliseconds for each identified neuron.

In order to identify the network bursts we followed the standard procedure of scanning the binary data of the network temporal spike activity in consecutive windows of 2 s, with a 50% overlap. Each window was divided into bins of 200 ms, and each bin was summed up over the number of active neurons. The timing of an SBE was defined as the time bin in which there were a maximum number of active neurons within the 2 s window. We ignored events that had less than 10–50% active neurons, or that were less than 5 s apart from the previously found SBE. Once an SBE was located, we used a pre-trigger and post-trigger of 2 s as the SBE time-support (Chiappalone et al., 2004; Raichman and Ben-Jacob, 2008).

IDENTIFICATIONS OF THE ACTIVITY PROPAGATION MODES

As was mentioned earlier, the APM are characterized by the order of activity propagation between the sub-networks (**Figure 1B2**). With four-coupled networks each APM is described by a permutation of the sequence [1234]. For example, $X = [1, 2, 3, 4]$ means that APM X was such that sub-network "1" fired first, then "2," "3" and lastly "4." Therefore, for four sub-networks there are $4! = 24$ different possible APMs.

Usually once a sub-network becomes active it does not relax and become active again within the same mutual SBE (representing a finite sub-network refractory period).

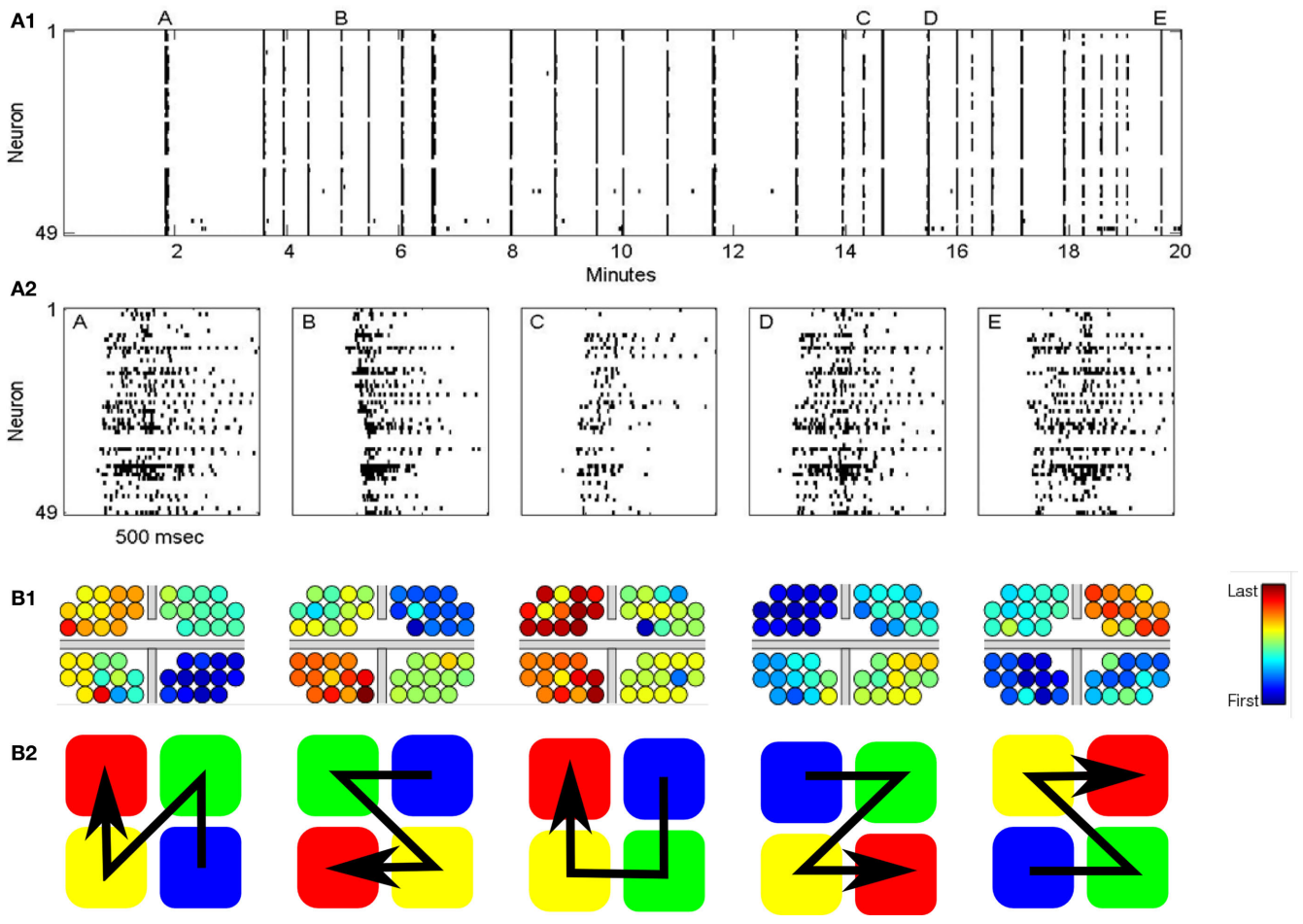


FIGURE 1 | (A1) Typical raster plot of the recorded activity of for coupled neural network**.** Each line corresponds to the recorded activity from a specific electrodes. Bars indicated neuronal firing. The results show the formation of mutual SBEs. **(A2)** Zoom in on the raster plot showing five distinct SBEs. **(B1)** Color code of the order of individual neuron firings within the different APMs from the first firing neuron (blue) to the last (red). Location and size of the electrodes is not in scale. Gray bars mark PDMS lines used to separate between the sub-networks. The order of activity of the four sub-networks follows a wave-like pattern, where the first firing group first activates each of the two nearest clusters. **(B2)** The activity order of the five leading (most frequent) activity propagation modes (APMs). **(A1)**, **(A2)**, and **(B1)** are reproduced with permission (Raichman and Ben-Jacob, 2008).

Timing the sub-network activity

Three different methods for timing the sub-network activity were tested: (1) average the firing time of first spikes ("first"), (2) center-of-mass of activity profile ("COM"), and (3) max firing rate ("max rate").

The "*first*" time is defined as: $t^i_{\text{first}} = 1/N_{\text{first}} \sum_{k=1}^{N_{\text{first}}} t^i_k$, where N_{first} is equals to the number of spikes that are considered to be "first" (this number is selected to optimize the measure), and t_k is the time of the kth spike in the ith sub-network.

The motivation to measure only the first spikes, is in line with results showing that spike timing is more accurate in the beginning of the spike-trains, both in spontaneous firing and in bursts generated as a response to electric stimuli. Moreover, it was suggested that bursts propagates as traveling waves where local networks act as the substrate of sequential firing patterns since activity which passes through a given point initiates similar local sequences. (Jimbo and Robinson, 2000; Bonifazi et al., 2005; Luczak et al., 2007; Raichman and Ben-Jacob, 2008; Shahaf et al., 2008).

The number of first spikes N_{first} introduces a tradeoff between robustness and accuracy. We chose the criterion for choosing N_{first} to be such that more bursts fall into the same motif, thereby identifying a smaller number of distinct APMs.

The "*center-of-mass*" time was defined as: $t_{\text{COM}} = 1/N_{\text{tot}} \sum_{k=1}^{N_{\text{tot}}} t_k$, where N_{tot} is the overall number of spikes fired by the sub-network. This method is based on the assumption that different sub-networks fire with similar patterns of firing rate. In this case, averaging the whole firing pattern can produce a fine and robust measure of the sub-network firing pattern. This method assigns larger weight to time periods with higher firing rates in the weighted average. The reason is that such time periods are relatively less noisy (assuming Poisson noise).

The "*max rate*" time was defined using a histogram: $b_m = \sum_{k=1}^{N_{\text{tot}}} \theta(t_k - m\Delta t) - \theta((m+1)\Delta t - t_k)$, where θ is the Heaviside function and Δt the histogram resolution (1 ms).

The estimated maximum rate time was defined at the center of the histogram maximum: $t_{\text{maxrate}} = \Delta t(\arg\max_m b_m + 0.5)$. The motivation for this measurement is to order activity of the different sub-networks by the delays of the maximum local activity. The idea is that the first spikes describe the propagation front of the neural signal, but once each sub-network is activated it has its own internal activity propagation.

In the results section we compare between these three timing methods. We then selected the method that yielded the least distinction entropy between the different APMs, following the idea of minimum-entropy data partitioning (Roberts et al., 2001).

Consistency test

In order to test for consistency of our method of identification of the APMs, we compared it with the correlation and delay similarity methods mentioned in the introduction.

The *correlation* method is based on a binary activity matrix $A^i{}_{N\times T}$ representation of a SBE where $N$is the number of neurons and T is the number of time bins. The element is $A^i_{n,t} = 1$ if neuron n fired during the time bin t in SBE $i = \{1 \ldots N_{\text{SBE}}\}$ (zero otherwise). First, the activity vector of each neuron $(A^i_n(t))$ is convolved with a normalized Gaussian kernel with width adjusted to the firing rate in order to obtain a smooth rate representation $D^i_n(t)$ (typically ~50 ms). Next we define a normalized Pearson's cross correlation $C^{i,j}_n(\tau)$ between the bursts couple (i, j) per neuron with time displacement τ:

$$C^{i,j}_n(\tau) = \left(\frac{\sum_{t=1}^{T} \left(D^i_n(t) - \bar{D}^i_n\right)\left(D^j_n(t-\tau) - \bar{D}^j_n\right)}{\sqrt{\sum_{t=1}^{T} \left(D^i_n(t) - \bar{D}^i_n\right)^2 \sum_{t=1}^{T} \left(D^j{}_{n(t)} - \bar{D}^j_n\right)^2}} \right)$$

Finally, a max correlation matrix EC is defined as the maximum (over τ) of the sum (over neurons) of the correlation $C^{i,j}_n(\tau)$

$$EC(i,j) = \arg\max_{\tau} \left(\sum_{n=1}^{N} C^{i,j}_n(\tau) \right)$$

This correlation matrix can be interpreted as representation of N_{SBE} bursts in a N_{SBE} dimension space. This metric space is then clustered using the dendrogram algorithm tree – an agglomerative hierarchical cluster technique based on distances (Mathworks, 2009). This clustering method allows sorting of the SBEs into different modes, each with its own pattern of correlations between the neuron firings.

The *delay similarity* method is based on a delay activation matrix B such that $B^i_{n,m}$ is the delay between the first spikes of neuron n and m in the ith SBE. Neurons that did not fire in the particular SBE are assigned a NULL value in the activation matrix. The similarity between bursts is than defined as:

$$S(i,j) = \frac{1}{N(N-1)} \sum_{n \neq m} \theta\left(\tau_0 - \left|A^i(n,m) - A^j(n,m)\right|\right)$$

Where θ is the Heaviside function and τ_0 is a threshold which is set to 30 ms following the average spike precision in bursts (Bonifazi et al., 2005). The method detects the center (motive) of SBE clusters with high similarity, by applying a two-stage method that uses a hierarchical clustering algorithm followed by an iterative search for independent cluster centers.

For comparison between the correlation and delay similarity methods with our approach, we used the previous methods' metric matrix and reordered these matrices in three different fashions: (i) by clustering with respect to the same metric space (e.g., correlation metric reordered according to the correlation space and vice versa), (ii), we repeated the procedure with the alternative metric space (e.g., delay metric reordered based on correlation metric and vice versa), (iii) we reordered the metrics by the APMs that our method identified.

In the result section we show that, while our method is consistent with the other two methods, it is more efficient.

POWER-LAW TESTING AND ESTIMATION

To quantify the finite scaling of the frequency of appearance of the different APMs we followed and extended the method of (Goldstein et al., 2004; Clauset et al., 2009). The estimation and significant testing of the power-law ($p(k) \propto k^{-\beta}$) distribution's parameter (β) have been extended for the case of the observed finite power-law. This is defined as a finite repertoire of motifs' (alphabet) distribution which follows the power-law only for $k < M$ where k is the event frequency rank.

We focus only on estimating the power parameter, while M was fixed and chosen such that it differentiated between APMs with frequency higher and lower than that calculated for the limit of uniform frequency of appearance.

The details of the estimation and testing is detailed in the Appendix (see Power-Law Testing and Estimation).

THE WRESTLING MODEL

We developed a semi-realistic model which recovers quite efficiently the observed statistical behaviors of the APMs repertoire. This "wrestling" model is an extension for the "boxing arena model" which was proposed for two coupled networks (Feinerman et al., 2007).

The central assumption is that each sub-network has several BIZs and they all "compete" to be the first to initiate a mutual SBE. We assume that each sub-network, had it been isolated from the other sub-networks, have its own innate mean time between SBEs. In other words, each sub-networks a stochastic SBEs generator with its own innate "clock." However, since the statistics of IBI (inter-bursts-intervals) follows a Lévy distribution (Segev et al., 2002; Ayali et al., 2004), the definition of the generator and the clock have to be done with extra care. In the model used here, we used a stochastic generator that generates a Lévy flight process (Chambers et al., 1976). The generators of the different sub-networks had the same α (slope) and γ (variability) parameters while each sub-network had its own most probable IBI (the δ parameter). The symmetry parameter β_{Levy} was set to zero (no drift). This difference in δ was generated from a Gaussian distribution with zero mean and STD of σ^2_{inter} and was rolled once before each simulation.

The normalized version of the model has two parameters only: first, the variability variance ratio (η) which is the IBI's intra-variability variance normalized by the inter-variability variance. Secondly, the Lévy's distribution power coefficient (α). The details of the model simulation and the procedure of parameter estimation are described in the Appendix (see Model Details and its Parameter Estimation).

RESULTS

CONSISTENCY TEST

We analyzed the activity of six cultures all having similar structure of four-coupled sub-networks (see Materials and Methods). All of these cultures showed global synchronization marked by the existence of mutual SBEs. First, we show a typical sorting of the different SBEs using the methods of correlation and delay similarity (see Materials and Methods). Then, we compared this similarity/correlation metric matrix when reordered by our new characterization approach.

The new method provides an efficient and clear sorting of SBE into distinct motives of APMs which can be seen as areas of strong intra-group and weak inter-group delay similarity (**Figure 2**). It is worthwhile noting that although this method achieved a

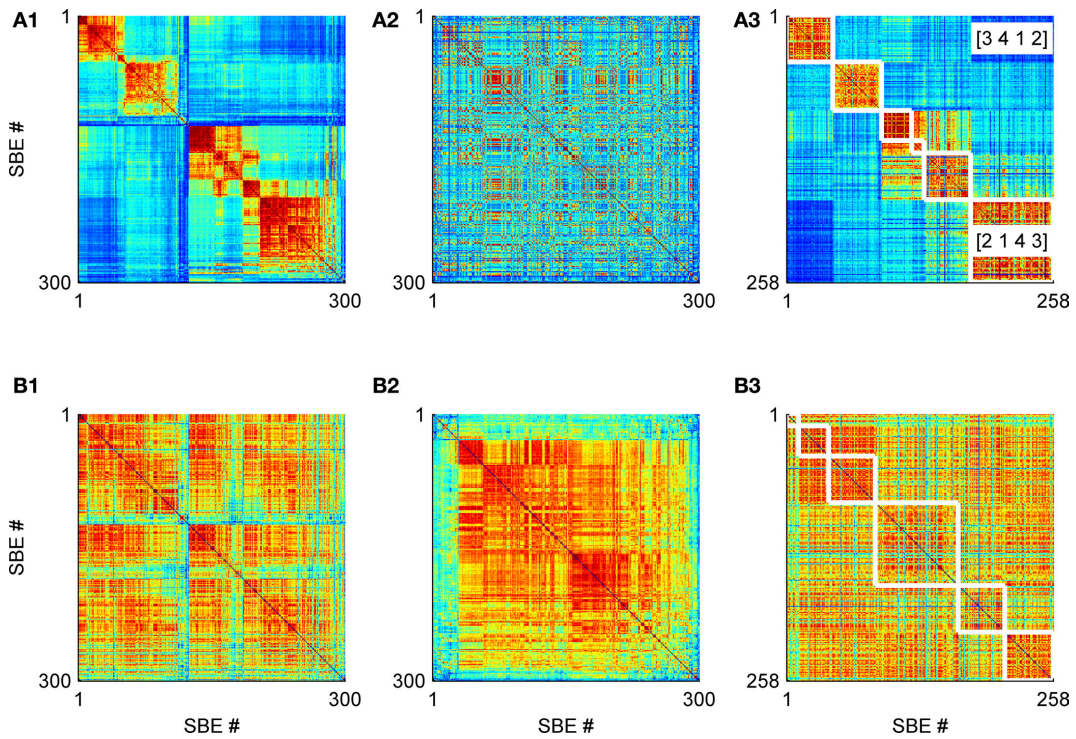


FIGURE 2 | Comparison of delay similarity (row (A)) and correlation (row (B)) metric matrices with culture #A (see Materials and Methods). The *x*- and *y*-axis are 300 SBEs indices while the color at each pixel represents the normalized metric value (red-high, blue-low). These metrics were reordered according to three different permutations (1–3 columns): (1) delay similarity, (2) correlation matrix, and (3) our APM method. The latter APMs were bordered (white line) and rearranged (without breaking groups) to fit bets the permutation of the relevant metric (row **(A)** or **(B)**). Notice the sequence [3412] and [2143] have opposite activation order and thus have low cross delay similarity metric values.

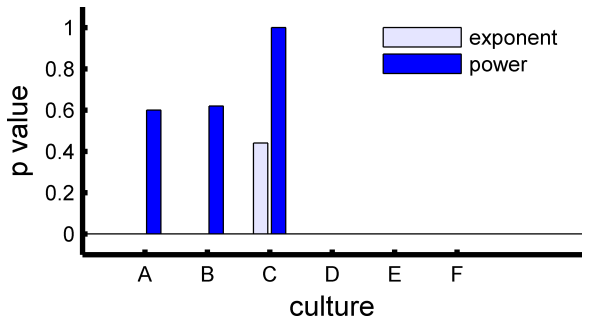


FIGURE 3 | Comparison between the significance levels of a power-law model versus an exponential one. Half of the networks showed a significance *p*-value such that the null hypothesis of a power-law distribution (with cutoff) cannot be declined. On the other hand, only 1/6 of the networks showed similar significance for the exponential model.

precise classification, it is much less computationally demanding than previous techniques based on metric estimation since the computation of the $O(N_{SBE}{}^2)$metric space is not needed.

We found that two APMs with a reverse order of activity propagation, such as the APMs [2, 1, 4, 3] and [3, 4, 1, 2] (**Figure 2A3**), show low delay similarity. And, two APMs in which the activity starts at the same sub-network show high delay similarity.

ASSESSMENT OF THE ACTIVITY TIMING METHODS

Comparison between the different methods of activity timing revealed that the "first" timing method (based on the firing time of the first few spikes), yields the statistically most significant sorting of the APMs. The statistical significance was assessed by calculating the minimum-entropy data partitioning approach (Roberts et al., 2001). In this approach the entropy of the frequency of appearance of the APMs is calculated (the relative frequency of appearance of each APM is taken as its probability). A distribution with lower entropy corresponds to sorting that is more statistically significant (higher deviation from a uniform distribution).

We found that the best sorting by the "first" timing method is obtained when the first five spikes are taken ($N_{first} = 5$) as is shown in **Figure 4**.

FREQUENCY OF APPEARANCE

Half of the networks (three out of six) expressed scaling consistent with a finite power-law with *p*-value higher than 20% (**Figures 5A–C** cultures). The three other networks only showed power-law scaling for the leading APMs. Moreover, we compared the power-law with an alternative exponential model. The exponential model was rejected in five out of six networks (**Figure 3**).

Note that all networks deviated greatly from what would be expected from a uniform distribution of the frequency of appearance (black transparent patches in **Figure 5**).

EMPLOYING THE WRESTLING MODEL

The results of the wrestling model simulation are in good agreement with the observed distributions. The level of the agreement indicates that the model may explain some observed features and in particular the four orders of magnitude ratios in the frequency of appearance (see **Figure 5** *y*-axis) and the finite cutoff in the power-law scaling.

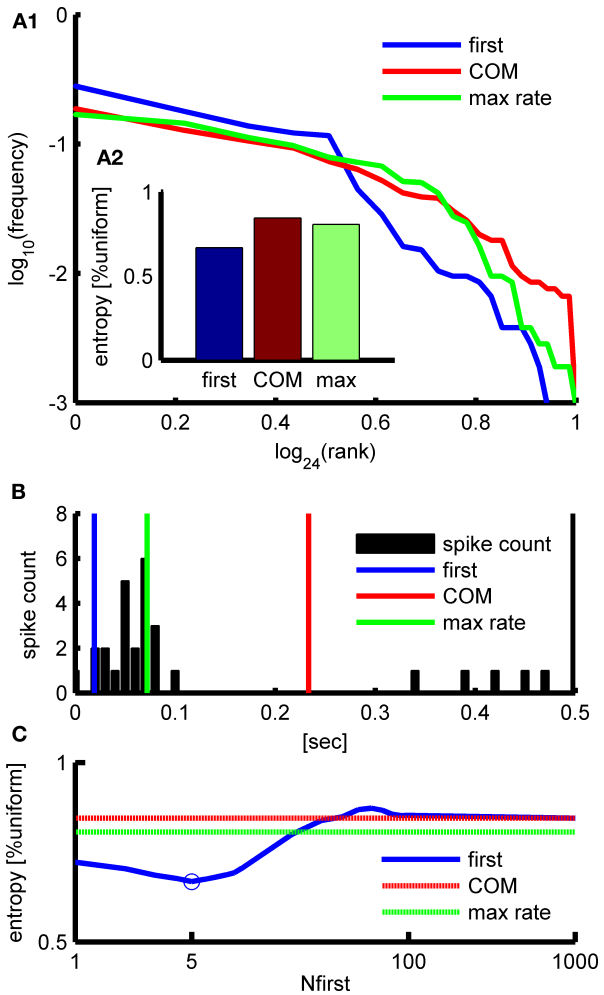


FIGURE 4 | Comparison between the different methods to estimate the APMs' timing on culture #A (see Materials and Methods). (A1) The frequency of appearance (log) of the frequency of appearance of the APMs using the different methods is similar. **(A2)** The "first" method shows the closer to power-law like behavior and the entropy. **(B)** The sub-network activity timing can be measured by averaging the first N_{first} spikes (blue), by the center-of-mass (COM) of activity (red), or by the maximum firing rate (green) as can be seen plotted on the activity profile histogram (bin size 100 ms). **(C)** The "first" method entropy is minimized at $N_{first} = 5$.

Although the different networks had very different time scales and activity, we estimated similar parameters for the different sub-networks.

To avoid confusion we note that there is the set of parameters of the Lévy distribution of the IBI generated by the stochastic generator and the set of parameters of the distribution of the frequency of appearance of the APMs.

The set of parameters for the IBI distributions are defined as: (1) the stability parameter (the slope of the tail of the distribution) – the α parameter of the Lévy distribution. (2) the scaling parameter η, that is equal to the ratio between the variability parameter of the generated IBI sequences (related to the generalization of the STD for Lévy distribution) and the variability (σ_{inter}) in the mean IBI of the generated IBI sequences of the four

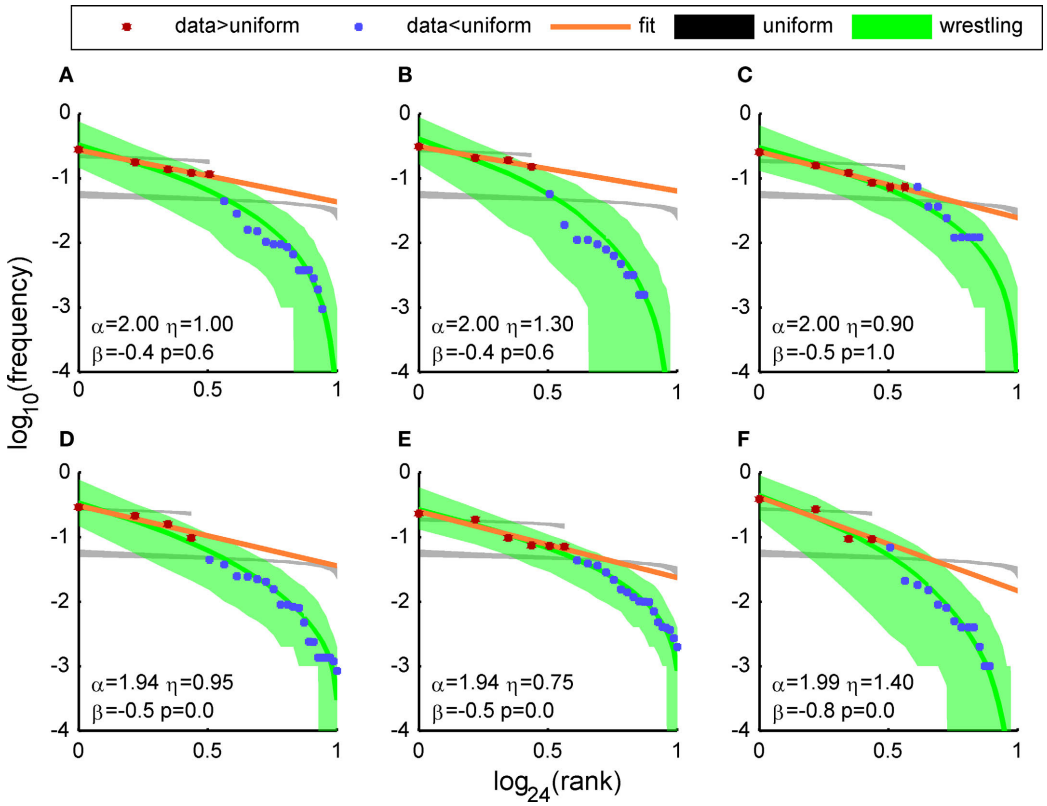


FIGURE 5 | Distributions of the frequency of appearance of the APMs of six different four coupled networks (labeled (A) to (F)). The dots represent the frequency of appearance of the APMs on a log scale, ranked from the most frequent one and down The red dots are the leading (most frequent) APMs – the ones which were more frequent than random (1/24) and blue are the APMs with lower frequency of appearance (including one additional point which was above it). The black patches represent the 1–99% Monte Carlo simulation of a ranking a uniform distribution with alphabet size equal to the maximum number of possible words (24) and to the number of words above uniform (red) (population size = number of culture SBEs and 10,000 repeated simulations). The orange line is the maximum likelihood fit to a power-law model of the points above uniform (red). The Green patches represent the 1–99% Monte Carlo simulation of wrestling model simulation with parameters which were estimated according to maximum likelihood measurement on a grid search (see Materials and Methods). The green line is the mean value of the patch. The number summarize the wrestling arena model simulation parameters (α, η), the power-law estimated slope (β) and the power-law test *p*-value (*p*). Notice that also the cultures had very different time properties (e.g., mean IBI lasting from 10 to 250 s; see Materials and Methods) the global distribution parameters are similar and consistent across cultures. Moreover, the maximum likelihood wrestling model fit gives higher values to more frequent events, but nevertheless the model fits nicely also to rare events upto 1 to 1000. In order to emphasize the break of symmetry in the frequency of appearance of the different SAO, we also compared the apparent distribution to a ranking of a uniform distribution. We repeated this first with an alphabet size equal to the maximize number of possible sequence (4! = 24) and secondly with an alphabet size equal to the number of SAO which were more frequent then uniform.

different sub-networks is. It is important to note that η equals to 1 for the case that the internal variability of the IBI sequences and the variability between the sub-networks are comparable. The symmetry parameter $\beta_{\text{Lévy}}$ was set to zero (no drift) and should not be confused with β used here that is the slope of the algebraic (power-law) part of the distribution of the APMs' frequency of appearance.

Employing the wrestling model, we found that the parameters that fits the observations were: $\overline{\alpha} = 1.98 \pm 0.01$, $\overline{\eta} = 1 \pm 0.1$ (±SEM) for the Levy distribution and $\overline{\beta} = -0.52 \pm 0.06$ for the power-law slope (± SEM). Note that the Lévy slope was almost 2 which is on the edge of Gaussian. We note the similarity across cultures by measuring the coefficient of variance of the model parameters: $CV(\alpha, \beta, \gamma) = (0.01, 0.28, 0.24)$. We also note that, five out of six cultures passed a leave-one-out multi-variant ANOVA test with 5% threshold with the null hypothesis being the same parameters mean.

An additional important observation is related to the value of the inter-burst-time interval (IBI) which preceded the appearance of each APM. The wrestling model predicts that the distribution of the observed bursts IBIs would be of the same order of magnitude for the different sub-networks in agreement with the experimental observations.

In other words, the mean IBI of a specific BIZ conditioned of it being the shortest one in the current round is smaller than the aggregated IBI mean and is comparable to the most frequent BIZ's IBI:

$$\left\langle \mathrm{IBI}_i^r \middle| i = \arg\min_j \mathrm{IBI}_j^r \right\rangle_r \sim \arg\min_j \left\langle \mathrm{IBI}_j^r \right\rangle < \left\langle \mathrm{IBI}_i^r \right\rangle_r$$

We compare the simulation result to the real data by treating each APM as it was generated by different IBI and in both the "winning" IBIs is relatively flat (**Figure 6**).

DISCUSSION AND SUMMARY

We showed that four-coupled cultured networks exhibit mutual SBEs with a reach repertoire of APM, each with a distinct order of activity propagation between the sub-networks. Investigations of the frequency of appearance of the APMS revealed power-law scaling between the several leading (most frequent) ones. In complex systems, power-law scaling can be a manifestation of hierarchy and robustness (Sornette, 2007). The non-uniform nature of Finite power-law suggests some kind of control mechanism that prevents a winner-takes-all scenario by the most active sub-network (so it does not generate almost all the mutual SBE).

We introduced a "wrestling model" to account for the observations. Simulations of the model to fit with the observations revealed that the scaling parameter η has to be close to 1. This result indicates that the intrinsic variability in the IBI sequences generated by the sub-networks is regulated to fit the variability in the mean IBI between the different sub-networks.

This result ($\bar{\eta} = 1 \pm 0.1$) suggests that there must be some unknown mechanism which can co-regulate the local intra-variability and the global inter-variability to be comparable.

One possible mechanism might be related to the propagation of calcium waves in the astrocytes. It has been proposed that astrocyte calcium waves may constitute a long-range signaling system within the brain (Cornell-Bell et al., 1990).

The calcium waves can be regulated by the rate of activity of the different sub-networks and in turn regulates the effective synaptic strengths. Since they have a long time scales and can propagate over long distances, the calcium waves might provide a mechanism that couples the intrinsic scaling of IBI and the global variability between the different sub-networks. The possible role of calcium wave can be tested experimentally by testing the effect of regulations of the astrocyte calcium wave's dynamics on the frequency of appearance of the APMs.

Finally we would like to note that the similarity in the model's parameters across cultures might suggests that these are invariants of the culture network.

APPENDIX

POWER-LAW TESTING AND ESTIMATION

To quantify the finite scaling of the frequency of appearance of the different APMs we followed and extended the method of (Goldstein et al., 2004; Clauset et al., 2009). The estimation and significant testing of the power-law ($p(k) \propto k^{-\beta}$) distribution's parameter (β) have been extended for the case of the observed finite power-law. This is defined as a finite repertoire of motifs' (alphabet) distribution which follows the power-law only for $k < M$ where k is the event frequency rank. We focus only on estimating the power parameter, while M was fixed and chosen such that it differentiated between APMs with frequency higher and lower than that calculated for the limit of uniform frequency of appearance.

Finite power-law estimation

We estimated the power-law only on a subset of the distribution at $k < M$. The justification for this approach is as follows: if the subset accumulates q fraction of the whole distribution P_0, the subset distribution P is related to it as $P = qP_0$. Since maximum likelihood estimation maximizes P and since q is independent on the distribution parameters (β), it can be omitted and P can be treated as if it was the real distribution.

Assuming that the one sample distribution is defined as:

$$P(k|\beta, M) = k^{-\beta}\left(\zeta^M(\beta)\right)^{-1} \qquad k \in \{1,2\ldots M\}$$

where $\zeta^M(\beta) = \sum_{k=1}^{M} k^{-\beta}$ is the partition function for the case of M discrete values and the exponent β. Note that this is not the real partition function and it does not normalize the whole distribution rather only the power-law subset part.

If the measurements are statistically independent, the log likelihood (λ) of β with N observations $\{k_i\}_{i=1}^{N}$ can be written as $\Lambda(\beta) = \log\lambda(\beta) = -N\log\zeta_\beta^M - \beta\sum_{n=1}^{N}\log k_n$. To find its minimum, we differentiate with respect to the parameter and get the ML estimator:

$$\frac{\partial\Lambda\left(\beta, M, \{k_i\}_{i=1}^{N}\right)}{\partial\beta} = -N\frac{\left(\zeta^M(\beta)\right)'}{\zeta^M(\beta)} - \sum_{n=1}^{N}\log k_n = 0$$

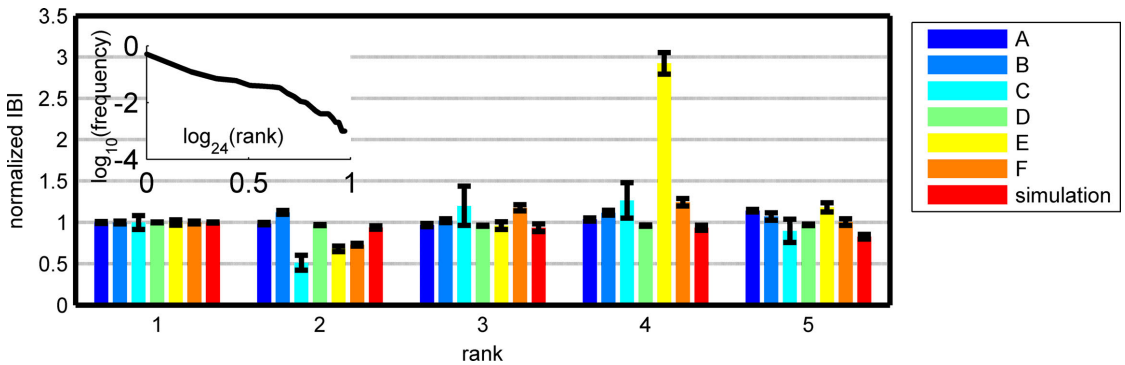


FIGURE 6 | Mean IBI just before one of the five most frequent (highest rank) APMs of six different networks and a simulation of the "wrestling arena." We normalized these IBI according the mean IBI of the most frequent APMs (thus the most frequent mean is "1") (errors are SEM). It seems that the IBI before each APM is the same for the different leading APMs despite the high variability of the IBIs. The wrestling model predicts this (typical simulation $\alpha = 1.9$, $\eta = 1$, $N = 1000$). The model predicts that if a slower BIZ wins over the most frequent APM its IBI is shorter than usual and of the same order of magnitude as the most frequent APM (or smaller). Note that the wrestling simulation can produce a finite power-law like behavior (not in all runs) with a cutoff similar to the observed data.

where we mark the partition function derivation as $(\zeta^M(\beta))' = \frac{d}{d\beta}\zeta^M(\beta) = -\sum_{k=1}^{M}\log(k)k^{-\beta}$.

Thus we get the implicit expression for the estimated power $\hat{\beta}$:

$$\frac{\zeta'_M\left(\hat{\beta}\right)}{\zeta_M\left(\hat{\beta}\right)} = -\frac{1}{N}\sum_{n=1}^{N}\log k_n$$

which can be solved numerically for every observation set $\{k_i\}_{i=1}^{M}$ and fixed M.

Power-law testing

Observing an approximately straight line on a log-log plot is a necessary but not sufficient condition to indicate power-law scaling (Clauset et al., 2009). Thus, we tested the hypothesis of power-law distribution in a statistically significant manner using goodness-of-fit test based on Kolmogorov–Smirnov (KS) statistic test (Goldstein et al., 2004). A significant p-value for this test (typical more than 0.05) means that the power-law null hypothesis cannot be rejected which means that the data is compatible with the null hypothesis.

To avoid estimation bias, we used a Monte Carlo calibration process in which we drew a large number ($n \sim 10^3$) of synthetic data sets from different power-laws distributions with uniform random slope α in the range [0, 1] ($\alpha \sim U(0, 1)$) of discrete alphabet size $M \in [3, 24]$. Then we fitted each one individually to the power-law model (see previous subsection) and calculated the KS statistic for each one relative to its own best-fit model. We then measured the test's p-value by estimating the fraction of trials which had a KS value larger than the observed one.

To summarize, in order to test the hypothesis that the observed data set is drawn from a power-law distribution one should: (1) Determine the best fit of the power-law to the data by estimating the scaling parameter β using the ML method, (2) Calculate the KS statistics for the goodness-of-fit of the best-fit power-law to the data, (3) generate a large number of synthetic data sets. Fit each according to the ML method, and calculate the KS statistic for each fit, (4) calculate the p-value as the fraction of the KS statistics for the synthetic data sets whose value exceeds the KS statistic for the real data, (5) If the p-value is sufficiently small, the power-law distribution can be ruled out (Clauset et al., 2009).

An analysis of the expected worst-case performance of the method produced a rule of thumb for determine the number of trials (n): if the p-values is to be accurate to within about ε of the true value, then at least $(2\varepsilon)^{-2}$ synthetic data sets should be generated (Clauset et al., 2009). If, for example, the p-value is accurate to about two decimal points, $\varepsilon = 0.01$ should be chosen, which implies a generation of about 2500 samples minimum. In our test, we used 10,000.

MODEL DETAILS AND ITS PARAMETER ESTIMATION

At each round ($r = [1, N_r]$, $N_r = 1000$) the ith BIZ produces a random IBI according to the sum of its own mean IBI $\mu_i \propto N(0, \sigma_{\text{inter}})$ and its centered Lévy distribution realization $l_i^r \propto L(\eta, \alpha)$ such that $\text{IBI}_i^r = \mu_i + l_i^r$. At each round the winning BIZ (W^r) is the one with the shortest IBI: $W^r = \arg\min_i(\text{IBI}_i^r)$. The histogram $H_i = \sum_{r=1}^{N_r}\delta(W^r, i)$ of winning BIZs is then sorted (ranked) and normalized where $\delta(i, j)$ being the Kronecker's delta. The probability distribution of the different ranks is thus simply $p_i = (\sum_{k=1}^{M} H_k)^{-1} H_i$.

We estimated the model parameters by a grid search ($\alpha = [1{:}0.01{:}2]$, $\eta = [0.5{:}0.05{:}2]$). For each couple of parameters, we estimated the probability distribution of each rank $\{p_k\}_{k=1}^{M}$. Then we computed the log likelihood by modeling the observed frequency of rank $\{x_k\}_{k=1}^{M}$ as multinomial distribution. We found $(\alpha_{\text{ML}}, \eta_{\text{ML}})$ which maximize the log likelihood Γ using the relation $\Gamma = \log\gamma \propto \sum_{i=1}^{M} x_i \log p_i$. For the ML estimated parameters' values we computed the 98% (1–99%) confidence interval of the frequency of appearance for each rank (from the same Monte Carlo simulation).

Obviously, when the scaling parameter η is large we expect only one BIZ to win ("winner-takes-it-all" scenario), thus producing a delta function in the ranked SAO distribution. However, if it is small, it would create a uniform distribution since all BIZ are equally likely to "win." We claim that only variability ratio of the order of one ($\eta \sim 1$) can explain the observed SAO distribution which is neither uniform nor exclusive.

We note that η is somehow problematic as for α lower than 2, the variance of the Lévy's distribution diverges. Therefore, we used the empiric variance and normalized it according to σ^2_{inter}.

ACKNOWLEDGMENTS

We are thankful to Liel Rubinsky and Mark Shein. One of us HS thanks Prof. Hagit Messer-Yaron for partial support during part of this research. This research has been supported in part by the Tauber Family Foundation and the Maguy-Glass Chair in Physics of Complex Systems at Tel Aviv University.

REFERENCES

Ayali, A., Fuchs, E., Zilberstein, Y., Robinson, A., Shefi, O., Hulata, E., Baruchi, I., and Ben-Jacob, E. (2004). Contextual regularity and complexity of neuronal activity: from stand-alone cultures to task-performing animals. *Complexity* 9, 25–32.

Bak, P. (1996). *How Nature Works: The Science of Self-Organised Criticality*. New York, NY: Copernicus Press.

Baruchi, I., and Ben-Jacob, E. (2007). Towards neuro-memory-chip: imprinting multiple memories in cultured neural networks. *Phys. Rev. E, Stat. Nonlin. Soft Matter Phys.* 75, 509011–509014.

Baruchi, I., Volman, V., Raichman, N., Shein, M., and Ben-Jacob, E. (2008). The emergence and properties of mutual synchronization in in vitro coupled cortical networks. *Eur. J. Neurosci.* 28, 1825–1835.

Beggs, J. M., and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177.

Beggs, J. M., and Plenz, D. (2004). Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures. *J. Neurosci.* 24, 5216–5229.

Bonifazi, P., Ruaro, E., and Torre, V. (2005). Statistical properties of information processing in neuronal networks. *Eur. J. Neurosci.* 22, 2953–2964.

Chambers, J. M., Mallows, C. L., and Stuck, B. W. (1976). A method for simulating stable random variables. *J. Am. Stat. Assoc.* 71, 340–344.

Chiappalone, M., Novellino, A., Vato, A., Martinoia, S., Vajda, I., and van Pelt, J. (2004). *Analysis of the Bursting Behavior in Developing Neural Networks*. 2nd International Symposium on Measurement, Analysis and Modeling of Human Functions, Genova.

Chiappalone, M., Vato, A., Berdindini, L., Koudelka-Hep, M., and Martinoia, S. (2007). Network dynamics and synchronous activity in cultured cortical neurons. *Int. J. Neural. Syst.* 17, 87–103.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.* 51, 661–703.

Cornell-Bell, A. H., Finkbeiner, S. M., Cooper, M. S., and Smith, S. J. (1990). Glutamate induces calcium waves in

cultured astrocytes: long-range glial signaling. *Science* 247, 470–473.

De Arcangelis, L., Godano, C., Lippiello, E., and Nicodemi, M. (2006). Universality in solar flare and earthquake occurrence. *Phys. Rev. Lett.* 96, 051102–051105.

Feinerman, O., Segal, M., and Moses, E. (2007). Identification and dynamics of spontaneous burst initiation zones in unidimensional neuronal cultures. *J. Neurophysiol.* 97, 2937–2948.

Goldstein, M. L., Morris, S. A., and Yen, G. G. (2004). Problems with fitting to the power-law distribution. *Eur. Phys. J. B* 41, 255–258.

Grishchenko, V. (2006). *Wikipeida/Zipf's_law. License, GNU Lesser General Public.* Available at http://en.wikipedia.org/wiki/Zipf%27s_law.

Hulata, E., Baruchi, I., Segev, R., Shapira, Y., and Ben-Jacob, E. (2004). Self-regulated complexity in cultured neuronal networks. *Phys. Rev. Lett.* 92, 198181–198104.

Hulata, E., Segev, R., and Ben-Jacob, E. (2002). A method for spike sorting and detection based on wavelet packets and Shannon's mutual information. *J. Neurosci. Methods* 117, 1–12.

Jimbo, Y., and Robinson, H. P. C. (2000). Propagation of spontaneous synchronized activity in cortical slice cultures recorded by planer electrode arrays. *Bioelectrochemistry* 51, 107–115.

Kinouchi, O., and Copelli, M. (2006). Optimal dynamical range of excitable networks at criticality. *Nat. Phys.* 2, 348–351.

Koch, C., and Laurent, G. (1999). Complexity and the nervous system. *Science* 284, 96–98.

Levina, A., Herrmann, J. M., and Geisel, T. (2007). Dynamical synapses causing self-organized criticality in neural networks. *Nat. Phys.* 3, 857–860.

Luczak, A., Bartho, P., Marguet, S. L., Buzsaki, G., and Harris, K. D. (2007). Sequential structure of neocortical spontaneous activity in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 104, 347–352.

Madhavan, R., Chao, Z. C., and Potter, S. M. (2006). *Spontaneous Bursts are Better Indicators of Tetanus-Induced Plasticity than Responses to Probe Stimuli*. Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference, Arlington, VA, V–VIII.

Marom, S., and Shahaf, G. (2002). Development, learning and memory in large random networks of cortical neurons: lessons beyond anatomy. *Q. Rev. Biophys.* 35, 63–87.

Mathworks. (2009). MATLAB Statistics Toolbox. Natick, MA: Mathworks.

Mukai, Y., Shina, T., and Jimbo, Y. (2003). Continuous monitoring of developmental activity changes in cultured cortical networks. *Electr. Eng. Jpn.* 145, 28–37.

Petermann, T., Thiagarajan, T. C., Lebedev, M. A., Nicolelis, M. A. L., and Chialvo, D. R. (2009). Spontaneous cortical activity in awake monkeys composed of neuronal avalanches. *Proc. Naatl. Acad. Sci. U.S.A.* 106, 15921–15926.

Potter, S. M. (2001). Distributed processing in cultured neuronal networks. *Prog. Brain Res.* 130, 49–62.

Raichman, N., and Ben-Jacob, E. (2008). Identifying repeating motifs in the activation of synchronized bursts in cultured neuronal networks. *J. Neurosci. Methods* 170, 96–110.

Raichman, N., Rubinsky, L., Shein, M., Baruchi I., Volman V., and Ben-Jacob, E. (2009). "Cultured neuronal networks express complex patterns of activity and morphological memory," in *Handbook of Biological Networks*, eds S. Boccaletti, V. Latora, and Y. Moreno Singapore: (World Scientific), 257–278.

Raichman, N., Volman, V., and Ben-Jacob, E. (2006). Collective plasticity and individual stability in cultured neuronal networks. *Neurocomputing* 69, 1150–1154.

Roberts, S. J., Holmes, C., and Denison, D. (2001). Minimum-entropy data partitioning using reversible jump markov chain monte carlo. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 909–914.

Segev, R., Baruchi, I., Hulata, E., and Ben-Jacob, E. (2004). Hidden neuronal correlations in cultured networks. *Phys. Rev. Lett.* 92, 118102.

Segev, R., Benveniste, M., Hulata, E., Cohen, N., Palevski, A., Kapon, E., Shapira, Y., and Ben-Jacob, E. (2002). Long term behavior of lithographically prepared in vitro neuronal networks. *Phys. Rev. Lett.* 88, 1181021–1181024.

Shahaf, G., Eytan, D., Gal, A., Kermany, E., Lyakhov, V., Zrenner, C., and Marom, S. (2008). Order-based representation in random networks of cortical neurons. *PLoS Comput. Biol.* 4, e1000228. doi: 10.1371/journal.pcbi.1000228.

Sornette, D. (2007). *Probability Distributions in Complex Systems.* arXiv:0707.2194, ARXIV.

Volman, V., Baruch, I., and Ben-Jacob, E. (2005). Manifestation of function-follow-form in cultured neuronal networks. *Phys. Biol.* 2, 98–110.

Zipf, G. (1932). *Selective Studies and the Principle of Relative Frequency in Language*. Cambridge, MA: MIT Press.

Zipf, G. (1935). *Psycho-Biology of Languages*. Boston, MA: Houghton-Mifflin (1934), MIT Press (1965).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 December 2009; paper pending published: 29 January 2010; accepted: 14 July 2010; published online: 09 September 2010.
Citation: Shteingart H, Raichman N, Baruchi I and Ben-Jacob E (2010) Wrestling model of the repertoire of activity propagation modes in quadruple neural networks. Front. Comput. Neurosci. ***4****:25. doi: 10.3389/fncom.2010.00025*