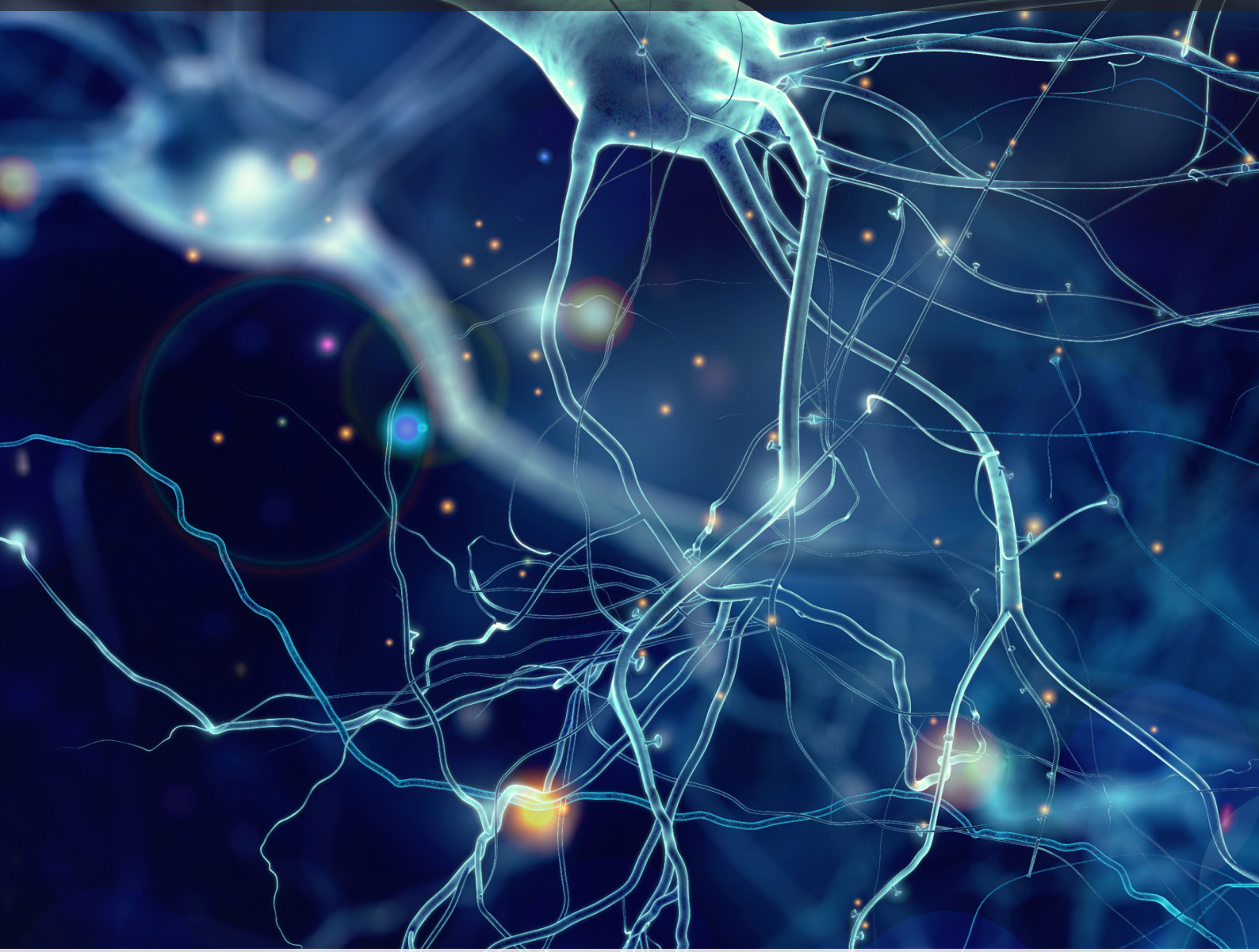
METASTABLE DYNAMICS OF NEURAL ENSEMBLES

EDITED BY: Emili Balaguer-Ballester, Ruben Moreno-Bote, Gustavo Deco and Daniel Durstewitz

PUBLISHED IN: Frontiers in Systems Neuroscience



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88945-437-2
DOI 10.3389/978-2-88945-437-2

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**

METASTABLEDYNAMICS OF NEURAL ENSEMBLES

Topic Editors:
Emili Balaguer-Ballester, Bournemouth University, United Kingdom; Bernstein Center for Computational Neuroscience Heidelberg-Mannheim, Germany
Ruben Moreno-Bote, Universitat Pompeu Fabra, Parc Sanitari Sant Joan de Déu, Spain
Gustavo Deco, Universitat Pompeu Fabra, Spain; Institució Catalana de la Recerca i Estudis Avançats (ICREA), Spain; Max Planck Institute for Human Cognitive and Brain Sciences, Germany; Monash University, Australia
Daniel Durstewitz, Bernstein Center for Computational Neuroscience Heidelberg-Mannheim, Central Institute of Mental Health, and Heidelberg University, Germany

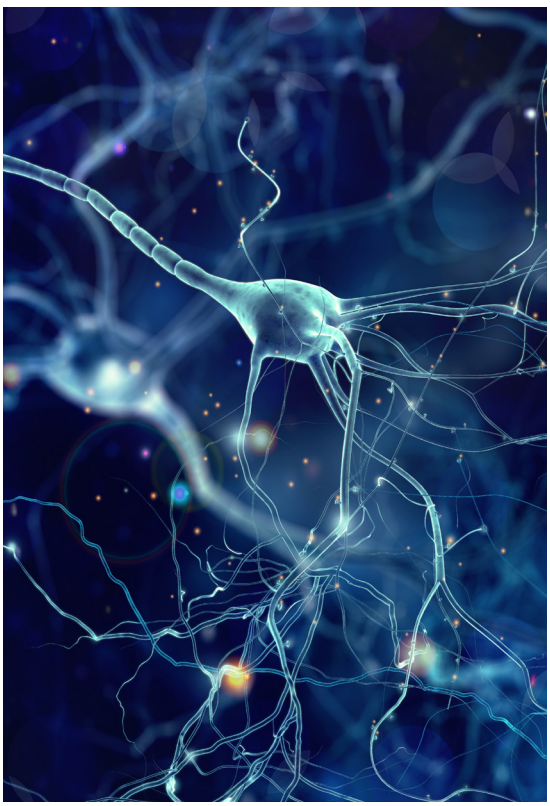


Image: whitehoune/Shutterstock.com

A classical view of neural computation is that it can be characterized in terms of convergence to attractor states or sequential transitions among states in a noisy background. After over three decades, is this still a valid model of how brain dynamics implements cognition?

This book provides a comprehensive collection of recent theoretical and experimental contributions addressing the question of stable versus transient neural population dynamics from complementary angles. These studies showcase recent efforts for designing a framework that encompasses the multiple facets of metastability in neural responses, one of the most fascinating topics currently in systems and computational neuroscience.

Citation: Balaguer-Ballester, E., Moreno-Bote, R., Deco, G., Durstewitz, D., eds. (2018). Metastable Dynamics of Neural Ensembles. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-437-2

Table of Contents

frontiers in Systems Neuroscience

EDITORIAL
published: 26 January 2018
doi: 10.3389/fnsys.2017.00099

Editorial: Metastable Dynamics of Neural Ensembles

Emili Balaguer-Ballester[1,2]*, *Ruben Moreno-Bote*[3], *Gustavo Deco*[4,5,6,7] and *Daniel Durstewitz*[8]

[1] Department of Computing and Informatics, Faculty of Science and Technology, Bournemouth University, Poole, United Kingdom, [2] Bernstein Center for Computational Neuroscience Heidelberg-Mannheim, Mannheim, Germany, [3] Center for Brain and Cognition and Department of Information and Communications Technologies, Pompeu Fabra University, Research Unit, Parc Sanitari Sant Joan de Déu, Barcelona, Spain, [4] Center for Brain and Cognition, Computational Neuroscience Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain, [5] Institució Catalana de la Recerca i Estudis Avançats (ICREA), Barcelona, Spain, [6] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [7] School of Psychological Sciences, Monash University, Melbourne, VIC, Australia, [8] Department of Theoretical Neuroscience, Bernstein Center for Computational Neuroscience Heidelberg-Mannheim, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

Keywords: metastability, trial-to-trial variability, transient dynamics, attractor dynamics, neural noise

Editorial on the Research Topic

Metastable Dynamics of Neural Ensembles

A classical view of neural computation is that it can be characterized in terms of convergence to fixed-point-type attractor states (representing for instance memory patterns in Hopfield, 1982) or limit-cycle-like sequential transitions among states (mapping e.g., motor or syntactical sequences in Elman, 1990). After over three decades, is this still a valid model of how brain dynamics implements cognition? The idea that neuro-computational dynamics is mainly deterministically driven by convergence to emergent stable states in a synaptic/network noisy background has been lively debated, and recently challenged both empirically and by computational work. This question touches on the very basics of our understanding of neural computation; and hence it is one of the most exciting topics currently in systems and computational neuroscience.

This e-book comprises a comprehensive collection of recent theoretical and experimental contributions addressing the question of stable versus transient neural population dynamics, and its implications for the observed variability in neural activity, from diverse, complementary angles.

METASTABILITY IN MODELS

A connecting theme for the multiple contemporary views on metastability in the brain was proposed first by Tognoli and Kelso. In their foundational approach, the authors discuss classical and recent views on how information transfer between brain regions could be accomplished through synchronization and collective neural responses. They frame these ideas in terms of the *coordination dynamics* concept, potentially a key aspect for understanding metastability in neuronal populations.

Metastability and its possible functional role both within and outside of behavioral task contexts is further addressed in four specific modeling approaches situated at different spatial scales, ranging from macro/mesoscopic levels (Schwappach et al.; Stratton and Wiles; Aguilera et al.) to a biophysically detailed level of neuronal systems description (Mazzucato et al.).

The balance between global segregation and integration at a macroscopic scale is theoretically analyzed by Stratton and Wiles. They propose a computational model focused on how the thalamo-cortical loop may underlie long-range segregation between brain regions, producing metastable

OPEN ACCESS

Edited by:
Mikhail Lebedev,
Duke University, United States

Reviewed by:
Hiroshi Okamoto,
RIKEN Brain Science Institute (BSI), Japan

***Correspondence:**
Emili Balaguer-Ballester
eb-ballester@bournemouth.ac.uk

Received: *27 November 2017*
Accepted: *22 December 2017*
Published: *26 January 2018*

Citation:
Balaguer-Ballester E, Moreno-Bote R, Deco G and Durstewitz D (2018) Editorial: Metastable Dynamics of Neural Ensembles. Front. Syst. Neurosci. 11:99. doi: 10.3389/fnsys.2017.00099

responses observed at large spatial scales. Metastability at macroscopic levels could also stem from sensimotor interactions, as suggested by Aguilera et al. These authors designed a new theoretical framework and implemented it in an agent-based model which interacts with the environment. According to this model, metastability arises from the dynamics of sensori-motor feedback interactions, beyond what would be expected from considering brain activity just in isolation.

At mesoscopic scales, neural population models have been constructed that produce metastability through attracting chains of heteroclinic orbits, generating transient dynamics through a sequence of saddle points along which one or several axes are stable (the stable subspace). Following up on this theory, Schwappach et al. demonstrate, using a novel neural field model, how such heteroclinic subspaces can account for part of the observed trial-to-trial variability at the mesoscopic level. Hence, this variability may partly stem from sources other than neuronal or synaptic noise.

At microscopic (biophysical) scales, using a clustered spiking model (in which connectivity patterns are heterogeneous) which exhibits metastable states, Mazzucato et al. show that variation in neuronal ensemble activity may be confined to small subspaces of the whole state space spanned by all the individual units' firing-rates. Moreover, the dimensionality of these subspaces is smaller during stimulus-evoked activity than in the absence of a task. This is in line with empirical studies which report the reduction of neuronal variability upon stimulus presentation (Churchland et al., 2010).

EMPIRICAL STUDIES

Metastability was also addressed in four studies which provide novel analytical tools and empirical evidence. Tošić et al. proposed a new data analysis technique to identify metastability empirically, which was used to infer metastable states in local field potentials evoked by visual stimuli in anesthetised ferrets.

Interestingly, *visual scan paths* (Wilkinson and Metta) reveal complex dynamics which possibly reflects underlying metastable neural activity. In Wilkinson and Metta, the authors proposed a theoretical framework, termed the *singularity hypothesis*, which relies on transient spiral waves which govern persistent neural activity states underlying oculomotor postural control. In general, motor control strategies may be represented in neuronal activity patterns as a complex, distributed spatiotemporal code, which may not be revealed by looking just at neuronal firing rates within recorded ensembles. This is shown in Mao et al. who study behaving rats performing a directional choice task, using a nonlinear decoder to demonstrate how spatiotemporal activity patterns in motor areas increasingly discriminate the animal's choices as learning progresses.

Finally, spatiotemporal patterns generated by synchronized spontaneous activity in the idle brain are analyzed at different spatial scales in two studies (Liu et al.; Yada et al.) which further illustrate the rich diversity in methodological approaches to the empirical identification of metastable and transient dynamics. Specifically, Yada et al. propose that repeating spatiotemporal patterns emerging during synchronized bursts of activity *in vitro* may have their origin in specific sub-populations that become sequentially active in a reproducible temporal order. This result suggests an orchestrated activation of such ensembles that depends on the global state of the network, consistent with spontaneous transitions between metastable states.

At a macroscopic level, Liu et al. develop a new method for identifying co-activity patterns of fMRI responses which is robust to non-stationarity. In this study, a novel type of cluster analysis suggests a richer repertoire of co-activation states beyond the resting-state networks identified previously, and hence perhaps a more specialized functional organization.

In summary, this book provides a comprehensive collection of current modeling and data analysis approaches related to the metastable behavior of cortical ensembles. These studies showcase recent efforts for designing a fundamental framework that encompasses the multiple facets of metastability in neural responses, beyond the original use of the concept in the context of statistical mechanics. To conclude, the last chapter of the book reflects on this plethora of approaches and connects them with the question of the functional role of experimentally observed trial-to-trial variability (Balaguer-Ballester), one of the most intriguing topics currently in systems neuroscience (Moreno-Bote, 2014).

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378. doi: 10.1038/nn.2501

Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1016/0364-0213(90)90002-E

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558. doi: 10.1073/pnas.79.8.2554

Moreno-Bote, R. (2014). Poisson-like spiking in circuits with probabilistic synapses. *PLoS Comput. Biol.* 10:e1003522. doi: 10.1371/journal.pcbi.1003522

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.

frontiers in
SYSTEMS NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 25 June 2014
doi: 10.3389/fnsys.2014.00122

Enlarging the scope: grasping brain complexity

***Emmanuelle Tognoli*[1]* and *J. A. Scott Kelso*[1,2]**

[1] The Human Brain and Behavior Laboratory, Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL, USA
[2] Intelligent System Research Centre, University of Ulster, Derry, Northern Ireland

Edited by:
Emili Balaguer-Ballester, Bournemouth University and Bernstein Center for Computational Neuroscience Heidelberg-Mannheim, UK

Reviewed by:
Nicholas M. Dotson, Montana State University, USA
Karl Friston, University College London, UK

***Correspondence:**
Emmanuelle Tognoli, The Human Brain and Behavior Laboratory, Center for Complex Systems and Brain Sciences, Florida Atlantic University, 777 Glades Road, Boca Raton, FL-33431, USA
e-mail: tognoli@ccs.fau.edu

To further advance our understanding of the brain, new concepts and theories are needed. In particular, the ability of the brain to create information flows must be reconciled with its propensity for synchronization and mass action. The theoretical and empirical framework of Coordination Dynamics, a key aspect of which is metastability, are presented as a starting point to study the interplay of integrative and segregative tendencies that are expressed in space and time during the normal course of brain and behavioral function. Some recent shifts in perspective are emphasized, that may ultimately lead to a better understanding of brain complexity.

Keywords: metastability, brain coordination dynamics, integration, segregation, functional connectivity

THEORIES OF THE BRAIN: A CONCISE HISTORY

How does the brain work? This nagging question is an habitué from the top ten lists of enduring problems in Science's grand challenges. Grasp this paradox: how is one human brain—a chef d'oeuvre of complexity honed by Nature—ever to reach such a feast as to understand itself? Where one brain may fail at this notorious philosophical riddle, may be a strong and diversely-skilled army of brains may come closer. Understanding of the local principles at play has emerged due to the combined efforts of many scientists: neurons talk to their partners by teasing them with charged particles of either excitatory or inhibitory effect, as Nobel laureate Sir John Eccles demonstrated (Eccles, 1966). Targeted release of ions was later shown at sites that seem designed for the exchange of information: typically the axonal termination of the emitting neuron facing the dendrites of a receiving partner (Hodgkin and Huxley, 1952). Many of those two-some neural interlocutors build into a reticulum with remarkable emergent properties. A booming network science followed, generalizing microscale principles on a large-scale. David Rumelhart and James McClelland, Stephen Grossberg and many others pursued this connectionist endeavor (Grossberg, 1976; Rumelhart and McClelland, 1986). Putting function first, they aimed to model specific aspects of human cognition and behavior such as visual perception or language. Yet others, such as Olaf Sporns have devoted much effort to the neurobiological fidelity of their inquiries, conceiving behavior as an emergent phenomenon from an appropriate connectional design (Sporns et al., 2004). The latter may be probed both with theoretical connectivity models where brain complexity is carefully thought through (Sporns et al., 2000) along with empirically-derived models that borrow their connectional blueprints from images of "real" brains (Honey et al., 2010). Though brief, this historical overture suggests that the brain has demonstrable mechanisms for both point-to-point communication of information—most obviously at the microscale—as well as emergent phenomena arising from network properties at the meso- and macroscale.

NEURONAL RELAYS AND THE PROPAGATION OF INFORMATION

The principle of synaptic transmission proved to be picture-perfect for a theory of communication, boosted by the influential work published in 1948 by Claude Shannon (Shannon, 1948). Transfer of information became a principal tenet of brain function, and theories went so far as to conceive of "centers" (Charcot, 1878; see perspectives in Catani and ffytche, 2005) as final destinations for information to be communicated (the concept has now retreated, although it remains perniciously present in neuroscientists' conceptions of brain hierarchies[1] (e.g., Meunier et al., 2010); an alternate view is that it is the journey,

[1]Clearly, the brain is not a diffuse web of randomly coupled dynamical systems. It is a highly structured organ that has been crafted by evolution. The outcome of phylogenesis, a slow timescale pattern-forming process (Bressler and Tognoli, 2006) is the progressive emergence of functional specialization, where some brain regions take on a more integrated role than others, a phenomenon that has been named hierarchical organization (see Robert, 1999; Zilles et al., 2002 for anatomo-functional patterning, see also Honey et al., 2007 for a related network perspective). The notion of hierarchy is popular in neuroscience and may be relevant (see also Kelso and Tuller, 1981 for a discussion of heterarchical and coalitional forms of brain organization): hierarchical systems induce a distinction between ascending and descending connections that could contribute to symmetry breaking which may then lead to directed coupling and information transfer (Tognoli and Kelso, 2014).

but not an elusive final destination, that really matters). This theory of information processing in the brain raises a question that may not have received enough attention: can we readily transpose findings from the smallest synaptic level—findings that speak of only a pair of neurons—to larger spatial scales such as neural areas or the whole brain? In all justice, countless emergent phenomena were discovered through this extrapolation, both in empirical and theoretical work (e.g., Sperry, 1969; Hopfield, 1982; McLeod et al., 1998). But it remains an uneasy feeling that so much of Brain Science is built upon the foundation of a pair of neurons, outside the context of their networks, and with two open-ended areas of darkness at either of their extremities that must be thought of as the entire remainder of the organism's brain (and body).[2]

COLLECTIVE POWER OF NEURONAL SYNCHRONY

We will come back to information transmission later, but let us now explore the matter of spatial scales. As humans tend to agree, increased size makes up for smarter brains (disclosure: both authors are human), and those bigger brains have room to organize themselves at multiple levels, coalescing into functional ensembles at several steps along the way up from neurons to functional areas and to the entire brain (Kelso, 1995; Bressler and Tognoli, 2006; Buzsáki, 2007; Kelso et al., 2013; Tognoli and Kelso, 2013). At larger and more integrated levels of description, other ordering phenomena were discovered that brain scientists conceive in terms of information exchange. In the late 1980's, two groups of scientists, one with Reinhard Eckhorn (Eckhorn et al., 1988) and another with Charles Gray and Wolf Singer (Gray and Singer, 1989), discovered that perceptual integration (or *Gestalt*) elicited transiently synchronous action potentials amongst neurons that had shared-stakes in the sensory object being viewed. Those neurons dealt with separate parts of the visual field, and they generally disagreed on when to elicit their action potentials in the regular course of their participation in visual function. Somehow however, through the complex labyrinth of the visual cortex and despite the fact that some finite amount of time was required to get from any one to any other of them (delays and frustrations manifested in their usual asynchrony), they managed to coincide when they responded to the same object. What we knew from those neurons is that they "responded" strongly to orientation, fragments of contours with sharp luminance gradients. Their synchrony it seems, was a trace of their joint participation in the construction of something bigger (the object) than what each of them was about (pieces of contour). These discoveries resonated with earlier theorizing regarding the organized behavior of neurons such as Donald Hebb's cell assemblies (Hebb, 1949) or Walter Freeman's mass action (Freeman, 1975). The findings by Eckhorn, Singer and Gray launched a relentless quest for synchrony in all parts of the brain and for numerous functions (von der Malsburg et al., 2010), and took the form of several variants (the most basic being coincidence of action potentials and phase-locking of neural oscillations).

IRRECONCILABLES

Theories and dedicated experimental paradigms were built upon both discoveries of synaptic transmission and neural synchronization. And from each side, supporting evidence abounded (see previously mentioned references for evidence and reviews). In spite of their prominence and ubiquity though, the theories carefully avoided confrontation with each other, remaining mostly in the separate territories of distinct research groups.[3] One may note already some difficulties in reconciling them. Let us follow the two extreme views: perfect synchronization and perfect transfer. If all neurons were completely synchronized, they would remain in a changeless state of simultaneity. It is unclear how this system could have flows of information from one place to another. On the other end, if each neuron relayed information in a strict sense, the system would lack basic simultaneity through which synchronous phenomena could emerge. In their radical form it seems, the theories of information exchange *qua* synaptic transfer or neural synchrony are mutually exclusive.

CAN WE FIND DIRECTIONS IN THE BRAIN?

The tension is also visible in some empirical facts. Although directed flows of information in the Shannonian spirit do most certainly occur in neural networks, it is indeed quite challenging to track information otherwise than in local or statistical sense (by tracking, we mean to follow the path of information on a brain map as one would follow any object in motion on a symbolic representation of its spatial domain—see **Figure 1**). The

[2]Although much of neuroscience and its methodology focus on communication between pairs of neurons at the synaptic level, there is a long tradition of trying to understand the brain in terms of self-organization and dynamic coordination. For instance, Ross Ashby theorized on regulatory brain systems, considered under an equilibrium (homeostatic) perspective (Ashby, 1962) and he contributed to developing Cybernetics ("steermanship" born from control theory). From a more physical perspective, Arthur Iberall's Homeokinetics (formulated as Homeodynamics for biology) stressed the persistent, marginally stable nature of open thermodynamic systems which necessarily organize themselves as cyclic, physical action modes at all scales (see Yates, 2008, for review; also Eigen and Schuster, 1979 for a more formal but related approach). More recent years have seen the emergence of Synergetics as a foundational mathematical framework for nonequilibrium self-organization in physics, chemistry and biology (Haken, 1983). Synergetic self-organization is one of the cornerstones of Coordination Dynamics which stresses adaptive, informationally meaningful, bidirectional couplings on multiple levels (Kelso, 1995; Kelso and Haken, 1995; Kostrubiec et al., 2012; Kelso et al., 2013). A key notion of Synergetics that overcomes simple directional interactions between pairs of neurons is so-called circular or reciprocal causality (see discussions in Kelso, 1995; Kelso and Engstrøm, 2006; Tognoli and Kelso, 2013). In a multiscale perspective, circular causality manifests itself in terms of the micro and macroscopic states of a system reciprocally affecting each other; such that the macroscopic (ensemble) behavior enslaves the microscopic states, the ensemble average of which in turn produces the macroscopic states. It is possible that the formalism called for in this essay will rest upon outstanding developments along these lines. On the other hand, since "macro" and "micro" are relative terms in neurobiological dynamical systems where the timescales are not well-separated, a more systematic level by level approach may be required (see Kelso et al., 2013).

[3]There are few exceptions, for instance Fries' (2005) theory of communication through coherence that uses the phase of macroscopic patterns of coherence as a scaffold for selective modulation of information transfer at the microscopic level.

FIGURE 1 | A teasing figure aimed at marking difference between directionality in well-formed Shannonian systems (as in this imaginary brain map) and in complex systems such as the brain. In the latter, direction is less intuitive past immediate spatial and temporal neighborhoods, and it can reverse across spatial scales of observation (Tognoli and Kelso, 2013). The question is highly relevant though, when one is concerned with where and how to effect changes in the system.

brain network after all, is a web, as Francisco Varela et al. (2001) emphasized, and one gets quickly lost with all the branchings, loops and loops within loops (Edelman, 1993; Gallos et al., 2012); structural features that "distribute" information (albeit unlike a postmaster distributes mail). So it seems that transmission principles do not scale well upward from simple "channels" of synaptic interactions to the larger and more complex web of evolved brains. Thus, it is without surprise that the brain betrays an essential communicational etiquette: its parts do not behave in a sequential—one-talks-at-a-time—manner (as opposed to the humoristic illustration of **Figure 1**). It is also overwhelmingly clear that "inputs" from the environment do not enter a silent system. Brain parts constantly exchange information about their current and past affairs, and what comes in at a given time works more as a "perturbation" to an already established ballet, an event that weaves itself within a broader scheme of coordinated brain behavior rather than the sole commander of all things present (Kelso, 1995). All of these nuances differentiate the brain from a channel in which information is transferred from sender to receiver. This situation creates mounting complications. The quest for directional flows in the brain has proved difficult both conceptually and methodologically, yet, it has not deterred efforts toward understanding. Mathematical and empirical studies aimed at resolving these questions are an active area pursued by many, including our own colleagues (Bressler et al., 2007; Bressler and Seth, 2011).

BRAIN ORGANIZATION: SYNCHRONIZATION OR COORDINATION?

The second concept, synchrony, also bears its share of ambiguities. The firm ground on which we stand is that the timing of neural activity is not left to hazard (as if parts of the brain behaved independently, and were totally oblivious of what the others were doing). "*When*" one brain part behaves influences *when* others do. And like social creatures, neurons also use the power of their numbers to increase their impact, creating collective structures that speak from a common voice. A generic name for such behavior is *coordination* (Kelso, 1995). Synchrony is a narrower concept, one of several ways for a system to coordinate itself. Though synchrony has multiple meanings (and though its study uses a variety of tools across the board), it is easy to conceive and to model, perhaps explaining its systematic resort. To be rigorous however, synchrony requires two important and inter-related characteristics: first, that the underlying temporal

order in which the system is embedded be frequency-locked, and second, that attractors have emerged in the system's coordination dynamics (attractors are mathematical structures that entrap the system's coordination dynamics into persistent—hard to break—states). How to examine if there are attractors in the system from empirical data? We do know how to go from theoretical descriptions of dynamics with- and without-attractors to their phenomenology (using models to create data at will), but we are not very successful at taking the return path: demonstrating attractors or lack thereof from looking at data, at least for complicated systems like the brain. The other criterion, frequency-locking, is a little bit easier. And what the data say is that brain parts exhibit tendencies toward frequency-locking without going all the way to being perfectly identical. So it seems, we could be dealing with synchronous *tendencies* rather than synchrony (Kelso, 2008; Tognoli and Kelso, 2014). The difference may seem subtle to some (and some might be tempted to brush it aside as noise or measurement uncertainty), but mathematically and conceptually, it is enormous: it speaks of two entirely different species of dynamical systems, as said before, one with and the other without attractors. We are turning the spotlight to this distinction because we believe it to be paramount for progress in understanding the brain.

The mathematical formalism of "generalized synchronization" (Rulkov et al., 1995; see also Barreto et al., 2000) is apt to quantify such regimes of coordination sans attractors. Generalized synchronization assumes the existence of a functional mapping between two dynamical processes, without imposing the strong constraint of identical synchronization usually sought for in classical measures of coherence and correlation. The notion of generalized synchrony is set to grow in importance in the context of a coordination dynamic between weakly coupled components—a beneficial situation that enhances system complexity. The absence of complexity is a nefarious situation that André Gide related to tyranny.

AT THE CROSSROADS OF PROPAGATION AND SYNCHRONIZATION

We hope that the previous exposé motivated the thought that neural networks neither operate on perfect synchrony nor on strict transfer, which is good news as each prevents expression of important features of the other. What then is the link between them? Some attempts at studying synchrony and transfer in a common formalism have emerged, exceptions to their usual avoidance of one another. One is the quest for quantifying directional coupling as discussed above. Another attempt is functional (relatedly, effective) connectivity,[4] a daring concept that Karl Friston created on his way to developing theoretical and computational tools for the analysis of functional images of the brain (Friston, 1994). Connectivity deals with ways for information to go from one place to another. Ideally, we would be able to measure the connection (the "traffic" between two sites) independently from the state of those sites where said traffic imparts effect (as one would measure how many cars travel on the road between two cities). If independently measured, large scale connectivity and local activity would be amenable to reveal their effect on each other. Since we do not have adequate tools to measure the flow of information in living fiber tracts at large though, connectivity is not measured directly; rather it is inferred from the way brain components behave. Interaction, it is postulated, has to leave detectable traces in the behavior of its participants. Of course, contemporaneous theories have shaped the lens through which scientists have tried to see this influence. To make things practical, the assumption was often made that regions exchanging information must be correlated or synchronized (connectivity→correlation). Flipping things around for the operational goal of quantifying the unquantifiable information flow, "how much regions were correlated" became the proxy for how much they exchanged information (correlation→connectivity). But with only this concept of synchronization under the scope, we may see a mere fraction of the brain at work, the tip of the iceberg. What if most coordinated behaviors in the brain do not fall under our definition of correlation or synchrony? Depending on the methods used, that would mean for instance brain regions that are coordinated yet not temporally coincident; or assemblies in which self-organization favors a fluid coordination regime sans-attractor (such as metastable tendencies, to be discussed below) over rigid states of phase-locking. Can we see dynamics in which no absolute "order" emerges in space (synchronization) or in time (transfer), and still make sense of it as a means for the brain to function? Those are the dark and uncharted areas in the spatiotemporal organization of complex systems—those for which we sorely lack concepts and methods (**Figure 2**).

ENLARGING THE SCOPE: METASTABILITY

The set of questions above resonates with a recent shift in perspective on brain function, from a primary focus on neural synchronization to the broader—and deeper—problem of dynamic coordination. This shift was salient in the editorial introduction to a special issue of Nature Review Neuroscience in February 2010, where the word "coordination" occurred six times in a short text of 250 words (From The Editors, 2010). And this is a concept that is growing (Kelso, 1995, 2012; Bressler and Kelso, 2001; Kelso and Tognoli, 2007; Werner, 2007; Rabinovich et al., 2008; Tognoli and Kelso, 2009, 2013, 2014; von der Malsburg et al., 2010; Farmer, 2011; Kelso et al., 2013). Coordination includes synchronization as one possible collective behavior, but it also considers many other ways for components of the brain to interact. In particular, under certain conditions partially synchronized behaviors arise. In them, the parts exhibit simultaneous tendencies to temporarily couple and to segregate as independent entities. Such metastable regimes, we and others have shown, constitute a recipe for complexity (Kelso, 1995, 2012; Friston, 1997; Freeman and Holmes, 2005; Kelso and Tognoli, 2007; Werner, 2007; Rabinovich et al., 2008; Tognoli and Kelso, 2009, 2013, 2014; Bhowmik and Shanahan,

[4]Karl Friston (1994) intended functional connectivity to expose temporal correlations between neurophysiological events—a descriptive approach without *a priori* specification that the originating structures would affect one another—whereas effective connectivity is meant to address explicit influences between brain parts (how much does A affect B, an estimation based on *a priori* model of those areas' relations, later reframed as causality).

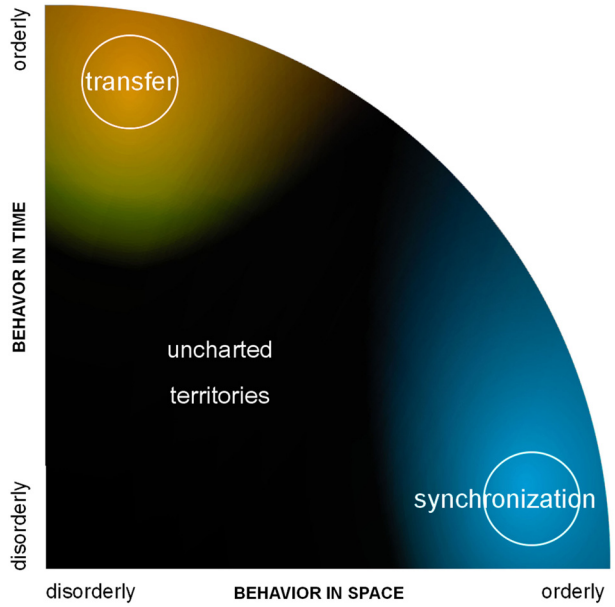


FIGURE 2 | A graph of spatiotemporal organization. The horizontal axis quantifies the degree of order in space; order in time is on the vertical axis. Examples of orderly phenomena are propagation and synchronization. There are numerous hints that the brain is more efficient with less than complete order in space and time. The dark area represents this region with incomplete spatial and temporal order, for which concepts and tools are sorely missing.

2013; Kelso et al., 2013). Why is metastable coordination dynamics of high interest to understanding how the brain works? First, it typically arises when the parts are no longer perfect clones of one another (e.g., as in computational models built from collections of identical neurons). When symmetry is broken and interacting parts are recognized in the diversity of their intrinsic behavior, a more realistic picture of brain function emerges (indeed a trend toward studying more diverse associations in the brain may explain the shift in perspective, for instance the interactions between neurons and astrocytes, see, e.g., Wade et al., 2013). Second, incomplete synchronization is more adaptive than pure forms. A fundamental nonlinearity in brain self-organization exists. Too much autonomy (parts of the brain hardly ever affected by what others are doing) prevents emergence, integration and mass action (Uhlhaas and Singer, 2006). Yet, too much integration (for instance the whole brain engaged in a giant common behavior, Cruikshank and Connors, 2008) is inadequate too, because the respective parts can no longer do what they are supposed to do in contributing to collective behavior. The parts then have no choice but to behave exactly like each other and the richness of their individual dispositions is lost to the ensemble. It is enough to note—as many have—that excess synchronization is pathological in the brain, for instance in epilepsy or Parkinson's disease (Uhlhaas and Singer, 2006; Hammond et al., 2007; Lehnertz et al., 2009). As a result, the ideal place for a brain to exhibit a rich set of meaningful behaviors is in-between integration and segregation. This is where the "incomplete" synchronization tendencies or metastable coordination comes into play (Tononi et al., 1994; Kelso, 1995, 2012; Friston, 1997; Werner, 2007; Chialvo, 2010; Tognoli and Kelso, 2013, 2014). Elsewhere, we have also speculated on the tremendous functional advantages that metastability would confer to a system, including speed, flexibility and resilience (Kelso and Tognoli, 2007; Kelso, 2012).

CREATING A NEW CONCEPTUAL FRAMEWORK

The gap between our current understanding of the brain and the miracles of our mental life and behavioral achievements (for example, consciousness and capacity for invention) remains abysmal. Looking through the history of science, several paradigms of brain function flourished and then dried up following the ebbs and flows of scientific metaphors. The ultimate model, the one that allows to forecast all matters of brain action and to design an artificial counterpart of multiple functional prowesses, remains out of sight. Two lines of thinking have been much explored in recent times: information transfer and synchronization. Their success owes much to the fact that they are special cases and open to quantification. When examined together though, they reveal some incompatibilities that seem to require a relaxation of both principles: less stringent temporal order and less complete spatial order. To advance our understanding of the brain, Neuroscience must open up avenues to study functional behavior in a broader sense. We face two alternatives: to leave it all within the current framework, with the approximate truth derived from current theories (the brain "sort of transfers information", and it operates with "near synchrony"), or to face the issue head on with a different theoretical mindset. In the latter case, a new phenomenology is up for grasp. It will be difficult to conceptualize, and even more so to observe, since it points toward a void in understanding. Modeling approaches can lead the way, by informing which observables we can expect to encounter in the coordinating living brain. And tools will have to be revised or built to adapt to this new world, tools that will say for instance, when "more synchrony" is "too much synchrony" (astonishingly, this simple question is not built into our current enquiries, despite obvious evidence of the ills of excess synchrony). We note that Brain Science is reaching a turning point that may make this renewal possible: it shows many signs of its readiness to enlarge the scope on brain function, not least of which is a recent outburst of interest in segregation phenomena (Poulet and Petersen, 2008; Ecker et al., 2010; Renart et al., 2010). A new paradigm would help to integrate principles that seem contradictory in their radical form: transfer and synchronization, as well as integration and segregation. Those pairs of concepts are reconciled under the dynamical regime of metastability (Kelso, 1995; Kelso and Engstrøm, 2006; Kelso and Tognoli, 2007; Tognoli and Kelso, 2014). Parts making-up the brain (for instance molecules, neurons, or brain areas) have simultaneous tendencies for independence and cooperation. As a result, they engage in the double-duty of lending their help to the collective behavior undertaken by several brain parts, and of performing their own independent behavior. Under a metastable regime, information is continuously created,

preserved and annihilated by spatiotemporally changing coalitions among parts and processes. This is a source of dynamic complexity, and the likely origin of the human brain's many prowesses.

ACKNOWLEDGMENTS

This work was supported by NIMH (MH080838), NSF (BCS0826897), the US ONR (N00014-09-1-0527), the Davimos Family Endowment for Excellence in Science and the Chaire d'excellence Pierre de Fermat. Suggestions by both reviewers are gratefully acknowledged.

REFERENCES

(2010) From The Editors. *Nature Reviews Neuroscience* 11:71. doi: 10.1038/nrn2802

Ashby, W. R. (1962). "Principles of the self-organizing system," in *Principles of Self-Organization: Transactions of the University of Illinois Symposium*, eds H. Von Foerster and G. W. Zopf, Jr. (London, UK: Pergamon Press), 255–278.

Barreto, E., So, P., Gluckman, B. J., and Schiff, S. J. (2000). From generalized synchrony to topological decoherence: emergent sets in coupled chaotic systems. *Phys. Rev. Lett.* 84, 1689–1692. doi: 10.1103/physrevlett.84.1689

Bhowmik, D., and Shanahan, M. (2013). Metastability and inter-band frequency modulation in networks of oscillating spiking neuron populations. *PLoS One* 8:e62234. doi: 10.1371/journal.pone.0062234

Bressler, S. L., and Kelso, J. A. S. (2001). Cortical coordination dynamics and cognition. *Trends Cogn. Sci.* 5, 26–36. doi: 10.1016/s1364-6613(00)01564-3

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. doi: 10.1002/sim.2935

Bressler, S. L., and Seth, A. K. (2011). Wiener–granger causality: a well established methodology. *Neuroimage* 58, 323–329. doi: 10.1016/j.neuroimage.2010.02.059

Bressler, S. L., and Tognoli, E. (2006). Operational principles of neurocognitive networks. *Int. J. Psychophysiol.* 60, 139–148. doi: 10.1016/j.ijpsycho.2005.12.008

Buzsáki, G. (2007). The structure of consciousness. *Nature* 446:267. doi: 10.1038/446267a

Catani, M., and ffytche, D. H. (2005). The rises and falls of disconnection syndromes. *Brain* 128, 2224–2239. doi: 10.1093/brain/awh622

Charcot, J. M. (1878). *Leçons sur les Localisations dans les Maladies du Cerveau.* Paris: Delahaye et Co.

Chialvo, D. R. (2010). Emergent complex neural dynamics. *Nat. Phys.* 6, 744–750. doi: 10.1038/nphys1803

Cruikshank, S. J., and Connors, B. W. (2008). State-sanctioned synchrony. *Nature* 454, 839–840. doi: 10.1038/454839a

Eccles, J. C. (1966). The ionic mechanisms of excitatory and inhibitory synaptic action. *Ann. N Y Acad. Sci.* 137, 473–494. doi: 10.1111/j.1749-6632.1966.tb50176.x

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. doi: 10.1126/science.1179867

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., et al. (1988). Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. *Biol. Cybern.* 60, 121–130. doi: 10.1007/bf00202899

Edelman, G. M. (1993). Neural Darwinism: selection and reentrant signaling in higher brain function. *Neuron* 10, 115–125. doi: 10.1016/0896-6273(93)90304-a

Eigen, M., and Schuster, P. (1979). *The Hypercycle: A Principle of Natural Self-organisation.* Heidelberg: Springer-Verlag.

Farmer, S. (2011). Review of dynamic coordination in the brain: from neurons to mind. *J. Psychophysiol.* 25, 201–202. doi: 10.1027/0269-8803/a000055

Freeman, W. J. (1975). *Mass Action in the Nervous System.* New York: Academic Press.

Freeman, W. J., and Holmes, M. D. (2005). Metastability, instability, and state transition in neocortex. *Neural Netw.* 18, 497–504. doi: 10.1016/j.neunet.2005.06.014

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapp.* 2, 56–78. doi: 10.1002/hbm.460020107

Friston, K. J. (1997). Transients, metastability and neuronal dynamics. *Neuroimage* 5, 164–171. doi: 10.1006/nimg.1997.0259

Gallos, L. K., Sigman, M., and Makse, H. A. (2012). The conundrum of functional brain networks: small-world efficiency or fractal modularity. *Front. Physiol.* 3:123. doi: 10.3389/fphys.2012.00123

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. doi: 10.1073/pnas.86.5.1698

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biol. Cybern.* 23, 121–134. doi: 10.1007/bf00344744

Haken, H. (1983). *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology.* New York: Springer-Verlag.

Hammond, C., Bergman, H., and Brown, P. (2007). Pathological synchronization in Parkinson's disease: networks, models and treatments. *Trends Neurosci.* 30, 357–364. doi: 10.1016/j.tins.2007.05.004

Hebb, D. O. (1949). *The Organization of Behavior.* New-York: John Wiley.

Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544.

Honey, C. J., Kötter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. U S A* 104, 10240–10245. doi: 10.1073/pnas.0701519104

Honey, C. J., Thivierge, J. P., and Sporns, O. (2010). Can structure predict function in the human brain? *Neuroimage* 52, 766–776. doi: 10.1016/j.neuroimage.2010.01.071

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U S A* 79, 2554–2558. doi: 10.1073/pnas.79.8.2554

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior.* Cambridge, Massachusetts: The MIT Press.

Kelso, J. A. S. (2008). An essay on understanding the mind: the A.S. Iberall Lecture. *Ecol. Psychol.* 20, 180–208. doi: 10.1080/10407410801949297

Kelso, J. A. S. (2012). Multistability and metastability: understanding dynamic coordination in the brain. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 906–918. doi: 10.1098/rstb.2011.0351

Kelso, J. A. S., and Engstrøm, D. A. (2006). *The Complementary Nature.* Cambridge, MA: The MIT Press.

Kelso, J. A. S., Dumas, G., and Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Netw.* 37, 120–131. doi: 10.1016/j.neunet.2012.09.003

Kelso, J. A. S., and Haken, H. (1995). "New laws to be expected in the organism: synergetics of brain and behavior," in *What is Life? The Next Fifty Years: Speculations on the Future of Biology*, eds M. P. Murphy and L. A. J. O'Neill (Cambridge, UK: Cambridge University Press), 137–160.

Kelso, J. A. S., and Tognoli, E. (2007). "Toward a complementary neuroscience: metastable coordination dynamics of the brain," in *Neurodynamics of Higher-level Cognition and Consciousness*, eds R. Kozma and L. Perlovsky (Heidelberg: Springer), 39–59.

Kelso, J. A. S., and Tuller, B. (1981). Toward a theory of apractic syndromes. *Brain Lang.* 12, 224–245. doi: 10.1016/0093-934x(81)90016-x

Kostrubiec, V., Zanone, P.-G., Fuchs, A., and Kelso, J. A. S. (2012). Beyond the blank slate: routes to learning new coordination patterns depend on the intrinsic dynamics of the learner —experimental evidence and theoretical model. *Front. Hum. Neurosci.* 6:222. doi: 10.3389/fnhum.2012.00222

Lehnertz, K., Bialonski, S., Horstmann, M. T., Krug, D., Rothkegel, A., Staniek, M., et al. (2009). Synchronization phenomena in human epileptic brain networks. *J. Neurosci. Methods* 183, 42–48. doi: 10.1016/j.jneumeth.2009.05.015

McLeod, P., Plunkett, K., and Rolls, E. T. (1998). *Introduction to Connectionist Modelling of Cognitive Processes.* Oxford, UK, New-York, NY: Oxford University Press.

Meunier, D., Lambiotte, R., and Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* 4:200. doi: 10.3389/fnins.2010.00200

Poulet, J. F. A., and Petersen, C. C. H. (2008). Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* 454, 881–885. doi: 10.1038/nature07150

Rabinovich, M. I., Huerta, R., Varona, P., and Afraimovich, V. S. (2008). Transient cognitive dynamics, metastability and decision making. *PLoS Comput. Biol.* 4:e1000072. doi: 10.1371/journal.pcbi.1000072

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. doi: 10.1126/science.1179850

Robert, A. (1999). *Lamination and Within-Area Integration in the Neocortex.* San Diego: University of California, Doctoral dissertation.

Rulkov, N. F., Sushchik, M. M., Tsimring, L. S., and Abarbanel, H. D. (1995). Generalized synchronization of chaos in directionally coupled chaotic systems. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 51, 980–994. doi: 10.1103/physreve.51.980

Rumelhart, D. E., McClelland, J. L., and The PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations.* Cambridge, MA: MIT Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 623–656. doi: 10.1002/j.1538-7305.1948.tb00917.x

Sperry, R. W. (1969). A modified concept of consciousness. *Psychol. Rev.* 76, 532–536. doi: 10.1037/h0028156

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cogn. Sci.* 8, 418–425. doi: 10.1016/j.tics.2004.07.008

Sporns, O., Tononi, G., and Edelman, G. M. (2000). Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cereb. Cortex* 10, 127–141. doi: 10.1093/cercor/10.2.127

Tognoli, E., and Kelso, J. A. S. (2009). Brain coordination dynamics: true and false faces of phase synchrony and metastability. *Prog. Neurobiol.* 87, 31–40. doi: 10.1016/j.pneurobio.2008.09.014

Tognoli, E., and Kelso, J. A. S. (2013). "On the brain's dynamical complexity: coupling and causal influences across spatiotemporal scales," in *Advances in Cognitive Neurodynamics (III)*, ed Y. Yamaguchi (Netherlands: Springer), 259–265.

Tognoli, E., and Kelso, J. A. S. (2014). The metastable brain. *Neuron* 81, 35–48. doi: 10.1016/j.neuron.2013.12.022

Tononi, G., Sporns, O., and Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. U S A* 91, 5033–5037. doi: 10.1073/pnas.91.11.5033

Uhlhaas, P. J., and Singer, W. (2006). Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron* 52, 155–168. doi: 10.1016/j.neuron.2006.09.020

Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550

von der Malsburg, C., Phillips, W. A., and Singer, W. (Eds). (2010). *Dynamic Coordination in the Brain: From Neurons to Mind, Strüngmann Forum Report, Vol. 5.* Cambridge, MA: MIT Press.

Wade, J. J., McDaid, L. J., Harkin, J., Crunelli, V., and Kelso, J. A. S. (2013). Biophysically-based computational models of astrocyte ∼ neuron coupling and their functional significance. *Front. Comput. Neurosci.* 7:44. doi: 10.3389/fncom.2013.00044

Werner, G. (2007). Metastability, criticality and phase transitions in brain and its models. *Biosystems* 90, 496–508. doi: 10.1016/j.biosystems.2006.12.001

Yates, E. (2008). Homeokinetics/Homeodynamics: a physical heuristic for life and complexity. *Ecol. Psychol.* 20, 148–179. doi: 10.1080/10407410801977546

Zilles, K., Palomero-Gallagher, N., Grefkes, C., Scheperjans, F., Boy, C., Amunts, K., et al. (2002). Architectonics of the human cerebral cortex and transmitter receptor fingerprints: reconciling functional neuroanatomy and neurochemistry. *Eur. Neuropsychopharmacol.* 12, 587–599. doi: 10.1016/s0924-977x(02)00108-6

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: 20 March 2014; accepted: 09 June 2014; published online: 25 June 2014.
Citation: Tognoli E and Kelso JAS (2014) Enlarging the scope: grasping brain complexity. Front. Syst. Neurosci. ***8****:122. doi: 10.3389/fnsys.2014.00122*
This article was submitted to the journal Frontiers in Systems Neuroscience.

frontiers in Systems Neuroscience

ORIGINAL RESEARCH
published: 25 August 2015
doi: 10.3389/fnsys.2015.00119

Global segregation of cortical activity and metastable dynamics

Peter Stratton[1,2]* *and Janet Wiles*[3]

[1] Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia, [2] Centre for Clinical Research, The University of Queensland, Brisbane, QLD, Australia, [3] School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia

Cortical activity exhibits persistent metastable dynamics. Assemblies of neurons transiently couple (integrate) and decouple (segregate) at multiple spatiotemporal scales; both integration and segregation are required to support metastability. Integration of distant brain regions can be achieved through long range excitatory projections, but the mechanism supporting long range segregation is not clear. We argue that the thalamocortical matrix connections, which project diffusely from the thalamus to the cortex and have long been thought to support cortical gain control, play an equally-important role in cortical segregation. We present a computational model of the diffuse thalamocortical loop, called the competitive cross-coupling (CXC) spiking network. Simulations of the model show how different levels of tonic input from the brainstem to the thalamus could control dynamical complexity in the cortex, directing transitions between sleep, wakefulness and high attention or vigilance. The model also explains how mutually-exclusive activity could arise across large portions of the cortex, such as between the default-mode and task-positive networks. It is robust to noise but does not require noise to autonomously generate metastability. We conclude that the long range segregation observed in brain activity and required for global metastable dynamics could be provided by the thalamocortical matrix, and is strongly modulated by brainstem input to the thalamus.

Keywords: thalamocortical matrix, autonomous metastable dynamics, cortical segregation, default mode network, sleep, wakefulness, complexity, spiking networks

OPEN ACCESS

Edited by:
Emili Balaguer-Ballester, Bournemouth University, UK; Bernstein Center for Computational Neuroscience Heidelberg/Mannheim, Germany

Reviewed by:
Joachim Hass, Central Institute for Mental Health, Germany
Loreen Hertäg, Bernstein Center for Computational Neuroscience Heidelberg-Mannheim, Germany

***Correspondence:**
Peter Stratton, Queensland Brain Institute, The University of Queensland, QBI Building (#79), St Lucia, QLD 4072, Australia
p.stratton1@uq.edu.au

Received: *20 March 2015*
Accepted: *07 August 2015*
Published: *25 August 2015*

Citation:
Stratton P and Wiles J (2015) Global segregation of cortical activity and metastable dynamics. Front. Syst. Neurosci. 9:119. doi: 10.3389/fnsys.2015.00119

Introduction

In wakefulness and in rapid eye movement (REM) sleep, cortical activity exhibits persistent ongoing complex dynamics (Breakspear et al., 2003; Honey et al., 2007). In this state, activity shifts continuously throughout the cortex, and cortical regions couple (integrate) and decouple (segregate) across multiple spatial and temporal scales (Sporns et al., 2000a; Varela et al., 2001; Shanahan, 2008; Tognoli and Kelso, 2014). Each individual episode of integration and segregation is transient, but a continuous superposition of such episodes through space and time results in the observed persistent metastability of cortical dynamics (Tognoli and Kelso, 2014). This dynamical complexity is hypothesized to support the brain's flexibility and sophisticated processing capabilities (Breakspear et al., 2003; Buzsáki and Draguhn, 2004; Fries, 2005; Tognoli and Kelso, 2014), including memory retrieval, planning and problem solving (Binder et al., 1999; Mazoyer et al., 2001). During times when the brain is not actively processing sensory stimuli or task-related events, and as such is in a state known as the "resting" or "default-mode" state, cortical activity is

concentrated in a well-defined sub-network including regions of frontal and association cortices (specifically the ventromedial prefrontal cortex, posterior cingulate cortex, ventral precuneus, and parts of the medial temporal and medial, lateral and inferior parietal cortices) (Greicius et al., 2003; Uddin et al., 2008). Activity of this default-mode network (DMN) is anticorrelated with activity in much of the rest of the cortex—that is, activation of the DMN and of those cortical centers used for sensory and task-related processing is largely mutually exclusive (Greicius et al., 2003; Uddin et al., 2008; Tomasi and Volkow, 2011).

Metastable cortical states, and the functions these dynamics presumably underpin (Binder et al., 1999; Mazoyer et al., 2001; Breakspear et al., 2003; Buzsáki and Draguhn, 2004; Fries, 2005; Tognoli and Kelso, 2014), cannot exist without the myriad, often overlooked, sub-cortical areas that provide the cortex with controlling and modulatory input. Projections from the pedunculopontine nucleus (PPN) and the laterodorsal tegmental nuclei (LDT), parts of the brainstem network collectively known as the ascending arousal system (AAS—but previously known as the reticular activating system), enter the intralaminar nuclei of the thalamus (IL) and thence on to the cortex through the thalamocortical matrix connections. The AAS is thought to modulate wake and sleep states as well as arousal and vigilance levels (Moruzzi and Magoun, 1949; Reese et al., 1995; Jones, 2003). The IL matrix connections project diffusely and somewhat non-specifically to large portions of the cortex, which in turn project back to the IL through the thalamic reticular nucleus (RN). Notably, these projections have opposing effects: whereas the cortex excites the RN, the RN exerts an inhibitory influence on the IL. Hence, rising global cortical activity increases RN activation, which in turn inhibits the IL, reducing its input to cortical activation and ultimately countering the activity rise in the cortex. Similarly, a decrease in cortical activity can cause an increase in thalamocortical input from the AAS. This diffuse matrix thalamocortical loop can therefore potentially dynamically control overall activity levels in the cortex (Steriade and McCarley, 1990). Effectively, it implements a mechanism similar to k-winner-take-all (WTA) across the entire cortex, where the allowed activity level k is controlled by AAS input to the IL. WTA networks are known to be able to implement powerful computational functions (Maass, 2000).

Brain integration is served by long range excitatory connections, but the paucity of long range inhibition in the brain has meant that the mechanisms to support long range segregation of brain activity are less obvious and are not well understood. Long range inhibition has been discovered within the visual cortex (McDonald and Burkhalter, 1993), between the hippocampus and entorhinal cortex (Melzer et al., 2012), and between the prefrontal cortex and nucleus accumbens (Lee et al., 2014). However, currently these long range inhibitory connections are known to exist only in or between a limited number of specific structures; they certainly do not approach the abundance of the long range excitatory projections forming the large fiber tracts that criss-cross the brain. The mechanism by which, for example, the DMN is segregated from other cortical regions is unclear (Greicius et al., 2003; Uddin et al., 2008). The contribution of the AAS, IL and RN to global cortical activation control is widely suspected (Steriade and McCarley, 1990; Taylor and Farrukh, 1996; Saper et al., 2001). In this paper we refine this view and argue for the specific functions of long range competition and segregation of cortical activity, and thereby the support of cortical metastability. To demonstrate this potential we present a computational model of the diffuse thalamocortical loop, called the Complex Cross Coupling (CXC) spiking network (see **Figure 1**). The CXC network includes input from the AAS and RN to the IL, local and long range corticocortical connections and local inhibitory interneurons. Simulations demonstrate that the model requires no extrinsic noise to exhibit its full range of dynamical states. We use the model to show how different levels of input from the AAS to the IL could support a range of dynamical states in the cortex, including states with high dynamical complexity.

Methods

Neuron and Synapse Models

All simulations were conducted using the Parallel Circuit Simulator (PCSIM) (Pecevski et al., 2009), a comprehensive software package for the simulation of large neural networks. Simulations used Euler integration with a time-step of 0.1 ms.

The Complex Cross Coupling Spiking Network was constructed using two types of Izhikevich model cortical neurons (regular spiking, RS, and fast spiking, FS) (Izhikevich, 2003). These neurons provide for realistic neuron membrane dynamics such as spike frequency adaptation, intrinsic bursting, resonance and bistability, whilst being computationally tractable for large network simulations (Izhikevich, 2004). The Izhikevich model is defined by three equations over two variables, the membrane potential v (nominally in mV) and the membrane recovery variable q, which are updated as follows:

$$v' = 0.04v^2 + 5v + 140 - q + I \quad (1)$$

$$q' = a(bv - q) \quad (2)$$

$$\text{if } v \geq 30 \text{ then}$$

$$v \leftarrow g$$

$$q \leftarrow q + h \quad (3)$$

where I is the summed synaptic input current and a, b, g, and h are neuron-specific model parameters. In this study, all excitatory neurons were modeled as regular-spiking (RS) cells with the parameters $(a, b, g, h) = (0.02, 0.2, -65, 8)$ and inhibitory neurons as fast spiking (FS) cells with $(a, b, g, h) = (0.1, 0.2, -65, 2)$. These dimensionless parameters are set as specified in Izhikevich (2003) to obtain membrane dynamics (modeled by v) that closely resemble the modeled classes of neuron (i.e., RS and FS cells). In this study, pyramidal cells are modeled as regular spiking, and inhibitory interneurons as fast spiking.

The synapses received presynaptic spikes that would initiate postsynaptic currents that decayed with characteristic time constants. Synapses were conductance-based, meaning they each had a reversal potential E_{rev} at which current flow ceased and beyond which the direction of current flow reversed. Reversal

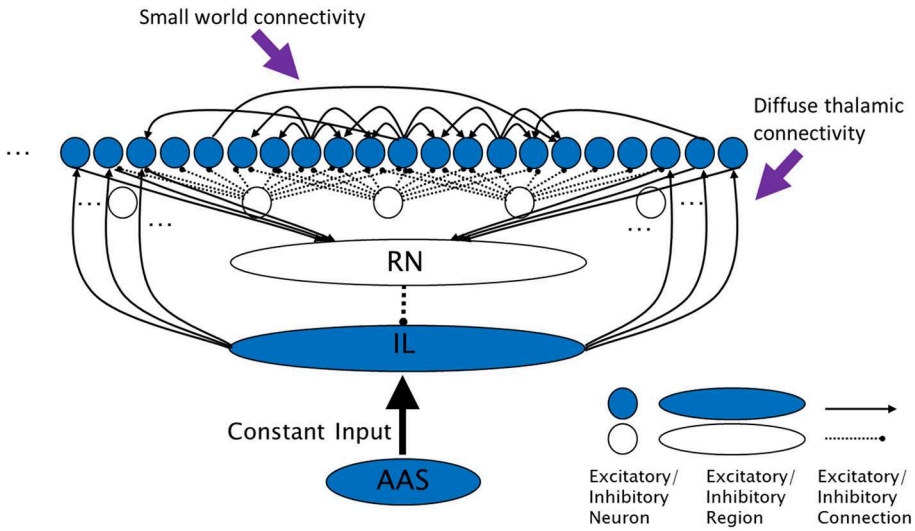


FIGURE 1 | The complex cross coupling (CXC) spiking network (not all connections shown). Filled circles are cortical principal cells (pyramidal neurons). Open circles are inhibitory interneurons. Excitatory cortical pyramidal neurons are innervated by diffuse thalamic (intralaminar nuclei) projections (from shaded oval). The intralaminar nuclei receive continuous steady, not time-varying, input current from the AAS. Cortical neurons project back to the thalamic reticular nucleus which inhibits the intralaminar nuclei, completing the diffuse thalamocortical loop. Cortical neurons excite each other within a small neighborhood radius, inhibit each other through local interneurons within a larger inhibitory radius, and project a small number of random long range excitatory synapses to other cortical neurons. RN, thalamic reticular nucleus; IL, diffuse (intralaminar) thalamic nuclei; AAS, ascending arousal system.

potential for excitatory and inhibitory synapses was 0 mV and −90 mV respectively.

To implement short term synaptic dynamics (which was applied to synapses between all excitatory neurons in one experiment, the results of which are shown in **Figure 5D**), a combination of synaptic depression and facilitation was used (Markram et al., 1998). For synaptic depression, synaptic efficacy was assumed to be a finite resource, of which a proportion p was in use at any given time. Only an amount u of the currently-available proportion of synaptic efficacy, $(1 - p)$, was used at the occurrence of each spike; p recovered back to zero with time constant d (p was therefore bounded in [0,1)). The rate of change of p for synapse i, $\dot{p}_i$, was given by:

$$\dot{p}_i = \delta(t - t_i)(1 - p)_i u_i - \frac{p_i}{d_i} \tag{4}$$

where δ is the Dirac delta function and t_i was the time of the last spike from the neuron that was presynaptic to synapse i. The initial value of u for dynamic synapses was set to U, then for synaptic facilitation u was increased on the occurrence of each spike, recovering back to U with time constant f (u was therefore bounded in $[U,1)$). The rate of change of u for synapse i, $\dot{u}_i$, was given by:

$$\dot{u}_i = \delta(t - t_i)(1 - u_i)U + \frac{(U - u_i)}{f_i} \tag{5}$$

The synaptic usage factor U was set to 0.15 with a recovery time constant of $f = 1000$ ms for all dynamic synapses. Total synaptic current, $I_{dynamic}$, for a postsynaptic neuron was given by the sum of all the currents from its afferent synapses:

$$I_{dynamic} = \sum_i \left[w_i(1 - p_i)u_i(v - E_{rev}) \right] \tag{6}$$

where w_i was the total efficacy of synapse i and v was the postsynaptic membrane potential.

The synaptic recovery time constant d was set to 50 ms for local excitatory-to-excitatory and 500 ms for long-range excitatory-to-excitatory connections. The difference was motivated in part by the different recovery times for AMPA and NMDA receptors. Long range connections in this model were loosely associated with feedback connections within the hierarchically-organized, massively recurrent cortical connectome. These feedback connections may be NMDA-rich, as against feedforward and local connections which may be predominantly AMPA-based (Thiele, 2012). The time constants used were longer than typically acknowledged for these receptor types to compensate for the absence of other network influences in the model, such as cholinergic neuromodulation, which are known to enhance NMDAR function.

For static synapses (synapses of fixed efficacy, i.e., with no short term synaptic dynamics), synaptic efficacy was assumed to be an infinite (non-depleting) resource, so u was always 1 and the rate of change of p for synapse i, $\dot{p}_i$, simplified to:

$$\dot{p}_i = \delta(t - t_i) - \frac{p_i}{d_i} \tag{7}$$

Total synaptic current, I_{static}, for a postsynaptic neuron was then:

$$I_{static} = \sum_i \left[w_i(1 - p_i)(v - E_{rev}) \right] \quad (8)$$

Static synapses were used in all simulations in this study except for **Figure 5D** (for which dynamic synapses were used as above). For static excitatory-to-excitatory synapses the synaptic recovery time constant d was set to 50 ms, for excitatory-to-inhibitory 5 ms and for inhibitory-to-excitatory 40 ms. No inhibitory-to-inhibitory connections existed.

Network

The CXC network was devised for this study in order to investigate the relationship between the diffuse thalamocortical matrix loop and cortical dynamics, although precursor models have been published (Stratton and Wiles, 2010a,b). The CXC model is an abstraction of the essential computational components of the complex thalamocortical connection structure. It utilizes some of the known characteristics of these regions (such as dense local and more sparse long-range connections), while ignoring others (such as cortical layers). As such, it is a general representation only, with no intended specific spatial scale. Its purpose is to show that thalamocortical-like connectivity between simple neuron-like elements can result in non-trivial dynamics, and to point to some of the general neuron and network properties that may be involved.

The network consisted of $n_e = 1000$ regular-spiking (RS) excitatory pyramidal neurons and $n_i = 250$ fast-spiking (FS) inhibitory interneurons connected linearly (i.e., in a 1-dimensional network) with directional synapses. Synaptic efficacies were set such that several presynaptic spikes in close succession were required to cause an output spike in a postsynaptic neuron. Connection structure was set to be small world-like, with dense local connectivity and sparse random long range connections, similar to the cortex. Each RS neuron was connected to each of its closest j neighbors with local excitatory efficacy w_n/j where $w_n = 2$ (the end neurons connected circularly to the opposite end of the neuron vector). Random long-range connections of weight w_r/k where $w_r = w_n$ were then made from each RS neuron to each other RS neuron with probability k/n_e, giving an average of k long-range connections per neuron. In all simulations $j = 4$ and $k = 10$ (these numbers of connections were smaller than in actual cortex due to the limited size of the modeled network, and synaptic efficacies and currents were therefore scaled up accordingly). Inhibitory FS neurons were spread uniformly between the RS neurons. Each FS neuron received excitatory input from and projected inhibitory output to each of its closest l RS neurons with efficacy w_i/l where $w_i = 1$ and $l = 20$.

The thalamic reticular (RN) and intralaminar (IL) nuclei were each implemented as a single analog (non-spiking) neuron with output equal to the sum of its input currents. These nuclei are modeled as analog because their function is to balance cortical activity, and while the brain uses spiking neurons in these nuclei, we hypothesize that the summed action of many spikes from many neurons results in this homeostatic balance, so for the CXC model the analog representation is sufficient. RS neurons were innervated by IL projections with synaptic efficacy of 0.4. The IL received continuous steady (not time-varying) input current from the AAS (I_{AAS}) which ranged from 0 to 10 in different experiments. This upper bound on I_{AAS} was set to strongly depolarize the connected neurons without causing them to reach spiking threshold. Cortical neurons projected to the RN with efficacy 1, with synaptic currents decaying with time constant d, as above (Equation 7). The RN inhibited the IL with efficacy -1, completing the diffuse thalamocortical loop. Network activity was initiated by a single input pulse at time zero into a random selection of $n_e/2$ RS neurons. Conduction delays were set in a uniform random distribution between 1 and 25 ms for long range connections and to 1 ms for all other connections.

To test the sensitivity of the results to numerical factors, additional test simulations were conducted with very high temporal resolution (1 μs time-steps) and larger scales (100,000 neurons in the network). Additionally, simulations of 10.000 s duration were run to ensure that network dynamics neither failed nor entered a short limit cycle attractor state. All these simulations exhibited metastable network dynamics across their entire spatiotemporal extents.

Analysis

Input current from the AAS to the IL (I_{AAS}) was varied from 0 to 10 in steps of 0.01. For each input current level, a simulation was conducted for 10 s of simulated time. Spike times and membrane potentials of all neurons were recorded at each time-step of 0.1 ms. The first 1 s of activity was deemed to be the network settling period during which time the network activity would transition from synchronous bursting (initiated by the input pulse) to sustained activity (or to quiescence depending on the input level). The neural activity for this first second was omitted from further analysis. The remaining 9 s of the recordings were used to calculate the mean firing rate, r, of the network ($r = T/n_e/9$, where T is the total number of spikes generated by all RS neurons).

We estimated the local field potential (LFP) that would be recorded from this network at any moment in time, $lfp(t)$, by summing the membrane potentials of all neurons:

$$lfp(t) = \sum_m v_m(t) \quad (9)$$

where $v_m(t)$ was the membrane potential of neuron m at time t. Due to the vertical alignment of cortical dendrites in real brains, real dendritic potentials, when summed from many cortical neurons, contribute to the LFP (Nunez and Srinivasan, 2006). Neurons in this study were simulated as point entities with no dendritic processes, so the neuron (soma) membrane potential was the closest estimate of the dendritic potential. This process of calculating the estimate of the LFP is similar to that used by Beim Graben and Kurths (2008).

An indication of the complexity of the dynamical network state was based on deviations of the interspike interval (ISI) distribution from exponential to multi-modal. During metastable dynamics, three modes (short, medium and long) were evident

in the ISI distribution; these modes arose explicitly due to the processes of active integration and segregation in the network. Short ISIs were caused by bursty firing of a neuron, which occurred when the neuron was receiving strong synaptic input from neighboring neurons and from long range connections from distant neurons, and hence indicated that the neuron was being integrated into network activity. Long ISIs were caused by long periods when a neuron did not fire, which occurred when the neuron was receiving weak or no input from neighboring neurons and from long range connections, and hence when the neuron was segregated from network activity.

Spikes firing within the central ISI mode, midway between the short and long modes, were simply following the dominant oscillation frequency of the network, exhibiting neither enhanced integration nor segregation from network activity. Small integration and segregation peaks, when compared to the central peak, therefore indicated that most spikes were entrained to the dominant network oscillation and neuron firing was predominantly periodic, whereas simultaneously-large integration and segregation peaks indicated that the network was in a complex metastable dynamical regime.

Finally, the network *trapping time* was calculated (Marwan et al., 2002). Trapping time quantifies the amount of time a network remains in a given state before transitioning to a new state. The trapping time was determined by first constructing a matrix S of states of the network spiking activity over 0.2 s non-overlapping time windows. Within each time window j, the number of times each RS neuron i emitted a spike was counted:

$$S_{i,j} = \left|\{0.2(j-1) \leq t_i < 0.2j\}\right| \text{ for all } i, j \tag{10}$$

where t_i is the set of spike times of neuron i and $|\{...\}|$denotes the length of a set.

The correlation matrix of S was calculated and the resulting matrix was thresholded at 0.5, yielding the state recurrence matrix R:

$$R = H(\text{corr(S)} - 0.5) \tag{11}$$

where H is the Heaviside step function. R is a symmetric matrix that reveals, for all states j, which other states were similar (i.e., which other states had a similar pattern of firing neurons). Finally, the mean width of the super-threshold region surrounding the main diagonal of R was determined (i.e., for each 0.2 s window represented by an element on the main diagonal of R, the width of the super-threshold region around this element was measured perpendicular to the diagonal by stepping outwards from the element until a sub-threshold element, where $H_{i,j} = 0$, was found; the mean of these widths for all elements along the main diagonal of R is the trapping time of the network). Trapping time is interpreted as follows: If the network transitioned between states rapidly, then adjoining state vectors of S had sub-threshold similarity and the diagonal of the recurrence matrix R (the trapping time) was only one state wide (i.e., each state was similar only to itself). However, if the state transitions were slow, then adjoining state vectors of S had super-threshold similarity and the width around the diagonal of R was greater than one. Longer trapping times indicated that the network state was evolving more slowly. Trapping times approaching 10 s (the length of most simulations in this study) indicated that the network dynamics were fixed in an attractor state in which activity did not evolve at all.

Results

Autonomous (Self-sustained) Metastable Dynamics

To establish the baseline complex dynamics supported by the CXC network, constant input from the AAS was first set to the baseline level of 1 and synaptic conductances were set to standard values for the network (see Methods). No noise sources were used within the network or in its input. The network exhibited metastable dynamics (**Figure 2**) despite being deterministic. We have previously investigated the network properties required for a spiking network (a precursor to the CXC) to exhibit such dynamics (Stratton and Wiles, 2010a,b)—these include small-world or scale-free cortical connectivity, a mechanism of global inhibition, and maintenance of dynamics in a critical (phase transition) state. The parameters of the CXC network of the current study were set to be in the critical region. Metastable network dynamics were robust to moderate network parameter changes, however significant modification of these parameters resulted in either collapse of network dynamics into a limit cycle resembling seizure or failure of activity to propagate causing the total network activity to fall to zero. These two states of seizure and quiescence are the low-complexity states between which the activity in the CXC network rapidly, partially and transiently switched to generate the metastability (Stratton and Wiles, 2010b)[1]. The cortico-thalamo-cortical feedback from the matrix connections played a homeostatic role, lowering cortical input as activity levels increased and raising input as activity levels decreased, ensuring that the network as a whole remained within the critical region of phase space. In all simulations, these autonomous, complex, non-periodic dynamics lasted indefinitely.

Inspection of the firing of neural assemblies (groups of neighboring neurons) in the spike raster plot shows that they occurred approximately 0.05 s apart (**Figure 2**). This oscillation period was determined by the interplay of the cortico-thalamo-cortical feedback and the time constant of the local synaptic inhibition. Despite the regular oscillation period, these assemblies formed at unpredictable, apparently random times and existed for unpredictable durations (**Figure 2**; also see Stratton and Wiles, 2010b). Dynamic formation of assemblies was unpredictable because it depended on three tightly coupled factors: (1) which assemblies were currently firing, (2) the precise pattern of random long-range connections from the currently active assemblies, since this pattern would determine which of the currently inactive neurons were receiving strongest synaptic input, and (3) the past history of firing activities, since neurons

[1]While the subcortical structure of the CXC is more complicated than that presented in this earlier paper, it fulfills the same function and the extra detail has been included for biological fidelity. Parameter sensitivity is therefore not significantly altered in the CXC.

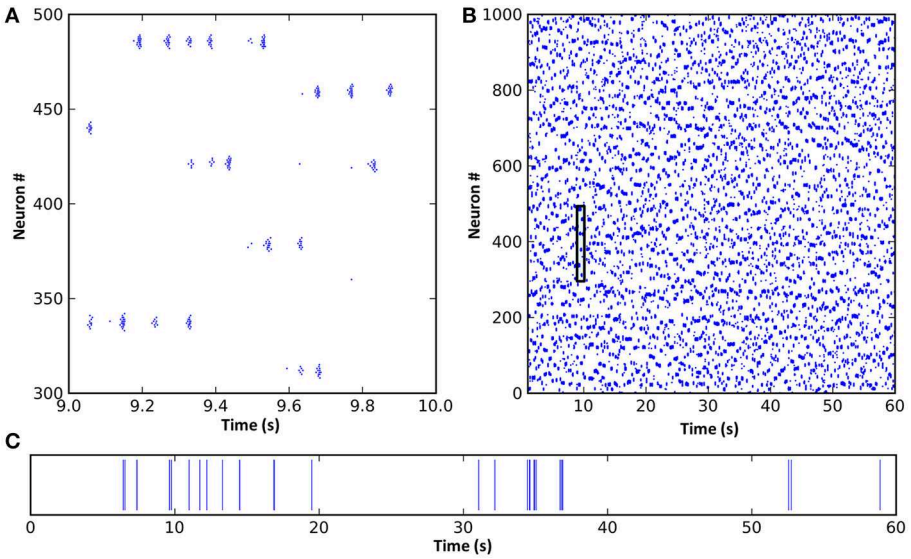


FIGURE 2 | A spike raster plot of CXC network activity shown at different zoom levels reveals characteristics of complex network dynamics. These include network-wide oscillations, groups of neighboring neurons unpredictably forming transient oscillating assemblies, and ongoing non-periodic activity. **(A)** Spiking activity of 200 neurons for 1 s. Each horizontal group is an assembly of bursting neurons. Each neuron within an assembly bursts at high frequency, with the bursts occurring every 0.05–0.1 s (10–20 Hz) while the assembly is active. Assemblies are transient, sometimes firing synchronously with other assemblies in the network and sometimes asynchronously. Examples of both synchronous and asynchronous firing can be seen. **(B)** The entire network shown for the full 60 s of simulation. The transient and unpredictable nature of assembly formation and duration is evident. The time period shown in panel **(A)** is boxed. Mean firing rate for all neurons was just 1.0 Hz. **(C)** Spike raster plot for one typical neuron over 60 s, showing random firing with burst activity punctuating extended periods of quiescence. Most neurons had lengthy periods of complete silence lasting 10 s or more despite having mean firing rates around 1 Hz and the network as a whole oscillating at 10–20 Hz.

which had been recently active could still have been in a relative refractory period, meaning other neurons could fire first even if they were receiving weaker synaptic input. Thus, even though the network dynamics were noiseless and deterministic, they were unpredictable unless the entire network state (both static and dynamic, i.e., the entire connectivity map, the complete properties of every neuron and synapse, and the precise current states of all membrane potentials and synaptic currents) was known with absolute accuracy. In other words, the only way to predict the ongoing activity would be to create an exact duplicate of the network, and any inputs impinging upon it, and simulate it in its entirety. (Interestingly, this inability to accurately predict model dynamics for the CXC network applies equally to real brains).

Integration and Segregation

For the network with baseline AAS input level (equal to 1, as in **Figures 2**, **3A**), LFP power was strongest at approximately 19 Hz, indicating the dominant oscillation frequency in this network (**Figure 3A** center). The ISI distribution appeared monotonic on a linear scale (**Figure 3A** right inset), but by exploiting the fact that very small ISIs are excluded by the refractory period, we plotted the ISIs on a log scale to give superior count resolution for short ISIs (**Figure 3A** right). On a log scale several peaks were apparent; the ISI distribution was clearly multi-modal. The central mode, occurring at just below 0.1 s (−1 on the log scale), was caused by the dominant 10–20 Hz oscillations. This mode could therefore be considered the "base mode" or default ISI of this network. In this respect, spikes occurring 0.05–0.1 s apart carried little information; these spike timings were predictable based on the observed dominant oscillation frequency, and in an information theoretic sense, the more predictable an event, the less information it conveys (Shannon et al., 1949). Unexpected, apparently random deviations from this base mode, however, can carry much larger amounts of information. Interestingly, for this network these deviations occurred in a manner suggestive of both integration and segregation of network elements, as follows:

The mode with shortest ISI, occurring at less than 0.01 s (−2.4 on the log scale), was caused when neurons integrated into assemblies. When a neuron received synaptic input from its local neighbors and from random long-range connections from other active assemblies elsewhere in the network, it would fire at much higher rates than baseline and become integrated by and into the network activity.

Similarly, the mode with longest ISI in the ISI distribution (**Figure 3A**, right), occurring at approximately 1 s (0 on the log scale) and beyond, was caused by neurons firing less often than expected given the network oscillation frequency, or equivalently, neurons being excluded from firing. The majority of the principal neurons in the network had lengthy periods of complete silence lasting 10 s or more despite the network as a whole displaying regular oscillations of much shorter periods.

Input Level from the Ascending Arousal System Changed Dynamical State

When constant input from the AAS was increased, several dynamical properties of the network were altered (see **Figure 3B**,

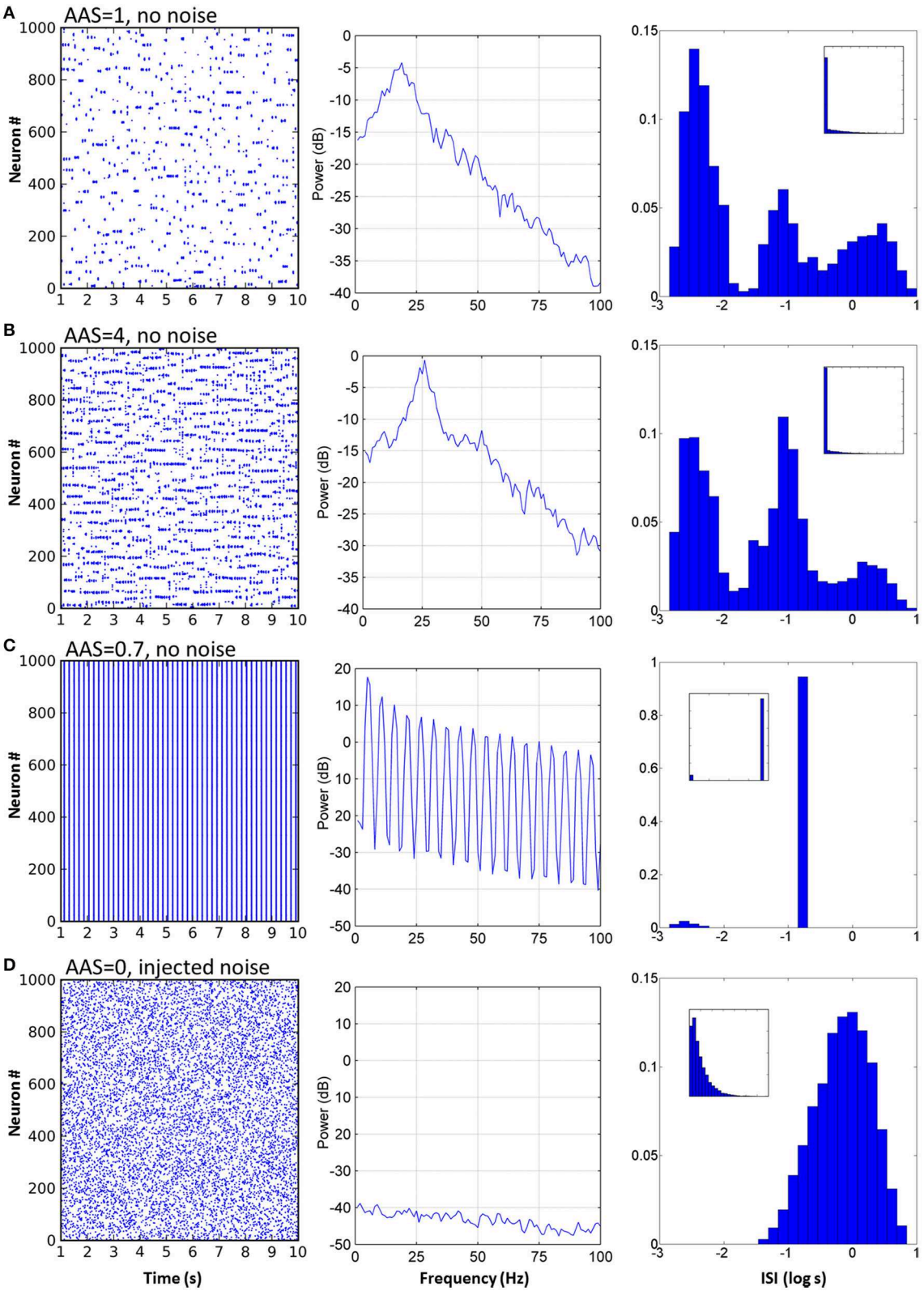


FIGURE 3 | Varying AAS input induced very different dynamical states in the CXC network. (A) With AAS input set to 1, complex dynamics emerged despite the absence of noise in the system. The spike raster plot for all the principal cells for 10 s of activity is shown (the first 1 s of network settling time was excluded) (left). The Local field potential (LFP) showed a

(Continued)

FIGURE 3 | Continued
dominant 19 Hz oscillation (center). Firing rate variability was high; neural activity occurred in combinations of strong bursts and long periods of inactivity. The ISI distribution, shown as a probability density, was therefore multi-modal (right), indicative of integration (bursts), and segregation (extended periods of inactivity—see text). Displaying ISI with log time enhances resolution for short ISIs (main panel, right; inset on right shows the ISI distribution with linear time). Network parameters in this panel match those shown in **Figure 2** above. **(B)** Increasing AAS input levels resulted in increased firing rate (left), higher dominant oscillation frequency (center) and a decrease in the number of spikes in the high and low modes of the ISI distribution (right) compared with the central mode, indicating that more spikes were becoming entrained to the dominant oscillation as AAS input increased. **(C)** Low but non-zero AAS input and no noise resulted in a very different dynamical state characterized by all neurons being entrained to a single low oscillation frequency (left). The LFP was characteristic of a rectangular wave (similar to a square wave but with a shortened or lengthened duty cycle) (center). Almost all ISIs were identical (right). **(D)** AAS input and all synaptic conductances in the network were set to zero and noise was added to induce random firing at 1 Hz (left). Mean firing rate was identical to that in **(A)** but the LFP was consistent with filtered noise, attenuated slightly at higher frequencies by the neural membranes (center). The ISI distribution was nominally exponential (inset, right) with the spike refractory period excluding short ISIs (main panel, right).

with four times the AAS input). The network firing rate increased as a direct result of the increased input allowing a greater number of neuron assemblies to be simultaneously active. The dominant network oscillation frequency also increased, due to the stronger input causing faster depolarization of the neurons after each oscillation cycle; the oscillation peak shifted from 19 to 26 Hz (**Figure 3B** center). In the ISI distribution, more spikes occurred in the central baseline mode and fewer in the outlying short and long ISI modes (**Figure 3B** right). This change in distribution indicated greater predictability of the spike train—a greater proportion of spikes were being entrained to the overall dominant network rhythm and fewer were exhibiting enhanced integration with or segregation from this overall network activity. The reason for the reduced segregation is straightforward—with more neurons firing there are fewer opportunities for neurons to be silent for extended periods. On the other hand, the comparatively lower integration into assemblies is due to stronger inhibition (which itself is due to the increased overall activity in the network); each oscillation cycle is shorter, so fewer spikes are generated by each neuron in each cycle before the assembly is silenced by inhibition.

For some small but non-zero levels of input from the AAS, the network entered a very different state (**Figure 3C**, AAS input lowered to 0.75). The dominant oscillation frequency decreased markedly and virtually all neurons were entrained to the global oscillation. This state had a very low complexity, as can be seen from the single large peak in the ISI distribution at the dominant frequency. In this state, the spike times of all the neurons verged on wholly predictable. It is not immediately evident why this state arose. However, two observations may be relevant: (1) the state was not strongly stable, with dynamics settling into either low or high complexity regimes under initial conditions with very small differences, and (2) the state was abolished by noise (see Section Robustness to Noise, below). These observations suggest that the network dynamics were bistable in this region of phase space (a not-uncommon occurrence in non-linear dynamical systems).

To establish a control condition with Poisson firing, which reveals the differences in dynamics when network activity is dependent on extrinsic noise rather than on the network structure, AAS input and all synaptic conductances in the network were set to zero efficacy and sufficient noise was added to the neural membranes to induce random spiking activity at 1 Hz (**Figure 3D**, left). The LFP showed a power spectrum characteristic of low-pass filtered noise; higher frequencies were filtered by the membrane capacitance of each neuron (**Figure 3D**, center). The ISI distribution took the expected exponential shape curtailed by the spike refractory period. Poisson firing highlights the contrasts in dynamics between random activity and metastability, where the average firing rates were equal but the higher order statistics were dissimilar.

With the extrinsic noise removed and synaptic conductances reinstated (i.e., the network restored to standard baseline), AAS input was then varied from 0 to 10 in steps of 0.01, the mean firing rate for each instance was recorded, and the network *trapping time* was calculated (see Methods for details). Trapping time quantified the amount of time the network tended to remain in a given state before transitioning to a new state.

Mean firing rate increased linearly as AAS input was increased (**Figure 4**, top), except for low input values where firing rate was typically either zero or extremely high. The regions of zero network firing were generally for very low AAS input values near zero, while regions of very high firing rate occurred for slightly higher input levels between 0.6 and 0.8. The transformation between low and high firing rates was sudden, with no intermediate states. The dynamics in the high firing rate regions were as shown in **Figure 3C**—low complexity, global entrainment, and slow oscillations around 5 Hz. In these high firing rate regions the trapping time was high (**Figure 4**, bottom), signifying that the network state was not changing (marked as point **c** in **Figure 4**).

With increasing AAS input to just above 0.8, network dynamics again switched dramatically, in this instance to a state of low trapping time (high complexity) and low firing rate, as shown in **Figure 3A** (marked as point **a** in **Figure 4**). Trapping time and firing rate increased with further rises in AAS input (point **b** and beyond). The firing rate increase has been explained above. The trapping time increase is related to, and perhaps caused by, the increase in firing rate; specifically, by the increase in the number of simultaneously-active assemblies. When a greater number of assemblies are active at any given time, there are fewer long-range connections that project to currently-inactive assemblies. Only inactive assemblies have a chance of switching to an active state and changing the overall state of the network. The consequence is that, as firing rate increases, there is decreasing probability of a change in state at any given time, as reflected in the increasing trapping time.

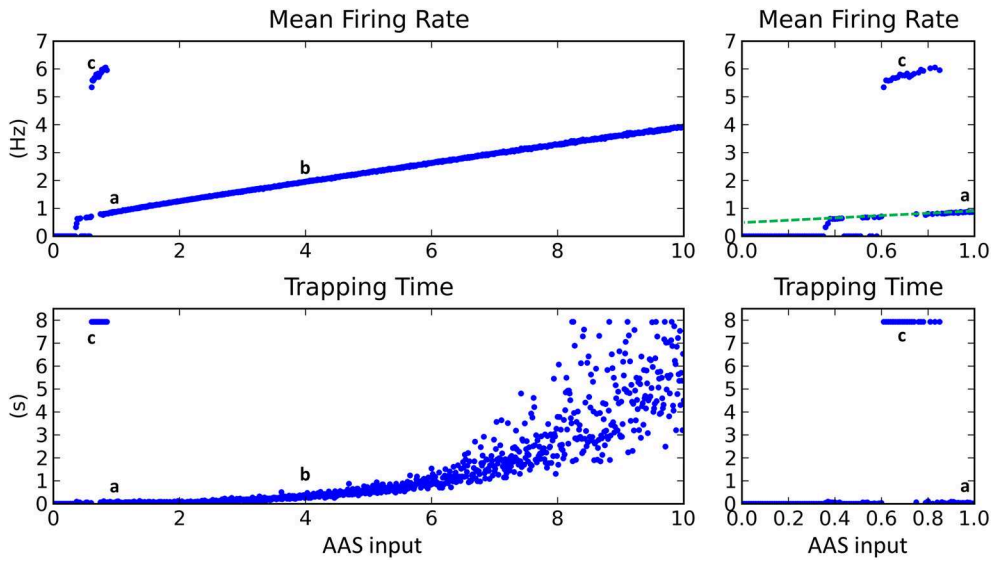


FIGURE 4 | Changing input level from the AAS dramatically affected network dynamics. Left panels—AAS input levels from 0 to 10. **Right panels**—close-up on AAS input from 0 to 1. Cases a–c from **Figure 3** are marked on the graph (top). Very low input levels below 0.6 usually resulted in no network activity, with sporadic instances of complex dynamics occurring for AAS input levels between 0.35 and 0.6 (see close-up panels on right). At input levels between 0.6 and 0.8, global entrainment at high firing rates but low oscillation frequencies emerged abruptly, again with some sporadic interspersed instances of complex dynamics. The sporadic large changes in firing rate and trapping time for low AAS input levels (between 0.35 and 0.8) are characteristic of network dynamics being bistable, with the random initial conditions for each network instance controlling which of the stable states the network settled into in each case. A small increment in AAS input could therefore result in a large change in dynamics, as can be seen in the close-up panels on the right. At AAS levels beyond 0.8, a sustained switch to complex dynamics occurred (i.e., the bistability vanished). At high levels of AAS input, firing rate and trapping time increased. Adding noise to the neural membranes removed the bistability and caused complex dynamics to be sustained for all levels of AAS input down to zero (dashed line, top right).

Connectivity Impacts Dynamics

In all of the above simulations, long range cortical connections were set randomly with uniform probability. It is also possible to base connection probability on distance between the cells, so that cells that are nearer have a greater chance of having a long range connection between them. Such a connection scheme resulted in "waves" of activity propagating through the cortical cells (**Figure 5A**). These waves arose due to the connections from any active region in the network projecting most densely to regions that were immediately adjacent. These adjacent regions, receiving the strongest input, were the most likely to become active next. As this process repeated, the result was a wave of activity propagating through the network. However, since the propagation was chaotic, the speed, number and even the direction of the propagating waves could vary unexpectedly. Alternatively, by dividing cortical neurons into two groups such that intergroup long range connections were less likely than intragroup, activity would unpredictably switch between the groups (**Figure 5B**). The activity switch occurred for a similar reason to the activity waves in **Figure 5A**, except that the connection probability within each group was fixed and equal between all neurons, causing all neurons within a group to have equal chance to become active.

Reducing the radius of local inhibition resulted in variable neuronal assembly sizes and much faster oscillations (**Figure 5C**). Assembly sizes varied because the reduced radius of inhibition was unable to contain the surrounding excitation, meaning that adjacent assemblies could link together and form larger super-assemblies. Oscillation frequency increased because local inhibition could no longer provide a stabilizing effect on the network dynamics; instead, local excitatory connections could form tight recurrent loops limited in frequency only by axonal conduction delays and membrane dynamics. Finally, increasing local inhibition and excitation strength and adding short-term plasticity (STP) to the excitatory synapses resulted in network dynamics which would spontaneously and intrinsically enter and exit seizure states (**Figure 5D**). STP seems to be involved in the generation of seizure-like dynamics in the model, since networks without STP do not exhibit these seizure states. Seizure onset was caused by chance synchronous firing, during one oscillation cycle, of a larger number of neurons than normal, simultaneously facilitating a large number of synapses, which pushed the network into a hyper-synchronous state that was then perpetuated (often beyond the facilitation time constant) by feedback through the thalamic matrix connections. Seizure termination also seemed to occur by chance, with a spike from a cell either not involved in the seizure or at a time not synchronized with the seizure oscillations. Such a "wayward" spike could sufficiently affect the timing of subsequent spikes in the network to break the hyper-synchronicity. However, a deeper understanding of the mechanism or mechanisms involved in seizure onset and termination would require further studies that are beyond the scope of the current paper.

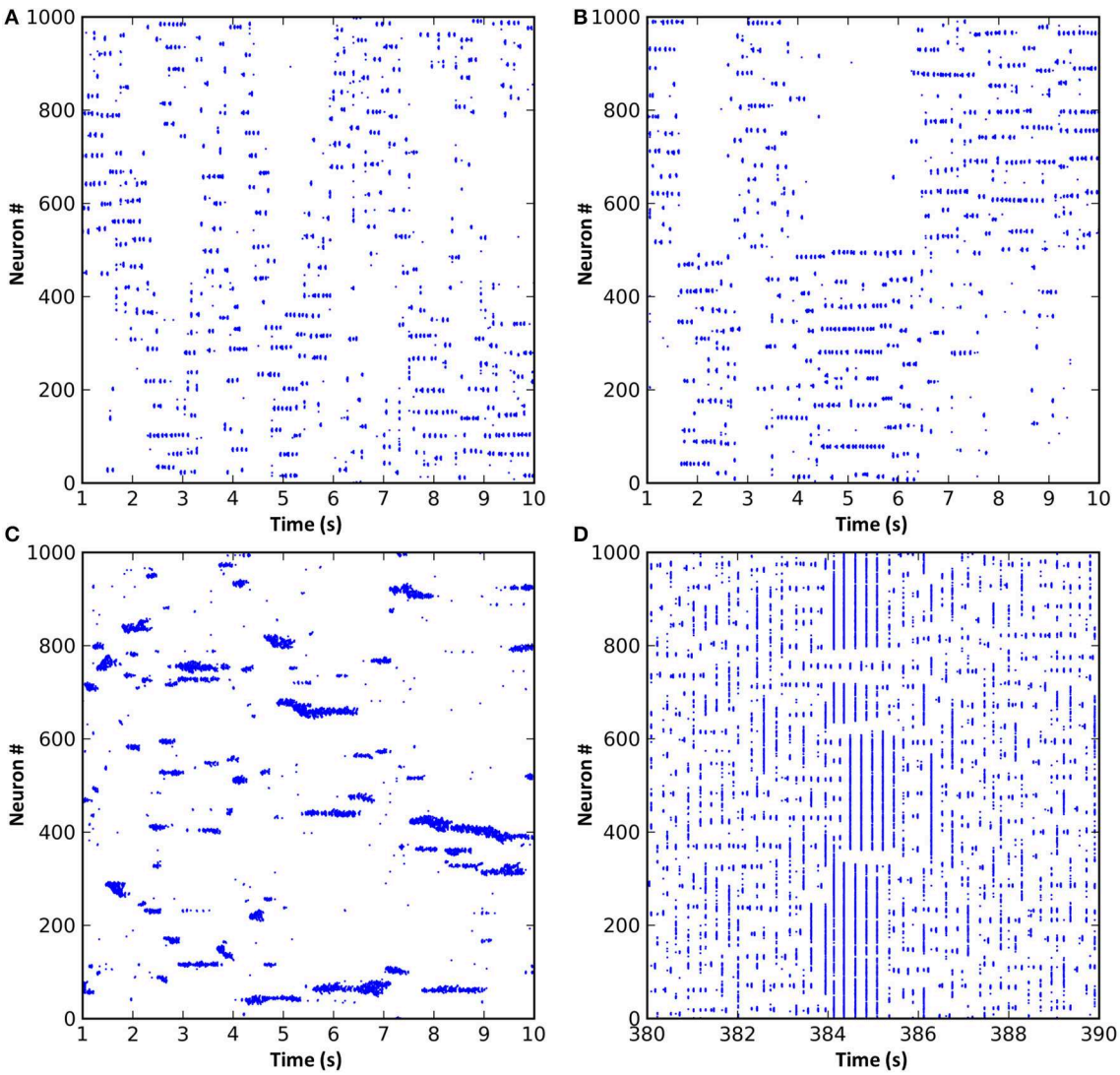


FIGURE 5 | Modifications to the synaptic connection schemes resulted in changes to the network dynamics. (A) Long range connection probability based entirely on distance between neurons resulted in propagating waves of activity. **(B)** Dividing the neurons into two groups where neurons within a group were more likely to be connected resulted in a dynamic switching of activity between groups. **(C)** Reducing the radius of local inhibition to less than that of local excitation resulted in the formation of variably-sized neuronal assemblies and much faster oscillations. **(D)** If short-term synaptic dynamics are modeled and excitation and inhibition strength increased, a network that exhibits spontaneous entry to and exit from seizure states can result.

In the version of the CXC network presented in this paper, the pedunculopontine, reticular and intralaminar nuclei are each modeled as unitary entities with no smaller components or sub-nuclei. While this simplification is adequate to demonstrate the functional involvement of these regions in long range segregation of cortical activity and facilitation of complex dynamics, in reality these nuclei are comprised of distinct components with distinct connection patterns. In particular, the "non-specific" matrix projections from the IL to the cortex may in fact have specific synaptic targets (Groenewegen and Berendse, 1994). This specificity has consequences for cortical dynamics: rather than cortical regions competing uniformly for activity, the competition (segregation) occurs in interconnected, hierarchically-arranged overlapping pockets of varying sizes across the entire cortex. With such a connection paradigm, total cortical activity remains controlled at all times, but inter-regional competition can vary in complex ways based on IL connection patterns and precise patterns of cortical activity at any given time. The result is that the repertoire of cortical dynamics is potentially even richer than presented here.

The network is not limited to linear (1-dimensional) layouts; 2-dimensional, and greater, network arrangements work equally as well (see online[2] for details). Overall, this and the above

[2]For an animated example of 2-dimensional cortical dynamics in the CXC see https://dl.dropboxusercontent.com/u/70159706/gtrajectory_out.avi.

preliminary results (**Figure 5**) suggest a strong impact of network connectivity on the ensuing neural dynamics that warrants further investigation. Understanding how neural dynamics are influenced by specific connection structures in the brain such as the directed connections between regions at the macro scale, intra-regional connectivity inside structures like the hippocampus at the meso scale, and detailed neural circuitry at the micro scale, is critical to understanding how these structures perform their diverse functions.

Robustness to Noise

To test the robustness of the network dynamics to noise, sufficient membrane noise was added to all principal cells to induce random spiking at 0.1 Hz in each cell; that is, spiking occurred at 100 Hz across the network of 1000 cells. (Note that this is a very different experiment to that depicted in **Figure 3D**, where the network was effectively disabled by setting synaptic efficacies to zero prior to the injection of noise.) In the case of adding noise to the functioning network, there was no significant change in the mean firing rate, integration and segregation, or the trapping time for most AAS input levels. However, for AAS input between 0 and 0.8, the transitions to low complexity states with high mean firing rates and high trapping times did not occur; instead, complex dynamics continued for all input levels down to zero (**Figure 4**, dashed line in top right panel). This continuation of complex dynamics was due to random spiking activity adding to the total activity in the network, similar to sustained higher AAS input, and holding the network in a metastable dynamical state. Crucially, the network's ability to maintain metastability and long range segregation of activity was not compromised by noise.

Discussion

Segregation, Metastability, and the Thalamocortical Matrix

The primary finding in the current study has been that the AAS-IL-RN circuit may have the ability to globally segregate cortical activity and maintain metastable cortical dynamics, with no need for injected noise or input perturbations, and that the global cortical state could be strongly modulated by brainstem input to the thalamus. In the absence of long-range inhibition, and the absence of long-range excitation onto local inhibitory interneurons, the long-range competition and network segregation observed in the CXC could only occur through the lowering of excitation from the IL. Local and long-range corticocortical connections in the CXC network were not strong enough to sustain activity at low firing rates—a depolarizing input from the IL was required to boost neural membrane potentials sufficiently close to threshold so that a small number of presynaptic spikes was capable of causing a spike in the postsynaptic neuron. If total activity in the CXC network increased, tonic depolarizing input from the IL decreased, and neurons or assemblies that would have fired due to convergent input from other active assemblies would then be unable to fire. This is competition through *effective* inhibition (inhibition that occurs via the withdrawal of tonic excitation), controlled by the total amount of activity across the network. Simultaneous, transient, unpredictable, recurring integration and segregation of activity in the CXC network resulted in the observed on-going metastability. Without these thalamocortical matrix connections, sustained metastability at low firing rates, long range competition, and global segregation of cortical activity would not be possible.

The activity states observed in the CXC network have analogs in real brains. The slow oscillations and global coupling seen at low AAS input levels in the CXC network are similar to the strong delta oscillations and very low dynamical complexity observed during deep (non-REM) sleep (Massimini et al., 2005; Murphy et al., 2009). In mammals, AAS input levels, particularly input from the pedunculopontine nucleus (PPN, or sometimes PPT or PPTg) are significantly reduced during periods of non-REM sleep (Moruzzi and Magoun, 1949; Reese et al., 1995; Jones, 2003), with slow delta oscillations at 0.5–4 Hz the predominant neural activity signature of this state. The transition between sleep and active cortical states is thought to be driven by cholinergic neuromodulation. However, there is likely to be more than one mechanism involved, and this result in the CXC network suggests that the transition could be assisted by the lowering of AAS input; supporting the switch from a state of high to low dynamical complexity.

At slightly higher AAS input levels than those required to generate sleep-like dynamics, the CXC network switches into a low firing rate, high complexity state. This state is analogous to awake states of low cortical arousal and metastability seen in real brains at intermediate levels of PPN activity, and observed during quiet relaxation (Stam et al., 1999). The characteristic neural activity signature of this state is relatively slow alpha oscillations (8–12 Hz, just lower than the dominant oscillation frequency observed in the CXC network) seen over large parts of the cortex. Increasing AAS input beyond this level in the CXC network resulted in increased firing rates, increased dominant oscillation frequency and increased trapping time—the length of time for which neural assemblies were stable—while decreasing the dynamical complexity. The decrease in complexity arose due to more spikes being entrained to the dominant oscillation and fewer spikes exhibiting either integration into assemblies or segregation from network activity. These states in the CXC network may correspond to states of attention, arousal and vigilance in real brains when AAS input is known to be maximal.

Default Mode Network

Free association and daydreaming tend to occur when the brain is not actively processing sensory stimuli or task-related events, and as such these states are correlated with activity in the default mode network (Mason et al., 2007; Buckner et al., 2008; Christoff et al., 2009). It follows that when the brain switches state from free association to attention, activity tends to switch from the DMN to other parts of the cortex. The regions of the cortex involved in the DMN are more strongly connected to each other than they are to the rest of the cortex (Buckner et al., 2008). We have shown that when a similar connection strategy is employed in the CXC network—dividing the cortical neurons into groups where intergroup connections are less likely than intragroup—activity dynamically and unpredictably switches between groups.

In the CXC network, activity tends to concentrate in one group at a time because the recurrent long range connections required to sustain ongoing activity (Stratton and Wiles, 2010b) are focused within groups. Switching between groups is driven by a combination of the long range connections and the input from the AAS; when several long range inputs to a new group are activated simultaneously, perhaps coinciding with a decrease in activity of the currently-active group due to either habituation or random activity fluctuation, then input from the AAS can be sufficient to ignite a larger number of assemblies in the new group. Activity in the new group then competes for persistence with the current group through the segregation process mediated by the AAS and the thalamic matrix connections, and if activity in the new group is strong enough then it will dominate and an activity switch occurs. This dynamic switching in the CXC network provides a parsimonious explanation and inherent mechanism for activity in the DMN and other cortical regions being mutually exclusive despite the limited connectivity (especially inhibitory connectivity) between them. In the brain, the timing of switches in activity between the default mode and task-positive networks is likely also influenced by task requirements and external events in the perceived environment, rather than being purely chaotic, but the segregation principle remains the same.

Consequences for Brain Function

Sustained firing of neural assemblies has been proposed as the neural substrate of working memory in the cortex (Wang, 1999; Pesaran et al., 2002; Jensen, 2006; Jensen et al., 2007). In the CXC network, the state of higher firing rates, increased trapping times and faster oscillations is analogous to a state of increased vigilance, attention and working memory in the cortex (Oken et al., 2006), driven by an increase of input from the AAS (specifically the PPN). There is ample evidence that increasing input from the AAS causes increasing cortical activation in general (Jones, 2003), and that firing rate and oscillation frequency also increase specifically for those neurons representing attended stimuli (Fries et al., 2001). There is also some evidence that increased vigilance does indeed reduce spiking variability (meaning that complexity of spiking patterns is also reduced) (Falkner et al., 2013). The converse state of low firing rates, short trapping times and high complexity in the CXC network is driven by lower AAS input levels and is associated with slower oscillations. This state is analogous to fluid states of mentation such as mind-wandering or daydreaming (Laufs et al., 2003; Mason et al., 2007; Buckner et al., 2008; Christoff et al., 2009) where working memory is not heavily utilized and mental associations arise freely and apparently randomly due to the increased variability of the neural dynamics.

The CXC network has a clear dichotomy between relaxed free association and vigilant attention; these modes cannot occur together because they are distinct dynamical states of the thalamocortical system, driven by changing input levels from the AAS. Dynamical states of low complexity are necessary for stable maintenance of working memory, with the trade-off that fewer potential states will be visited due to longer network trapping times. Dynamical states of high complexity and short trapping times are necessary for exploring more of the possible state space combinations of neural representations, supporting mind-wandering and free association, but are ineffective when focussed attention and working memory are required. Based on the model's behavior, we conjecture that states of high and low complexity are both useful but cannot co-occur, so the brain switches between them as need and opportunity arises.

Several predictions about neural dynamics can be made from the model:

1. Overall cortical activity should increase approximately linearly with increasing stimulation from the AAS. This correlation could possibly be measured experimentally using fMRI.
2. Increasing AAS input should cause increased trapping time in the cortex (i.e., longer activation of neural assemblies and slower transitions between cortical states). This relationship could be quantified using fMRI or multi-electrode electrophysiology recordings.
3. Increasing AAS input should also reduce the complexity (increase the predictability) of dynamic cortical activity patterns. Predictability can be quantified by calculating the entropy of or the mutual information in recorded spike trains (Dorval, 2008).

What can the CXC network tell us about computation in the brain, rather than simply brain dynamics? To adequately address this question requires consideration of both representation (how does brain activity represent information?) and learning (how does this information come to be in the brain?); these are substantive and intricately connected topics which are the focus of much research today. In its current form, the CXC network offers mechanisms behind the intrinsic, autonomous generation of metastability that seem to be required for complex thought (Binder et al., 1999; Mazoyer et al., 2001; Breakspear et al., 2003; Buzsáki and Draguhn, 2004; Fries, 2005; Tognoli and Kelso, 2014) and for the stable maintenance of dynamic neural assemblies required for working memory (Wang, 1999; Pesaran et al., 2002; Jensen et al., 2007). These apparently conflicting requirements are addressed through the control exerted over network dynamics by the level of input from the AAS to the thalamus. With these mechanisms in place, the dynamics observed in the CXC can potentially inform existing and perhaps even all-new theories of information transmission and transformation in the brain. Likewise, theories of learning, representation and mental processing can be tested in, and may lead to refinements of, the CXC. Initial questions include:

- Can repeated activity sequences be embedded in the network and spontaneously replayed—spontaneous sequence replay has been seen in animals during sleep and decision-making.
- Can plasticity mechanisms bias network dynamics to functionally integrate given cortical regions on demand (based on specific internal or external triggers)—this would help answer how the brain learns to functionally connect brain regions for the propagation of information as required.

Subsequent questions may address more complicated issues, such as how complex representations can be constructed in dynamical neural assemblies and utilized by the brain for purposeful computation. The ultimate goal of this future research will be to link the observed network dynamics with computational states in the brain. While realization of this goal is clearly distant, studying the CXC network will potentially lead in this direction.

Pathological Dynamics

Many brain disorders are associated with alterations to cortical dynamics, the DMN and the AAS (Garcia-Rill, 1997; Schnitzler and Gross, 2005; Uhlhaas and Singer, 2006; Buckner et al., 2008; Fröhlich et al., 2008; Fox and Greicius, 2010; Zhang and Raichle, 2010). We have demonstrated the dependence of autonomously-generated metastable dynamics on network connectivity in these regions. Because up to 99% of cortical connections derive from the cortex (Braitenberg and Schüz, 1998) and only a small fraction of the brain's energy consumption is used for the processing of external events (Raichle and Mintun, 2006; Zhang and Raichle, 2010), by far the majority of cortical activity originates in and is driven by other cortical activity. Changes in cortical structure driven by either learning or disease may therefore affect ongoing dynamics in unpredictable ways. Furthermore, these changes in dynamics may often lead to further structural change, creating a recursive dependency between dynamics and structure that could possibly lead to pathological states, such as in epilepsy, sleep disturbance, schizophrenia and many other brain disorders. By modeling autonomous metastable activity in the AAS-IL-RN circuit, this recursive chain of structural-dynamical co-dependence remains intact, allowing the investigation of how this dependence may lead to pathological states. Such studies are directions for future research with the CXC model.

Advances in the CXC Model

The thalamocortical matrix loop modeled in the CXC network is reminiscent of a winner-take-all (WTA) mechanism, however it is more accurately described as implementing winnerless competition (WLC). In WLC, a clear winning neuron or neural assembly never emerges from the competition; instead, each winner is immediately displaced by the next, resulting in continuously-evolving complex dynamic activity patterns (Akrami et al., 2012; Rabinovich et al., 2012). WLC can occur in networks of neurons mutually connected with inhibitory synapses, as observed in some simple animals during the generation of unpredictable behavior (Levi et al., 2005). It has also been shown analytically to occur in networks of neurons connected with slow global inhibition (Ermentrout, 1992). One of the insights offered by the CXC model is how the thalamocortical matrix can provide the necessary global inhibition to implement WLC for the control of complex dynamics across the cortex.

Random networks with balanced local excitation and inhibition have previously been shown to exhibit chaotic dynamics (Van Vreeswijk and Sompolinsky, 1996). Since then, studies have shown that networks connected using small world principles can also exhibit complex dynamics (Sporns et al., 2000b; Sporns and Tononi, 2001; Roxin et al., 2004; Riecke et al., 2007; Shanahan, 2008). However, for common topologies of these networks, the regions of parameter space where metastability was evident was small (Breakspear et al., 2003; Shanahan, 2008). More recently it has been shown that activity-dependent synaptic depression (a short-term decrease of synaptic efficacy based on postsynaptic activity) can massively enlarge the critical region where metastability occurs (Levina et al., 2007), and that the voltage-dependence of synaptic currents can stabilize complex dynamics for long periods of time (Kumar et al., 2008). Most recently, networks organized into hierarchical modules, where intra-module connections are abundant and inter-module connections are sparse, have been studied (Rubinov et al., 2011; Wang et al., 2011). Unlike previous networks, these latest networks (Kumar et al., 2008; Rubinov et al., 2011; Wang et al., 2011) can exhibit *irregular* sustained activity at low firing rates. For all of the above-mentioned networks, however, some or all of these questions remain open:

- How can complex dynamics be sustained indefinitely?
- How can low average firing rates, as seen in cortex, be obtained?
- Can dynamics be maintained without injection of extrinsic noise?
- Can very long interspike intervals (tens of seconds) be achieved?
- How can explicit segregation of activity (as against just a lack of integration) be accomplished?

In the current study, we have shown how the thalamocortical matrix connections control cortical dynamics and resolve the above issues. Complex metastable dynamics in the CXC network continue indefinitely with no need for injected noise, at low firing rates and with very long ISIs occurring frequently. Most importantly, the matrix connections can cause explicit dynamic segregation of network activity through withdrawal of tonic excitation—a process we have termed *effective* inhibition. Effective inhibition leads to activity between weakly-connected regions (such as the default mode and task positive networks) being significantly anticorrelated rather than simply uncorrelated. A mechanism by which effective inhibition can arise globally across the cortex has not previously been suggested.

Conclusion

The first and main result of this paper—that the global inhibition required for winner-take-all dynamics can be implemented by the diffuse thalamocortical loop—is not immediately obvious from neurophysiological observation, since there is little large-scale long-range inhibition within the loop. For this reason we have termed the process 'effective inhibition'. Secondly, we show how changing tonic input levels from the ascending arousal system to the thalamus can change the dynamical state in the cortex (Section Input Level from the Reticular Activating System Changed Dynamical State). This result explains previous observations concerning how AAS input affects cortical activity (such as state changes between sleep, wakefulness and vigilance)

and makes several novel predictions (Section Consequences for Brain Function). Thirdly, we show how cortical connectivity affects the sustained dynamics (Section Connectivity Impacts Dynamics). These sustained activity patterns can potentially be understood in terms of WTA dynamics, and we have demonstrated that the 'effective inhibition' paradigm, with its fundamentally different mechanism, is capable of supporting these patterns. When the network was structured into task-positive and DMN cortical regions with dense local and sparse long-range connections, we additionally showed how activity could intrinsically alternate between the groups based on chaotic dynamics with no extrinsic noise. We argue that such intrinsic alternation provides a plausible explanation for the dynamical segregation of the DMN from other cortical regions. We have shown that the network is robust to noise, but importantly does not require noise for the generation and maintenance of a complex, ongoing brain state.

Previous studies have examined networks that exhibited complex dynamics, but these networks were unable to achieve long range segregation of activity. In contrast, the CXC network achieves both segregation of activity and metastability through the global control of the RN acting through IL. Without global control, network activity reduces to numerous interconnected pockets of activity that can mutually integrate due to activity propagating through the long range connections, but cannot mutually segregate.

Acknowledgments

This work was supported by the Queensland Brain Institute, the Asia-Pacific Centre for Neuromodulation (a division of the University of Queensland Centre for Clinical Research), and an Australian Research Council (www.arc.gov.au) and National Health and Medical Research Council (www.nhmrc.gov.au) Thinking Systems Grant (Special Research Initiative: Thinking Systems; grant number: TS0669699). We thank our reviewers for their well-considered comments and suggestions which assisted us greatly.

References

Akrami, A., Russo, E., and Treves, A. (2012). Lateral thinking, from the Hopfield model to cortical dynamics. *Brain Res.* 1434, 4–16. doi: 10.1016/j.brainres.2011.07.030

Beim Graben, P., and Kurths, J. (2008). Simulating global properties of electroencephalograms with minimal random neural networks. *Neurocomputing* 71, 999–1007. doi: 10.1016/j.neucom.2007.02.007

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., and Cox, R. W. (1999). Conceptual processing during the conscious resting state: a functional MRI study. *J. Cogn. Neurosci.* 11, 80–93. doi: 10.1162/089892999563265

Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*. Berlin: Springer.

Breakspear, M., Terry, J. R., and Friston, K. J. (2003). Modulation of excitatory synaptic coupling facilitates synchronization and complex dynamics in a nonlinear model of neuronal dynamics. *Network* 52, 151–158. doi: 10.1016/s0925-2312(02)00740-3

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network. *Ann. N.Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., and Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8719–8724. doi: 10.1073/pnas.0900234106

Dorval, A. D. (2008). Probability distributions of the logarithm of inter-spike intervals yield accurate entropy estimates from small datasets. *J. Neurosci. Methods* 173, 129–139. doi: 10.1016/j.jneumeth.2008.05.013

Ermentrout, B. (1992). Complex dynamics in winner-take-all neural nets with slow inhibition. *Neural Netw.*5, 415–431. doi: 10.1016/0893-6080(92)90004-3

Falkner, A. L., Goldberg, M. E., and Krishna, B. S. (2013). Spatial representation and cognitive modulation of response variability in the lateral intraparietal area priority map. *J. Neurosci.* 33, 16117–16130. doi: 10.1523/JNEUROSCI.5269-12.2013

Fox, M. D., and Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Front. Syst. Neurosci.* 4:19. doi: 10.3389/fnsys.2010.00019

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011

Fries, P., Reynolds, J. H., Rorie, A. E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291, 1560–1563. doi: 10.1126/science.1055465

Fröhlich, F., Bazhenov, M., and Sejnowski, T. J. (2008). Pathological effect of homeostatic synaptic scaling on network dynamics in diseases of the cortex. *J. Neurosci.* 28, 1709–1720. doi: 10.1523/JNEUROSCI.4263-07.2008

Garcia-Rill, E. (1997). Disorders of the reticular activating system. *Med. Hypotheses* 49, 379–387. doi: 10.1016/S0306-9877(97)90083-9

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258. doi: 10.1073/pnas.0135058100

Groenewegen, H. J., and Berendse, H. W. (1994). The specificity of the 'nonspecific' midline and intralaminar thalamic nuclei. *Trends Neurosci.* 17, 52–57. doi: 10.1016/0166-2236(94)90074-4

Honey, C. J., Kötter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10240–10245. doi: 10.1073/pnas.0701519104

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Trans. Neural Netw.* 14, 1569–1572. doi: 10.1109/TNN.2003.820440

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Trans. Neural Netw.* 15, 1063–1070. doi: 10.1109/TNN.2004.832719

Jensen, O. (2006). Maintenance of multiple working memory items by temporal segmentation. *Neuroscience* 139, 237–249. doi: 10.1016/j.neuroscience.2005.06.004

Jensen, O., Kaiser, J., and Lachaux, J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends Neurosci.* 30, 317–324. doi: 10.1016/j.tins.2007.05.001

Jones, B. E. (2003). Arousal systems. *Front. Biosci.* 8, s438–s451. doi: 10.2741/1074

Kumar, A., Schrader, S., Aertsen, A., and Rotter, S. (2008). The high-conductance state of cortical networks. *Neural Comput.* 20, 1–43. doi: 10.1162/neco.2008.20.1.1

Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., et al. (2003). Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11053–11058. doi: 10.1073/pnas.1831638100

Lee, A. T., Vogt, D., Rubenstein, J. L., and Sohal, V. S. (2014). A class of GABAergic neurons in the prefrontal cortex sends long-range projections to the nucleus accumbens and elicits acute avoidance behavior. *J. Neurosci.* 34, 11519–11525. doi: 10.1523/JNEUROSCI.1157-14.2014

Levi, R., Varona, P., Arshavsky, Y. I., Rabinovich, M. I., and Selverston, A. I. (2005). The role of sensory network dynamics in generating a motor program. *J. Neurosci.* 25, 9807–9815. doi: 10.1523/JNEUROSCI.2249-05.2005

Levina, A., Herrmann, J. M., and Geisel, T. (2007). Dynamical synapses causing self-organized criticality in neural networks. *Nat. Phys.* 3, 857–860. doi: 10.1038/nphys758

Maass, W. (2000). On the computational power of winner-take-all. *Neural Comput.* 12, 2519–2535. doi: 10.1162/089976600300014827

Markram, H., Wang, Y., and Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5323–5328. doi: 10.1073/pnas.95.9.5323

Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., and Kurths, J. (2002). Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Phys. Rev. E* 66:026702. doi: 10.1103/PhysRevE.66.026702

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., and Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science* 315, 393–395. doi: 10.1126/science.1131295

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houdé, O., et al. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res. Bull.* 54, 287–298. doi: 10.1016/S0361-9230(00)00437-8

McDonald, C. T., and Burkhalter, A. (1993). Organization of long-range inhibitory connections with rat visual cortex. *J. Neurosci.* 13, 768–781.

Melzer, S., Michael, M., Caputi, A., Eliava, M., Fuchs, E. C., Whittington, M. A., et al. (2012). Long-range–projecting GABAergic neurons modulate inhibition in hippocampus and entorhinal cortex. *Science* 335, 1506–1510. doi: 10.1126/science.1217139

Moruzzi, G., and Magoun, H. W. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalogr. Clin. Neurophysiol.* 1, 455–473. doi: 10.1016/0013-4694(49)90219-9

Murphy, M., Riedner, B. A., Huber, R., Massimini, M., Ferrarelli, F., and Tononi, G. (2009). Source modeling sleep slow waves. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1608–1613. doi: 10.1073/pnas.0807933106

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

Oken, B. S., Salinsky, M. C., and Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin. Neurophysiol.* 117, 1885–1901. doi: 10.1016/j.clinph.2006.01.017

Pecevski, D., Natschläger, T., and Schuch, K. (2009). PCSIM: a parallel simulation environment for neural circuits fully integrated with Python. *Front. Neuroinform.* 3:11. doi: 10.3389/neuro.11.011.2009

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* 5, 805–811. doi: 10.1038/nn890

Rabinovich, M. I., Afraimovich, V. S., Bick, C., and Varona, P. (2012). Information flow dynamics in the brain. *Phys. Life Rev.* 9, 51–73. doi: 10.1016/j.plrev.2011.11.002

Raichle, M. E., and Mintun, M. A. (2006). Brain work and brain imaging. *Annu. Rev. Neurosci.* 29, 449–476. doi: 10.1146/annurev.neuro.29.051605.112819

Reese, N. B., Garcia-Rill, E., and Skinner, R. D. (1995). The pedunculopontine nucleus - auditory input, arousal and pathophysiology. *Prog. Neurobiol.* 47, 105–133. doi: 10.1016/0301-0082(95)00023-O

Riecke, H., Roxin, A., Madruga, S., and Solla, S. A. (2007). Multiple attractors, long chaotic transients, and failure in small-world networks of excitable neurons. *Chaos* 17:026110. doi: 10.1063/1.2743611

Roxin, A., Riecke, H., and Solla, S. A. (2004). Self-sustained activity in a small-world network of excitable neurons. *Phys. Rev. Lett.* 92:198101. doi: 10.1103/PhysRevLett.92.198101

Rubinov, M., Sporns, O., Thivierge, J.-P., and Breakspear, M. (2011). Neurobiologically realistic determinants of self-organized criticality in networks of spiking neurons. *PLoS Comput. Biol.* 7:e1002038. doi: 10.1371/journal.pcbi.1002038

Saper, C. B., Chou, T. C., and Scammell, T. E. (2001). The sleep switch: hypothalamic control of sleep and wakefulness. *Trends Neurosci.* 24, 726–731. doi: 10.1016/S0166-2236(00)02002-6

Schnitzler, A., and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nat. Rev. Neurosci.* 6, 285–296. doi: 10.1038/nrn1650

Shanahan, M. (2008). Dynamical complexity in small-world networks of spiking neurons. *Phys. Rev. E* 78:041924. doi: 10.1103/PhysRevE.78.041924

Shannon, C. E., Weaver, W., Blahut, R. E., and Hajek, B. (1949). *The Mathematical Theory of Communication.* Champaign, IL: University of Illinois Press.

Sporns, O., and Tononi, G. (2001). Classes of network connectivity and dynamics. *Complexity* 7, 28–38. doi: 10.1002/cplx.10015

Sporns, O., Tononi, G., and Edelman, G. M. (2000a). Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. *Neural Netw.* 13, 909–922. doi: 10.1016/S0893-6080(00)00053-8

Sporns, O., Tononi, G., and Edelman, G. M. (2000b). Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cereb. Cortex* 10, 127–141. doi: 10.1093/cercor/10.2.127

Stam, C. J., Pijn, J. P. M., Suffczynski, P., and Lopes da Silva, F. H. (1999). Dynamics of the human alpha rhythm: evidence for non-linearity? *Clin. Neurophysiol.* 110, 1801–1813. doi: 10.1016/S1388-2457(99)00099-1

Steriade, M. M., and McCarley, R. W. (1990). *Brain Control of Wakefulness and Sleep.* New York, NY: Springer.

Stratton, P., and Wiles, J. (2010a). "Complex spiking models: a role for diffuse thalamic projections in complex cortical activity," in *Neural Information Processing. Theory and Algorithms*, eds K. W. Wong, B. S. U. Mendis, and A. Bouzerdoum (Berlin; Heidelberg: Springer), 41–48. doi: 10.1007/978-3-642-17537-4_6

Stratton, P., and Wiles, J. (2010b). Self-sustained non-periodic activity in networks of spiking neurons: the contribution of local and long-range connections and dynamic synapses. *Neuroimage* 52, 1070–1079. doi: 10.1016/j.neuroimage.2010.01.027

Taylor, J. G., and Farrukh, N. A. (1996). "A basis for long-range inhibition across cortex," in *Lateral Interactions in the Cortex: Structure and Function,* eds J. Sirosh, R. Miikkilainen, and Y. Choe (Austin, TX: UTCS Neural Networks Research Group). Available online at: http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96/taylor/

Thiele, A. (2012). NMDA receptors figure it out. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10749–10750. doi: 10.1073/pnas.1207975109

Tognoli, E., and Kelso, J. (2014). The metastable brain. *Neuron* 81, 35–48. doi: 10.1016/j.neuron.2013.12.022

Tomasi, D., and Volkow, N. D. (2011). Association between functional connectivity hubs and brain networks. *Cereb. Cortex* 21, 2003–2013. doi: 10.1093/cercor/bhq268

Uddin, L. Q., Kelly, A. M., Biswal, B. B., Castellanos, F., and Milham, M. P. (2008). Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Hum. Brain Mapp.* 30, 625–637. doi: 10.1002/hbm.20531

Uhlhaas, P. J., and Singer, W. (2006). Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron* 52, 155–168. doi: 10.1016/j.neuron.2006.09.020

Van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1725. doi: 10.1126/science.274.5293.1724

Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550

Wang, S.-J., Hilgetag, C. C., and Zhou, C. (2011). Sustained activity in hierarchical modular neural networks: self-organized criticality and oscillations. *Front. Comput. Neurosci.* 5:30. doi: 10.3389/fncom.2011.00030

Wang, X. J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J. Neurosci.* 19, 9587–9603.

Zhang, D., and Raichle, M. E. (2010). Disease and the brain's dark energy. *Nat. Rev. Neurol.* 6, 15–28. doi: 10.1038/nrneurol.2009.198

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.

frontiers in Systems Neuroscience

ORIGINAL RESEARCH
published: 23 September 2016
doi: 10.3389/fnsys.2016.00076

Extended Neural Metastability in an Embodied Model of Sensorimotor Coupling

Miguel Aguilera[1, 2, 3]*, *Manuel G. Bedia*[1, 3] *and Xabier E. Barandiaran*[4, 5]

[1] *Department of Computer Science and Systems Engineering, University of Zaragoza, Zaragoza, Spain,* [2] *Department of Psychology, University of the Balearic Islands, Palma de Mallorca, Spain,* [3] *ISAAC Lab, Aragon Institute of Engineering Research, University of Zaragoza, Zaragoza, Spain,* [4] *Department of Philosophy, University School of Social Work, University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain,* [5] *Department of Logic and Philosophy of Science, IAS-Research Center for Life, Mind and Society, University of the Basque Country (UPV/EHU), Donostia - San Sebastián, Spain*

The hypothesis that brain organization is based on mechanisms of metastable synchronization in neural assemblies has been popularized during the last decades of neuroscientific research. Nevertheless, the role of body and environment for understanding the functioning of metastable assemblies is frequently dismissed. The main goal of this paper is to investigate the contribution of sensorimotor coupling to neural and behavioral metastability using a minimal computational model of plastic neural ensembles embedded in a robotic agent in a behavioral preference task. Our hypothesis is that, under some conditions, the metastability of the system is not restricted to the brain but extends to the system composed by the interaction of brain, body and environment. We test this idea, comparing an agent in continuous interaction with its environment in a task demanding behavioral flexibility with an equivalent model from the point of view of "internalist neuroscience." A statistical characterization of our model and tools from information theory allow us to show how (1) the bidirectional coupling between agent and environment brings the system closer to a regime of criticality and triggers the emergence of additional metastable states which are not found in the brain in isolation but extended to the whole system of sensorimotor interaction, (2) the synaptic plasticity of the agent is fundamental to sustain open structures in the neural controller of the agent flexibly engaging and disengaging different behavioral patterns that sustain sensorimotor metastable states, and (3) these extended metastable states emerge when the agent generates an asymmetrical circular loop of causal interaction with its environment, in which the agent responds to variability of the environment at fast timescales while acting over the environment at slow timescales, suggesting the constitution of the agent as an autonomous entity actively modulating its sensorimotor coupling with the world. We conclude with a reflection about how our results contribute in a more general way to current progress in neuroscientific research.

Keywords: neural assemblies, metastability, criticality, synaptic plasticity, embodied cognition, sensorimotor coupling, evolutionary robotics

OPEN ACCESS

Edited by:
Emili Balaguer-Ballester, Bournemouth University, UK

Reviewed by:
Paul Miller, Brandeis University, USA
Heiko J. Luhmann, University of Mainz, Germany

***Correspondence:**
Miguel Aguilera
sci@maguilera.net

Received: *23 May 2016*
Accepted: *31 August 2016*
Published: *23 September 2016*

Citation:
Aguilera M, Bedia MG and Barandiaran XE (2016) Extended Neural Metastability in an Embodied Model of Sensorimotor Coupling. Front. Syst. Neurosci. 10:76. doi: 10.3389/fnsys.2016.00076

1. INTRODUCTION

Generally, neurodynamic approaches have focused in understanding what kind of neural organization is necessary to cope with the requirements of an external world. Assuming that the brain is subject to demanding conditions from its environment, the challenge is to explain what type of neural computation or what form of organization of neural spatiotemporal patterns might be capable of satisfying the requirements for adaptive, conscious, cognitive activity. This has led to progress in the definition of a framework able to account for the brain's ability to display a rich set of meaningful behaviors. Nowadays, a popular view in neuroscience holds that the human brain is structured into a large number of areas in which information is highly segregated into local clusters and, at the same time, functionally integrated (Damasio, 1989; Varela, 1995; Tononi and Edelman, 1998). One of the most plausible mechanisms hypothesized to be behind this equilibrium between integration and segregation is metastable phase locking between neural assemblies over multiple frequency bands. This mechanism has been proposed to explain how the brain flexibly enters and exits coherent spatiotemporal patterns of neural activity (Kelso, 1995; Varela et al., 2001; Le Van Quyen, 2011). Subsequently, the notion of metastable neural assemblies as building blocks of brain organization has become relatively widespread in large-scale neuroscience studies (e.g., Werner, 2007a; Buzsáki, 2010; Edelman et al., 2011; Ward, 2011).

Nevertheless, when analysing and modeling brain organization, a crucial aspect of cognitive dynamics is frequently neglected: the sensorimotor coordination that continuously feeds back into brain dynamics (from saccadic eye movements to proprioception; from perception to action; O'Regan and Noë, 2001; Aguilera et al., 2013; Engel et al., 2013). Mental processes such as perception, emotion or intention are not limited to neural processes inside the brain, but produced through a flexible integration of the dynamics of brain, body and environment in a distributed manner. Hypotheses addressing this issue propose that brain organization consists in a plastic system of open loops developed in the process of life and closed to full functional cycles in every interaction with the environment (Fuchs, 2011), being the role of the central nervous system to transform and diversify these loops. In addition, it has been proposed that the behavior neural tissue in isolation might be restricted to little more than exhibiting spontaneous synchronization and other behaviors common to nonlinear dynamical systems, and the brain may operate as a *metastable circuit breaker* flexibly switching between different dynamic fields of agent-environment engagement (Dotov, 2014).

Furthermore, enactive approaches to neurodynamics have proposed that the formation and dissolution of neural assemblies in the brain must be embedded in sensorimotor regulatory cycles, producing the emergence of global organism-environment processes, which in turn affect their constituent elements (Thompson and Varela, 2001; Varela and Thompson, 2003; Di Paolo et al., 2016). One of the central contributions to this issue has been the notion of an operational closure of the nervous system (Varela, 1997; Di Paolo and Thompson, 2014), illustrated in **Figure 1**. Operational closure implies a circular regulation in which the coordinated activity of the neural system gives rise to the emergence of neural ensembles (or "cell assemblies"; Hebb, 1952), driving the behavior of the organism, which in turn generates a sensory input into the neural system closing a double regulatory loop. According to Francisco Varela,

> The nervous system is organized by the operational closure of a network of reciprocally related modular subnetworks giving rise to ensembles of coherent activity such that: (i) they continuously mediate invariant patterns of sensorymotor correlation of the sensory and effector surfaces; (ii) give rise to a behavior for the total organism as a mobile unit in space. The operational closure of the nervous system then brings forth a specific mode of coherence, which is embedded in the organism. This coherence is a cognitive self: a unit of perception/motion in space, sensorymotor invariances mediated through the interneuron network (Varela, 1992, p.10).

Nevertheless, there are a lack of good models or precise characterizations of the operational closure of the nervous systems and the kind of interaction that takes place between neural, bodily and sensorimotor cycles (Barandiaran, 2016). Thus, it is far from clear how to characterize this sensorimotor specific form of coherence and how is it constitutive of cognitive activity. Clarifying this issue is of fundamental importance for embodied neurodynamic views in order to propose solid explanatory alternatives to internalist perspectives of brain organization.

In the case of metastability in cognitive processes, we know that metastable behavior is not restricted to the brain, but also extends to behavioral patterns (Kelso, 1995; Kelso et al., 1995). Nevertheless it is not clear what the relation is between behavioral metastability and underlying neural metastability. Consider, for example, the case of perceiving an ambiguous image (e.g., an image perceived either as a face or as meaningless

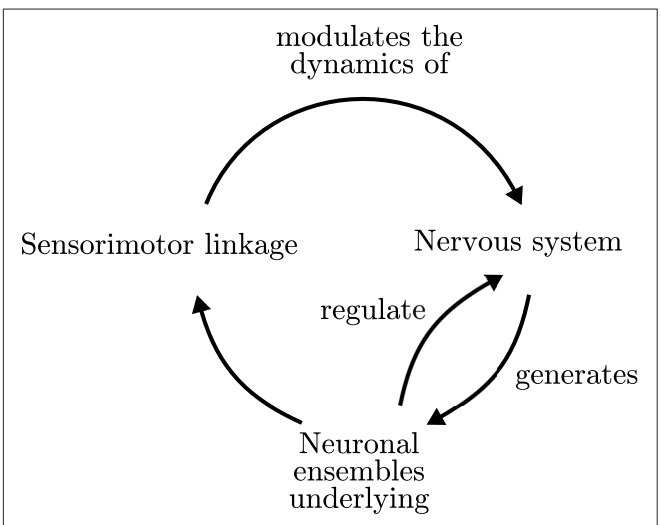


FIGURE 1 | Organizational closure of the nervous system. Adapted from Varela (1997).

shapes, or a Necker cube which can be alternatively seen as oriented in two different positions) in which perception can alternatively and spontaneously switch from one mode to another. If we take a look at the brain we can observe the emergence of transient assemblies of neural synchronization when one mode of perception arises (Rodriguez et al., 1999), and at the same time if we were to analyse the switching of perceptive patterns the subject is engaged in we can observe signatures of metastable behavior (Kelso et al., 1995, Section 3). This raises questions around whether long term behavioral metastability is the product of a direct mapping of intrinsic metastable states of the brain into behavior or, conversely, metastability of both the brain and behavior is a product of the whole brain-body-environment system coupled by sensorimotor processes (the interaction of neural dynamics, retinal activation, patterns of saccadic eye movements, etc.). These are difficult questions, requiring comparison between processes taking place at different scales (neurodynamic, sensorimotor, conductual, etc.) that pose important experimental difficulties. In the past, evolutionary robotics has been a highly productive tool for finding non-intuitive solutions to complex problems and understanding many-to-many relations between different scales of behavior (Nolfi and Floreano, 2000; Harvey et al., 2005). In the same tradition, in this paper we present an artificially evolved agent to explore the relation between neural and behavioral metastability in a simple bistable task. We choose a phototactic task[1] in which the agent alternatively develops a preference between two types of light (e.g., two different colors) as an example of a simple task involving metastable neural and behavioral dynamics.

Our hypothesis is that the slow modulation of synaptic plasticity over the sensorimotor coupling increases and sustains the metastability of neuro-behavioral integrated states in a manner that cannot be reduced to the dynamics of the brain in isolation, nor the brain receiving a structured input. These integrated metastable states are associated with specific modes of coherence in neural structures when they are engaged in bodily and environmental processes, as the brain-body-environment system becomes an operationally closed entity. The specificity of these modes of coherence is hypothesized to be related to the particularities of autonomous agency and the operational closure of the nervous system, i.e., the characteristics that allow us to describe an agent as an individual entity albeit in continuous interaction with its environment: such as the self-constitution of the agent as an entity or an asymmetrical interaction with its world in which it actively modulates its own sensorimotor coupling (see Barandiaran et al., 2009). Here, we propose a minimalistic approach to address some of the difficult questions arising from these ideas. We introduce a robotic model equipped with just three oscillatory units and synaptic plasticity in their connections. Because of its simplicity, this approach allows us to tackle the problem with a system about which we have complete knowledge and that is tractable using dynamical systems techniques of analysis.

In the following sections we first introduce the robotic agent and an artificial evolution process to obtain a model displaying metastable behavioral patterns in a bistable phototactic task. Then, we describe our methodology for analysing the role of the sensorimotor loop in the generation of metastable behavioral patterns, combining (1) the comparison of a *situated* agent interacting directly with its environment and a *passively-coupled* agent which is fed a signal identical to the one received by the situated agent, but it cannot influence its environment, and (2) a statistical description of the states of the agent and the environment together with the use of different tools from information theory to quantify the metastability in its behavior and the interaction between different scales of description of the robot (oscillatory activity, synaptic plasticity and behavioral patterns). In this framework we perform experiments showing that (1) the bidirectional coupling between agent and environment brings the system closer to a regime of criticality and triggers the emergence of additional metastable states, which are not found in the brain in isolation but extended to the whole system of sensorimotor interaction, (2) the synaptic plasticity of the agent is key to sustain open structures in the neural controller of the agent flexibly engaging and disengaging different behavioral patterns that sustain sensorimotor metastable states, and (3) this creates an asymmetrical circular loop of interaction between agent and environment, in which the agent is able to respond to variability of the environment at small timescales, while acting over the environment at large timescales. We conclude that metastability of neural dynamics can be extended to sensorimotor metastable states and that, in our model, this takes place when the agent establishes a specific circular relation with its environment, suggesting that the extension of metastable dynamics from the brain to interactive behavioral patterns is connected with specific forms of engagement with the world characteristic in autonomous agency.

2. METHODS

As we proposed above, our goal is to explore the relation between metastability in brain dynamics and behavior in a robotic model in order to test the hypothesis that some behavioral metastable states cannot be reduced to brain dynamics alone and are instead the product of an integration of brain, bodily and environmental dynamics. In order to do this, we design a model with the ability of presenting flexibility in both neural and behavioral patterns (which will be evolved using a genetic algorithm in order to reduce the constraints imposed onto the model) and we propose a framework of analysis allowing us to characterize metastable states and relations between components of the model, as well as a comparison of the behavior of the brain with and without the effect of the sensorimotor loop in equivalent conditions.

[1]Phototaxis implies climbing a light gradient. This task has been chosen because gradient climbing is a minimal task, which is widespread in nature. Many small scale adaptive behavior occurs along chemical gradients, and the microscopic world is full of gradients (like thermal gradients or light gradients but mostly chemical gradients). The adaptive behavior of small animals (e.g., C. elegans) and individual motile cells (e.g., bacteria but also animal cells migrating during development) is mostly a gradient-related adaptive behavior. Navigating smell or heat gradients are also stereotypical adaptive tasks for higher animals. Moreover many instances of higher-level behavior can also be interpreted as abstract gradient climbing (e.g., a human can move up a gradient of social popularity or economic wealth involving complex strategic decisions).

2.1. Model of a Neurodynamic Controller with Relational Homeostasis Embedded in a Robotic Agent

We propose a model of homeostatic adaptation inspired by previous work in evolutionary robotics (Iizuka and Di Paolo, 2007; Di Paolo and Iizuka, 2008), defining an adaptive mobile agent controlled by a plastic oscillatory neural system. This model is not intended to represent the activity of individual neurons but, more generally, to capture the dynamics of neural oscillations at a mesoscopic level, where integration mechanisms are hypothesized to be based on phase synchronization processes between neuronal groups (Varela et al., 2001), thus representing large-scale synchronization of brain regions that are anatomically far apart. Since the model is described in detail in Aguilera et al. (2015), we provide here a brief description.

The agent incorporates a neural controller defined as a fully connected Kuramoto network (Acebrón et al., 2005) with three units defined as:

$$\dot{\theta}_i = \omega_i + I_i + \sum_{j=1}^{N} K_{ij} \cdot sin(\theta_j - \theta_i) \quad (1)$$

where θ_i represents the phase of oscillator i, ω_i is its natural frequency (range $[0, 5]$), K_{ij} is the coupling strength between oscillators i and j, and I_i represents the sensory inputs. The behavior of the neural controller is modulated by plastic mechanisms preserving phase relational invariances of the system, defined as:

$$\delta\dot{K}_{ij} = \eta_{ij} \cdot p(\Phi_i - \Phi_i^0) \cdot sin((\theta_j - \theta_i) - \Phi_i^0) \quad (2)$$

where δK_{ij} are the connection weights, η_{ij} is the rate of plastic change (range $[0, 0.9]$) of each connection, and Φ represents the phase difference of oscillator i with respect to the sum of the oscillators connected to it weighted by the strength of their connections:

$$\Phi_i = \angle(\sum_{j=1}^{N} K_{ij} \cdot e^{\mathrm{i}(\theta_j - \theta_i)}) \quad (3)$$

where $\angle$ denotes the phase of a complex value and i is the imaginary unit. Φ_i^0 (range $[\frac{-\pi}{2}, \frac{\pi}{2}]$) stands for the preferred phase relation of the oscillatory node. Finally, the function $p(x)$ determines the level of plastic change for all incoming weights of a node, which is activated when the value of Φ_i is far from Φ_i^0 (**Figure 2B**). When plastic changes take place, connection strengths change following a continuous non-monotonic function $K_{ij} = \alpha \cdot F(\delta K_{ij})$ (**Figure 2C**) designed to explore the full configuration space[2], where α is a constant (range $[0, 5]$) that regulates the coupling strength.

In short, the model works under the assumption that large-scale neural oscillatory components try to maintain a preferred phase relation with respect to other oscillatory components by means of plastically regulating the strength of their connectivity. The model is designed to present the possibility of metastable behavior at different states. Kuramoto oscillator networks can display metastable states when the connection strengths are below a critical point of complete synchronization. Also, we defined the evolution of synaptic plasticity in such a way that the

[2]This continuous, non-monotonic change of synaptic strengths was introduced by Iizuka and Di Paolo (2008) as a simple mechanism to avoid weight saturation of Hebbian-like rules. This way, eschewing biological plausibility allows us to make sure that the robots is able to explore its whole range of weight configurations.

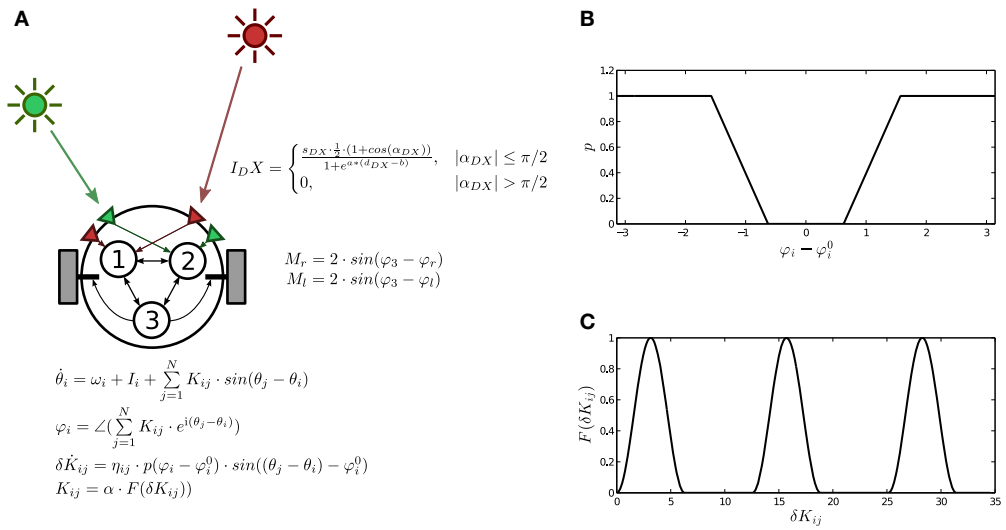


FIGURE 2 | The robotic agent with three plastic oscillatory units. (A) Schema of the agent, the environment, sensors and motors, and the neural controller. **(B)** Plastic function $p(\Phi_i - \Phi_i^0)$, in which plasticity depends on the difference between the weighted phase relation Φ_i of the neural oscillator i with respect to other oscillators and the preferred weighted phase relation Φ_i^0. **(C)** Mapping function $F(\delta K_{ij})$ which transforms weight values δK_{ij} into the actual value of coupling strengths between oscillators K_{ij}. Reproduced from Aguilera et al. (2015).

agent can potentially explore all different available possibilities of engagement with the environment, and furthermore be able to exploit forms of behavioral metastability.

The agent is modeled as a simulated wheeled robot with a circular body of radius 4 and two diametrically opposed motors (**Figure 2A**), driving the agent backwards and forwards. The agent's mass is assumed to be small enough for inertial resistance to be negligible, thus its translational speed is calculated as the vectorial average of the motor velocities, and the angular speed as the difference of the motor velocities divided by the body diameter. Motor outputs are calculated from the phase relation Φ_3 of the effector oscillator, with a gain parameter of value 2:

$$\begin{aligned} M_r &= 2 \cdot sin\,(\Phi_3 - \Phi_r) \\ M_l &= 2 \cdot sin\,(\Phi_3 - \Phi_l) \end{aligned} \tag{4}$$

where Φ_r and Φ_l (range $[0, 2\pi]$) are bias terms which map the motor output into the actual motor activation.

The agent has two pairs of sensors (right and left) for each of the different light sources A and B. Each sensor points to a direction at $\pi/3$ radians from the forward direction. Light A sensors are connected to oscillator 1 and light B sensors are connected to oscillator 2. The effects of both the angle and the distance on the sensor activation are represented by the following function:

$$I_{DX} = \begin{cases} \frac{s_{DX} \cdot 0.5 \cdot (1 + cos(\alpha_{DX}))}{1 + e^{a \cdot (d_{DX} - b)}}, & |\alpha_{DX}| \leq \pi/2 \\ 0, & |\alpha_{DX}| > \pi/2 \end{cases} \tag{5}$$

where X can represent either light A or B, D stands for either right or left sensor, α_{DX} is the angle of sensor DX to light X, d_{DX} is the distance between sensor DX and light X, and a and b have the arbitrary values of 0.03 and 100 respectively. The light intensity received at each sensor is multiplied by a gain parameter s_{DX} (range $[-8, 8]$), feeding the resulting value to the corresponding oscillator's input I_i. A full schema of the robot is represented in **Figure 2A**.

All parameter values (except where otherwise specified) are fixed by a genetic algorithm within the indicated range. A population of 20 agents is evolved using a rank-based genetic algorithm with elitism. Each of the agent parameters $\omega_i, s_{DX}, \alpha, \eta_{ij}, \Phi_r, \Phi_l$ and Φ_i^0 is encoded into a 5 bits string representing real numbers uniformly distributed within the specified range. For each generation, the best 4 agents (20% of the population) pass to the next generation without change. For the remaining slots, pairs of individuals are selected for crossover with a probability proportional to their fitness value, and new individuals are created mixing their genes (bit series) by adding a mutation probability of 3% for each gene.

The agents are evolved for displaying a metastable behavior in which the agent has to develop switching preferences toward two different types of light. That is, there is an environment with two types of light (e.g., two colors) and we want the agent to develop a preference toward one of them (e.g., repeatedly interacting with it) while being able to switch its preference to the other light depending on its internal configuration. This behavior is chosen because it demands the agent to present a robust phototactic behavior while at the same time presenting flexibility in the creation and dissolution of behavioral preferences. An evaluation procedure is proposed for the genetic algorithm in order to accomplish this objective, consisting in four different tasks designed by Iizuka and Di Paolo (2007): a single light A, a single light B, one light A and a blinking light B, one light B and a blinking light A (blinking lights illuminate with a probability of 0.15 for each time step). The agent gains fitness by approaching the non-blinking light. The objective of this configuration is to create a "dummy" that encourages the agent to learn to ignore one of the lights while approaching the other. Lights appear at a random distance, [100, 150]. When two lights are present, they appear, from the agent's point of view, with a random separation within the range $[\pi/2, 3\pi/2]$. The length of each trial is 125 s.

Each individual agent is tested for 12 independent runs (3 for each of the 4 tasks) consisting of a series of trials where a light or a pair of lights are presented to the agents for a fixed time. Synaptic weights δK_{ij} are reset to initial random values before each run. Each run consist of 8 trials in which the agent is presented with one or two lights for a specified time. Only the last 3 trials of each run are evaluated in order not to penalize slow plastic changes and bootstrap evolution. All simulations are run with an Euler step of 0.1.

Fitness for each trial is calculated in three terms, $F_{trial} = (F_D + F_p) \cdot F_H$. $F_D = 1 - d_f/d_i$, where d_f and d_i respectively correspond to the final and initial distances to the target light. F_p is equal to the proportion of time that the agent spends within a distance of less than 4 times its body radius (i.e., a distance of 16) to the target light during a trial. F_H represents the mean level of homeostasis in the system, computing the mean degree of homeostasis $\frac{1}{3}\sum_i(1 - p(\Phi_i - \Phi_i^0))$ (i.e., 1 minus the level of plasticity) for each oscillator. The genetic algorithm is run for 500 generations, reaching a stable level of fitness around 0.45. The best performing agent from the last generation (which is able to reach both lights ignoring the dummy) is selected. The code simulating the behavior of the agent and the parameters obtained from the genetic algorithm can be accessed from the following repository https://github.com/IsaacLab/HNA-robotic-model/tree/master/minimal-preference-task.

2.2. Conceptual Setup for Testing Sensorimotor Integration: Situated vs. Passively-Coupled Agents in An Open Environment

In order to explore how the agent exploits internal and sensorimotor metastability we simulate the agent in an open environment (which was never experienced during training) in which the agent can develop sustained preferences toward the two types of light. In this case, the two lights are always present with equal intensities, and a new pair of lights is generated periodically after a given time (starting a new trial). The best performing agent from the last generation of the genetic algorithm agent is simulated in a virtual environment which presents a series of pairs of lights, giving the agent a time of 1250 steps to choose and approach one of them. As analysed elsewhere (Aguilera et al., 2015), the agent is able to develop

stable preferences toward one of the lights, maintaining it for several trials until the preference is changed. The switching of preferences depends on the long-term interaction between internal plastic mechanisms and the encountered configurations of the environment. Different neural cell assemblies arise connected with particular patterns of behavior of the agent, and at slower timescales synaptic plasticity modulates the emergence and dissolution of these behavioral patterns. A video of the behavior of the agent (including plastic mechanisms) can be found at https://vimeo.com/53847420.

Once defined the agent in which we want to explore the emergence of metastable behavioral patterns, we propose a sensorimotor null model to be compared with our model in order to test the influence of the sensorimotor loop in the generation of metastable states in the agent. To this end, we propose an agent which maintains the structure of the received input but presents a disrupted coordination with its environment. Thus, we will compare:

1. A *situated* agent with normal sensorimotor interaction.
2. A *passively-coupled* agent, in which the input fed to the agent is recorded from the behavior of the situated agent. Thus, in the passively-coupled agent the received input is decoupled from the activation of the motors but it maintains an structure as if it was generated by real interaction.

With this comparison we can detect the effects that, despite being the result of sensorimotor coordination, cannot be replicated just by using an input with an adequate structure. This is a subtle difference, but if genuine sensorimotor coordination is constitutive of a cognitive process, the same process should not take place when the agent is passively processing an input with the rig as an input-structured process.

2.3. Discretization and Probability Density Function of the System

In the experimental setup defined above, we will use information theory tools in order to get a better understanding of how the different elements of the neural controller and the agent's behavior interact, using a symbolic representation of the system states. Understanding the coordination between neural ensembles and sensorimotor activity is not trivial due to the moderately high dimensionality of the system (9 dimensions of the neural controller, plus the dimensions of body and environment). Nevertheless, the system can be simplified by reducing both the state of neural ensembles and synapse configurations to discrete values representing the state of a network, plus a binary variable representing the behavior of the agent (reaching one light or the other).

To simplify the analysis, we are interested in a description of the system minimizing the number of symbols needed to describe the states of the robot, while maintaining the properties of the system. In this case, we find that a binary discretization is a good choice for describing the system. In order to ascertain a good discretization of the system, we test the validity of different possible discretizations by comparing the dimensionality of the original and discretized data through principal component analysis on their covariance matrix. We use the following definition of dimensionality (Abbott et al., 2009):

$$d = \Big(\sum_{i=1}^{N} \widetilde{\lambda}_i^2 \Big)^{-1} \tag{6}$$

where $\widetilde{\lambda}_i$ are the normalized eigenvalues of the covariance matrix expressing the fraction of the variance explained by the corresponding principal component. We find that different discretizations generally increase the dimensionality of the system (due to the introduction of noise in the form of discretization error). In **Figure 3** we can observe how the dimensionality of the discretized systems departs from the original dimensionality depending on the number of bins employed. Specifically, we find that the choice of a binary description with just two bins is a particularly good description of the covariances of the system, increasing the dimensionality of the system by just 1.32% of the original dimensionality. Therefore, for the rest of the paper we will employ a binary description of the variables of the system as described below.

In previous work (Aguilera et al., 2015) we have determined that the configuration of the oscillators in the assembly is the relevant variable for generating one or other type of behavior. Therefore, we define cell assemblies depending on the relative phase of the oscillators with respect to the mean phase of the Kuramoto network. We codify the state of the cell assembly with a string of three bits Θ:

$$\Theta = \{\Theta_1, \Theta_2, \Theta_3\}, \text{ where } \Theta_i = \begin{cases} 1, & \text{if } sin\,(\theta_i - \theta_m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $\theta_m = \angle(\frac{1}{3}\sum_i e^{i\theta_i})$ is the mean phase of the system at a particular instant.

Analogously to cell assemblies of neurons, we may consider a constellation of changing synaptic weights as an assembly

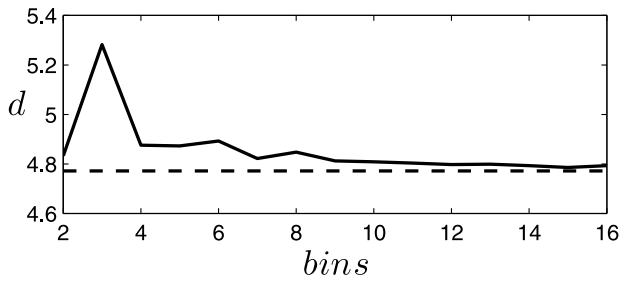


FIGURE 3 | Increment of dimensionality of the discretized system for different bin choices. The dimensionality of the data generated by the situated agent (the passively-coupled agent yields similar results) with continuous (dashed line) or discrete (continuous line) values using different numbers of equally spaced bins in the discretization. We observe how a binary description (2 bins) is a good discretization of the system, presenting a small increase in dimensionality only matched by discretizations with 7 or more bins. A discretization with 3 bins is particularly unfit since it cannot capture the covariance of small fluctuations around the central bin.

of synapses or "synapsemble" (Buzsáki, 2010, p. 372). Synapse assemblies have been hypothesized to be critical for building up and dissolving cell assemblies and linking together sequences of cell assemblies. To codify the activity of synapse assemblies, we define each synapse as active or inactive if the value of the synapse is higher[3] than the mean value of the synaptic strengths K_m, where $K_m = \frac{1}{6}\frac{1}{T}\sum_t \sum_{i,j} K_{i,j}(t)$. The state of the synaptic assembly Ψ is codified with a string of 6 bits:

$$\Psi = \{\Psi_{1,2}, \Psi_{2,1}, \Psi_{1,3}, \Psi_{3,1}, \Psi_{2,3}, \Psi_{3,2}\}, \text{ where}$$
$$\Psi_{i,j} = \begin{cases} 1, & \text{if } K_{ij} > \langle K_m \rangle \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

Similarly, for each trial we define a variable Λ which represents the behavioral pattern of the agent (i.e., what light it reaches):

$$\Lambda = \begin{cases} 1, & \text{if } d_{f,A} < d_{f,B} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where $d_{f,A}$ and $d_{f,B}$ are the final distances to each type of light at the end of the trial.

In summary, we define whether a particular cell or synapse ensemble is active by using a set of binary variables $s = \{\Theta, \Psi, \Lambda\}$ which represent if a specific ensemble is active at a particular moment of time and what behavioral pattern is being developed by the agent. All possible relations between oscillators give rise to 6 possible states for the cell assemblies, and all possible activated synapses give rise to 64 possible combinations or synapse assemblies, giving us a complete and discrete definition of the system that we can use to apply information theory tools. In order to do so 100 similar agents with random initial conditions are simulated for the situated and passively-coupled cases for 1000 trials with a duration of 1250 steps, generating a time series with $1.25 \cdot 10^6$ states. Sections 3.2 and 3.3 use one of these time series to compute mutual information through time and transfer entropy, although the other 99 series yield practically identical results. The calculations in Section 3.1 require us to accurately compute the whole probability density function of the system. Thus, in order to avoid correlations in our sampling we compose 100 series of 10^6 states, consisting in 10^4 random states extracted from each of the initial 100 time series. Each sample can provide an estimation of the frequency of the 2^{10} states of s, inferring the probability density function of the system, which we compute as $P(s) = n_s/n_T$, where n_s is the number of occurrences of state s, n_T is the total length of the sample.

2.4. Information Theory Tools

Having defined these variables of the system in a discrete manner, we can use information theory tools to determine the relation between variables. In our model, we can use these measures to quantify the relation between the state of different elements in the neural controller of the agent, or between such components and features of the environment surrounding the agent. The information contained in a random variable is quantified in terms of *entropy*, which is defined as:

$$H(X) = -\sum_{x \in X} P(x) log(P(x)) \quad (10)$$

where X is the set of states of the variable and $P(X)$ its density probability function.

A useful measure to compare two variables is the relative entropy or Kullback-Leibler divergence between their statistical distributions, which is a measure of the difference between two probability distributions X and Y. It is defined as:

$$D(X; Y) = \sum_{x \in X} \sum_{y \in Y} P(x) log \frac{P(x)}{P(y)} \quad (11)$$

Given a pair of variables X, Y and their marginal distributions the Kullback-Leibler divergence can be used to capture the information shared between the variables, defined as their *mutual information*:

$$I(X; Y) = H(Y) - H(Y|X) = \sum_{x \in X} \sum_{y \in Y} P(x, y)\, log\, \frac{P(x, y)}{P(x)P(y)} \quad (12)$$

By definition, $I(X; Y) = I(Y; X)$, thus mutual information cannot describe relations of causality. Instead, transfer entropy measures are typically employed to analyse causal relationships between variables. The decrease of uncertainty in the state of a variable derived from the past history of other variables is defined as the *transfer entropy* between two variables:

$$TE(X \rightarrow Y) = H(Y_{t+\tau}|Y_t^{(d')}) - H(Y_{t+\tau}|Y_t^{(d')}, X_t^{(d)}) =$$
$$= \sum_{x_{t+\tau}, x_t \in X} \sum_{y_t \in Y} P(x_{t+\tau}, x_t^{(d)}, y_t^{(d')}) log \frac{P(x_{t+\tau}, x_t^{(d)}, y_t^{(d')}) P(x_t^{(d)})}{P(x_{t+\tau}, x_t^{(d)}) P(x_t^{(d)}, y_t^{(d')})} \quad (13)$$

where $X_t^{(d)}$ denotes the past history of x counted from time t and length d (i.e., x_t, x_{t-1}, ..., x_{t-d}).

3. RESULTS

In this section we present results comparing a situated and a passively-coupled agent using a discrete description of the system and information theory tools. We first show that when the system is coupled to its environment it is brought closer to a regime of criticality and that the number of metastable states of the system is extended. These extended metastable states do not arise from the brain nor the agent in isolation but from the whole brain-body-environment system. We then use tools of mutual information to observe how neural plasticity in the agent in coordination is key for generating the structures that sustain flexible and metastable behavioral patterns. Finally, we use transfer entropy to characterize the loop of interactions between oscillatory dynamics, synaptic plasticity and behavioral patterns, describing the circular multiscale relation between the agent's neural controller and its behavioral dynamics necessary for generating extended sensorimotor metastable states.

[3] K_m is an arbitrary threshold for capturing when weights are active and sufficiently strong. Different values have been tested without altering the results.

3.1. Scale-free Statistical Distribution and Mestastability

We start by analysing the properties of the statistical distribution of the situated and passively-coupled agents. We compute the probability density function $P(s)$ for each agent, where $s = \{\Theta, \Psi, \Lambda\}$, by sampling 100 simulations of identical agents with random initial conditions during 1000 trials ($1.25 \cdot 10^6$ steps). We generate 100 samples, each one composed of 10^4 random states s randomly sampled from each simulation run (10^6 states in total) and calculate the frequency of occurrence of each state. The result from one randomly selected sample is shown in **Figures 4**, **5**, while in the text we provide the statistics of the complete set of 100 samples.

An initial finding about the probability density function of the agent is that it approximately follows the Zipf law (**Figure 4A**), for states with a probability larger than $2 \cdot 10^{-4}$. If the occurrence of the states of the system s is ordered by their decreasing frequency $P(s)$, Zipf's law states that $P(s)$ decays as the inverse of their rank $r(s)$ in the ordered sequence, making $P(s) \propto 1/r(s)$. The occurrence of Zipf-like distributions is considered to be a signature of criticality (Mora and Bialek, 2011), coinciding with previous analysis of self-organized critical patterns in the same robotic agent when it is coupled with its environment (Aguilera et al., 2015). We observe that, while the situated agent presents a pattern very close to Zipf's distribution, the passively-coupled agent diverges more from a perfect scale-free distribution (**Figure 4B**), especially in states with higher probabilities. We can quantify this divergence by computing the Kullback-Leibler divergence between the distribution of states of the agent $P(s)$ and Zipf's law distribution $P_{sf}(s) \propto 1/r(s)$. As we observe in **Figure 4C** the divergence from the Zipf distribution in the passively-coupled agent is more than twice that of the situated agent. Computing the average and standard deviation for the 100 generated samples reveals that this result is repeated for different agents, confirming that the Kullback-Leibler divergence to a Zipf distribution is much larger in the passively-coupled agent ($\mu = 0.381, \sigma = 8.44 \cdot 10^{-4}$) than in the situated case ($\mu = 0.141, \sigma = 5.46 \cdot 10^{-4}$).

Criticality in the brain is generally associated with the metastability of transiently formed neuronal assemblies (Werner, 2007b), although in general the exact relation between the existence of metastable states and criticality is still not well understood. The definition of a metastable state is a state whose energy is lower than any of its adjacent states while not being the state of minimum energy of the system. If we assume that the probability of each state follows a Boltzmann distribution[4], metastable states will be those with higher probabilities than those of their adjacent states. We define adjacency between two states when they are separated by a single flip[5] of an individual variable Θ_i, $\Psi_{i,j}$ or Λ. In short, we consider metastable states as local peaks in the probability landscape. If we compute the number of metastable states of the system $s = \{\Theta, \Psi, \Lambda\}$, we observe that the situated agent presents 17 metastable states for most of our data series ($\mu = 17.2, \sigma = 0.782$)), while the passively coupled agent presents typically over 13 metastable states ($\mu = 13.1, \sigma = 0.902$), indicating that the level of metastability is boosted when the agent is in interaction with its environment. However, if we only analyse the neural system of the agent, i.e., the system $s' = \{\Theta, \Psi\}$, we find that the situated agent presents 13 metastable states ($\mu = 13, \sigma = 0$) and the passively-coupled agent presents around 11 ($\mu = 11.0, \sigma = 0.243$). **Figure 5** portrays the number of metastable states of one random sample.

These results show how the critical scaling and the repertoire of metastable states of the agent is extended when it is coupled with its environment. Moreover, metastable states generated when coupled with the environment cannot be reduced to metastable states in the "brain" of the agent (i.e., in $s' = \{\Theta, \Psi\}$) but only appear when we analyse the distribution of the complete system ($s = \{\Theta, \Psi, \Lambda\}$). These results suggest that, aside from neural metastable states generated by oscillatory dynamics and neural plasticity, sensorimotor metastable states can appear from the coordination between behavioral patterns and internal

[4]The situated system can be considered an isolated system in thermodynamic terms, whereas the passively-coupled system may be considered a closed system exchanging energy but not matter with the situated system, thus making plausible the assumption of thermal equilibrium and a Boltzmann distribution of states, and therefore a direct mapping between probabilities and the energy of each state.

[5]Other definitions of metastability could be used, using an arbitrary number n of flips. Nevertheless, in this particular case only $n = 1$ yields the existence of mestastable states.

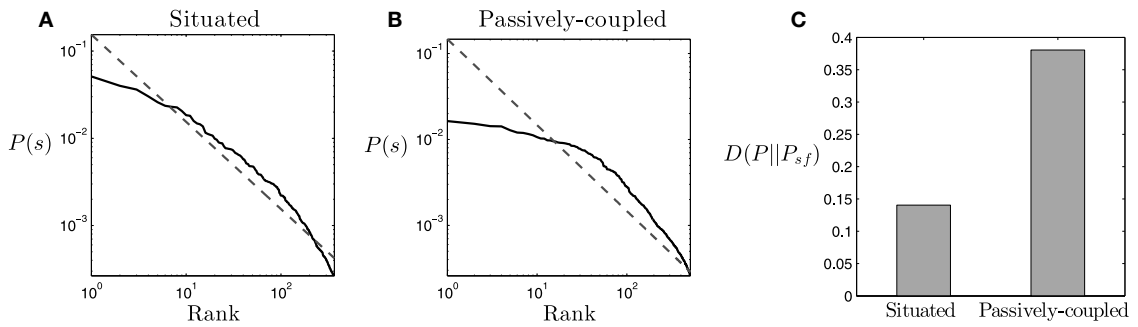


FIGURE 4 | Statistical distribution. (A,B) Ranked probability density function of the states $s = \{\Theta, \Psi, \Lambda\}$ for **(A)** the situated agent and **(B)** the passively coupled agent (solid line), compared to the distribution of a Zipf-like distribution (dashed line). **(C)** Divergence between the probability density function of each type of agent and a Zipf-like distribution. States with probability lower than $2 \cdot 10^{-4}$ are dismissed from the plot and calculations.

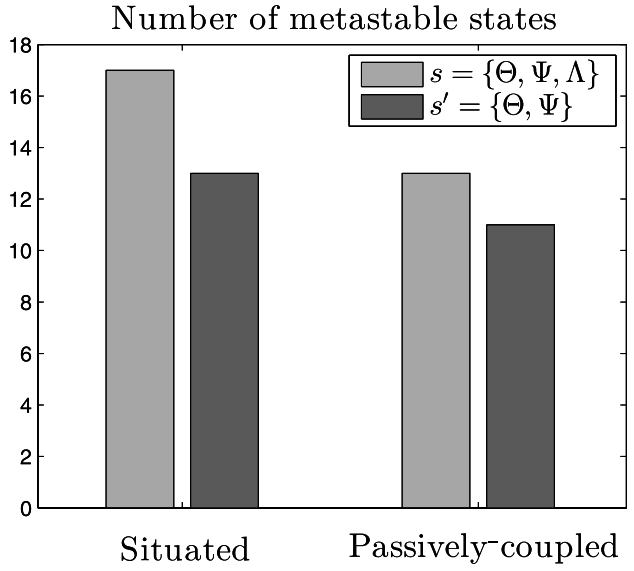


FIGURE 5 | Metastable states in the system. Count of metastable states of the probability density function of neural and behavioral patterns $P(s)$ ($s = \{\Theta, \Psi, \Lambda\}$), and of neural patterns alone $P(s')$ ($s' = \{\Theta, \Psi\}$), for both the situated and passively-coupled agents.

neural dynamics. Spin glass theory indicates that metastable states emerge when some of the couplings between variables are negative (Mezard et al., 1987), which can be translated to stating that in our case agent-environment effective coupling presents mechanisms of mutual inhibition between pairs of variables. The appearance of metastable states only existing for the situated case in the whole sensorimotor system suggests a complex regulation between neural and sensorimotor processes, inviting us to take a closer loop of how agent-environment relations take place in order to increase the metastability of the system.

3.2. Mutual Information Flows

Typically, features such as criticality and metastability have been linked to the idea of systems driven by interaction dynamics between its components (Jensen, 1998; Van Orden et al., 2003; Ihlen and Vereijken, 2010). We investigate what type of interaction takes place in our agent between oscillatory dynamics, neural plasticity and behavioral patterns to generate neural and sensorimotor metastable states. We use information theory tools to quantify the interactions between the components of the system in the situated and passively-coupled agents, simulated for 1000 trials with a duration of 1250 steps. Although we only show the results of the time series of an individual simulation (random sampling is not applicable if we want to maintain temporal correlations), the differences in the results analyzing different runs of the simulation were negligible.

First, we analyse what information is shared by the emergent cell assemblies Θ, synapse assemblies Ψ and behavioral patterns of the agent Λ by measuring mutual information along the time series of values of each variable. In **Figure 6** we can observe how the three variables share an important amount of information. The entropy of Λ (which is the variable with the lowest entropy) is 0.86, thus the shared entropy is in the same order of magnitude in most cases. In the case of the situated agent, we can observe in **Figure 6** (left) that all variables share a relevant amount of information. However, in **Figure 6** (right) we observe that the information shared between Ψ and Λ decreases dramatically, suggesting that most of the interaction between the two variables is lost.

The information analysis above shows a static picture of information flows on average, but it does not explain how these flows unfold over time. To overcome this limitation, Beer and Williams (2015) have proposed a framework combining information flow and dynamical analyses, exploring how a simulated model agent in a relational categorization task integrates information at different moments of time about a cue used for solving the task. Instead of analysing information as an average of the dependences between variables along a time series, they run the same task several times for different initial conditions and compute information measures for each time instant. Instead of using a series of temporal values of a variable, they use a series of values of a variable on each instant along different starting conditions. Similarly, we fold our time series into 1000 time series (one for each simulated trial) in which e.g., $\Theta'_T(t) = \Theta(t + (T-1)1250)$, where T is the trial number and 1250 is the duration of each trial. Consequently, for each value of t we have 1000 values of $\Theta'_T(t)$ we can use for computing mutual information with other variables. For six consecutive trials (i.e., $t = 1, ..., t_{end} = 6 \cdot 1250$) we compute the mutual information between the neurodynamic variables of the agent $\Theta'_T(t)$ and $\Psi'_T(t)$ and the behavioral pattern at the end of the sixth trial $\Lambda'_T(t_{end})$ (in order to observe how information about future behavioral patterns is accumulated). Also, we compute the joint mutual information that both $\Theta'_T(t)$ and $\Psi'_T(t)$ share with $\Lambda'_T(t_{end})$.

In **Figure 7** (left) we can observe the result for the situated agent. We see how the mutual information $I(\Theta'_T(t); \Lambda'_T(t_{end}))$ increases during the middle of the trial, fading out at the beginning and the end. That is, the activation of specific neural patterns when the robot is approaching a light contributes to its repetition in future trials (i.e., contributing to the habit of choosing that light). However, when the robot engages in other behavior (e.g., exploring its surroundings, or stopping after having reached a light) the habit is no longer being enacted and the information about it vanishes from its oscillatory patterns. Also, from one trial to the next the information at the peak increases, being maximal at the sixth trial. We can interpret this as a self-sustaining behavior of cell assemblies: when a cell assembly emerges, it reinforces itself and has more probabilities to reemerge in the next trial. Similarly, $I(\Psi'_T(t); \Lambda'_T(t_{end}))$ steadily increases until a cell assembly is activated at the middle of the sixth trial. Mutual information between $\Psi'_T(t)$ and $\Lambda'_T(t_{end})$ is continuously accumulated and does not decrease, thus we hypothesize that the configuration of the synapse ensembles "stores" information about the behavior that the agent will develop. Furthermore, when we analyse the joint mutual information $I(\Theta'_T(t), \Psi'_T(t); \Lambda'_T(t_{end}))$, we observe that it is always higher that the individual contributions. Moreover,

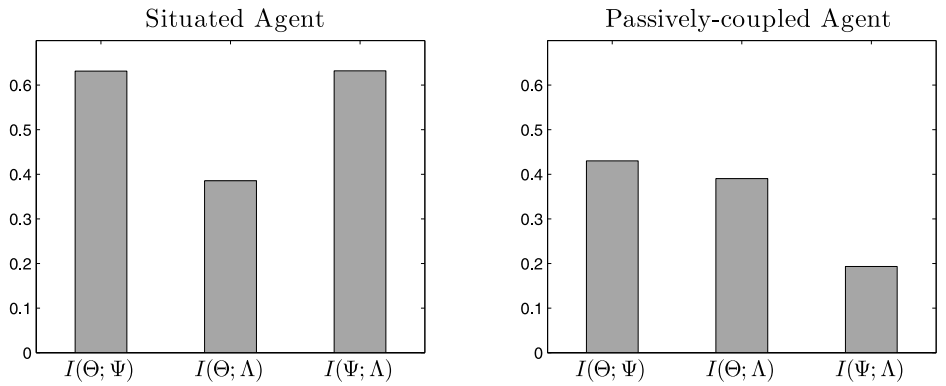


FIGURE 6 | Mutual information. Values of mutual information between cell assemblies (Θ), synaptic assemblies (Ψ) and behavioral patterns (Λ) for the situated agent (left) and passively-coupled agent (right).

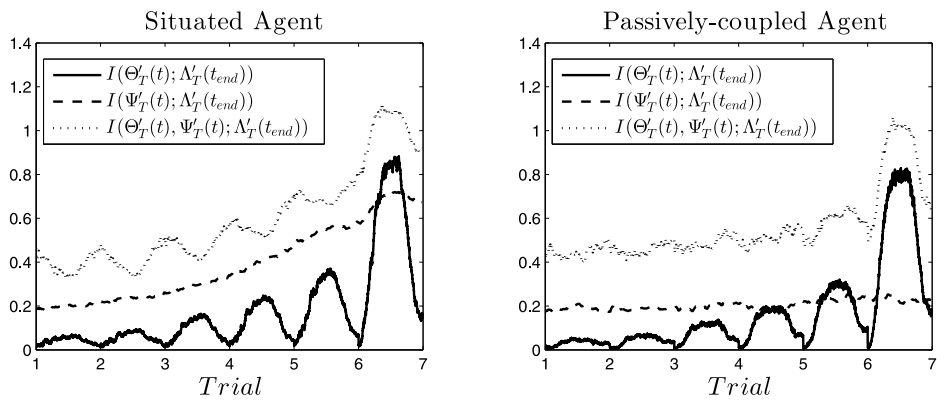


FIGURE 7 | Information flows. Mutual information at different instants between cell assemblies ($\Theta'_T(t)$), synaptic assemblies ($\Psi'_T(t)$) and the behavioral patter displayed at the end of the sixth trial ($\Lambda'_T(t_{end})$) for the situated agent (left) and passively-coupled agent (right).

it increases when $I(\Theta'_T(t);\ \Lambda'_T(t_{end}))$ decreases at the transitions between one trial and another. Also, $I(\Theta'_T(t),\ \Psi'_T(t);\ \Lambda'_T(t_{end}))$ decreases when at the middle of the trial a cell assembly is activated, except in the last trial. This portrays an interesting picture, where information flows back and forth between the emergent cell assemblies and the collective cell-synapse assemblies, until the sixth trial when an assembly emerges producing behavior $\Lambda'_T(t_{end})$.

If we analyse the passively-coupled agent we observe a quite different picture (**Figure 7**, right). Although $I(\Theta'_T(t);\ \Lambda'_T(t_{end}))$ is quite similar in both cases (its values are slightly smaller in the passively-coupled condition), $I(\Psi'_T(t);\ \Lambda'_T(t_{end}))$ does not integrate any information. This suggests that even when the input produced by behavior $\Lambda'_T(t_{end})$ is able to influence the cell assemblies that emerge, coordination between behavior and the stabilization of synapse assemblies does not take place. Furthermore, the joint information $I(\Theta'_T(t),\ \Psi'_T(t);\ \Lambda'_T(t_{end}))$ does not integrate much information either, and the anticorrelation between $I(\Theta'_T(t);\ \Lambda'_T(t_{end}))$ and $I(\Theta'_T(t),\ \Psi'_T(t);\ \Lambda'_T(t_{end}))$ disappears. This suggests that the passively-coupled agent does not capture the struggle between information flows through individual and collective variables, indicating that the important moments for generating the behavior of the agent are not only synchronizing moments of emergence of cell assemblies, but that most information is built during instants of desynchronization corresponding to transition from one assembly to the next.

These results show that synaptic plasticity, in coordination with behavioral patterns, plays a fundamental role in the situated agent, since it allows the agent to "store" information about future behavioral patterns the agent will engage it. Interestingly, the results in **Figure 7** suggest that neural assemblies behind the execution of a specific behavioral pattern reinforce the synaptic circuits sustaining that pattern, which store information about the repetition of that behavioral pattern. This resonates with the idea of the brain as a plastic system of open loops (Fuchs, 2011) created during previous interactions with the environment and functionally closed to full sensorimotor cycles in every new coupling with the environment. Sensorimotor metastable states could be precisely the transient closure of those loops, which in turn imprint and reinforce onto the brain the synaptic structures necessary to their reproduction.

3.3. Transfer Entropy

The analysis above shows the information shared by variables unfolding through time. As mutual information is a symmetric index, it is not a good tool to characterize causal interactions between parts of a system. Instead, we characterize directional interactions by measuring transfer entropy between variables using Equation 13, with $d = d' = 1$ as the length of the past history that we take into account[6] and a logarithmically distributed series of values of τ from 1 to 625, 000 steps (half the length of the 1000 trials), with multiplicative intervals of $10^{0.1}$. In **Figure 8** (left) we can observe a complex chart of information flows for the situated agent:

- $\Theta - \Psi$ transfer. We can observe transfer entropy from Θ to Ψ taking place at small and medium values of τ, whereas at larger values of τ the flow of the information is reversed. This suggests a circular causal chain in which, at short timescales, the structure of the current synaptic assembly determines the cell assemblies that can emerge, but at long timescales it is the self-sustainment of particular assemblies during different trials that determines the stability of the possible synaptic assemblies.
- $\Theta - \Lambda$ transfer. We observe that while there is an important transfer entropy flow from Λ to Θ at fast and medium timescales, the flow does not exist in the opposite direction (**Figure 8**, middle-left), suggesting that the behavioral pattern of the agent influences the cell assembly that emerges, but that the current cell assembly that is active at a particular moment of time is not decisive for the behavioral pattern that the agent will deploy.
- $\Psi - \Lambda$ transfer. There is an important bidirectional exchange of information between Ψ and Λ. This suggests that Ψ is the variable that determines the behavior that will be chosen by the agent. Also, we can observe that $TE_{\Lambda\to\Psi}$ and $TE_{\Theta\to\Psi}$ are very similar in value and shape (if we integrate the area of the difference between $TE_{\Lambda\to\Psi}$ and $TE_{\Theta\to\Psi}$ and divide it by $TE_{\Theta\to\Psi}$ the result is 0.11, showing that both functions coincide with almost 90% of accuracy). This is supported by the fact that there is no functional dependency from Λ to Ψ, since change in the weights $\delta\dot{K}$ is only a function of K and θ. As we can easily check, all the other information flows present in **Figure 8** correspond to actual functional dependencies depicted by the equations defining the systems[7]. This suggests (since Λ influences Θ and not otherwise), that Λ causally determines Θ which in turn influences Ψ in their circular mutual interaction[8].

[6]We tested different values of d and d' up to 5 for a series of representative values of τ with similar results, therefore we used $d = d' = 1$ to reduce the computational cost, since just varying the value of τ seems to be enough for capture the different timescales of information flows.

[7]Since $\dot{\theta}$ is a function of θ, K and s_{DX}, and $\dot{s}_{DX}$ is a function of θ, K and s_{DX}. The only functional dependency that presents no informational content is $TE_{\Theta\to\Lambda}$

[8]It is worthy noting that in the experiments depicted in **Figure 8**, the variable Λ is defined as the macroscopic behavior of the agent, whereas Θ and Ψ are patterns of low level variables of the agent. To test that this difference did not distort in any way the result of our analysis, we repeated the analysis substituting Λ by a variable

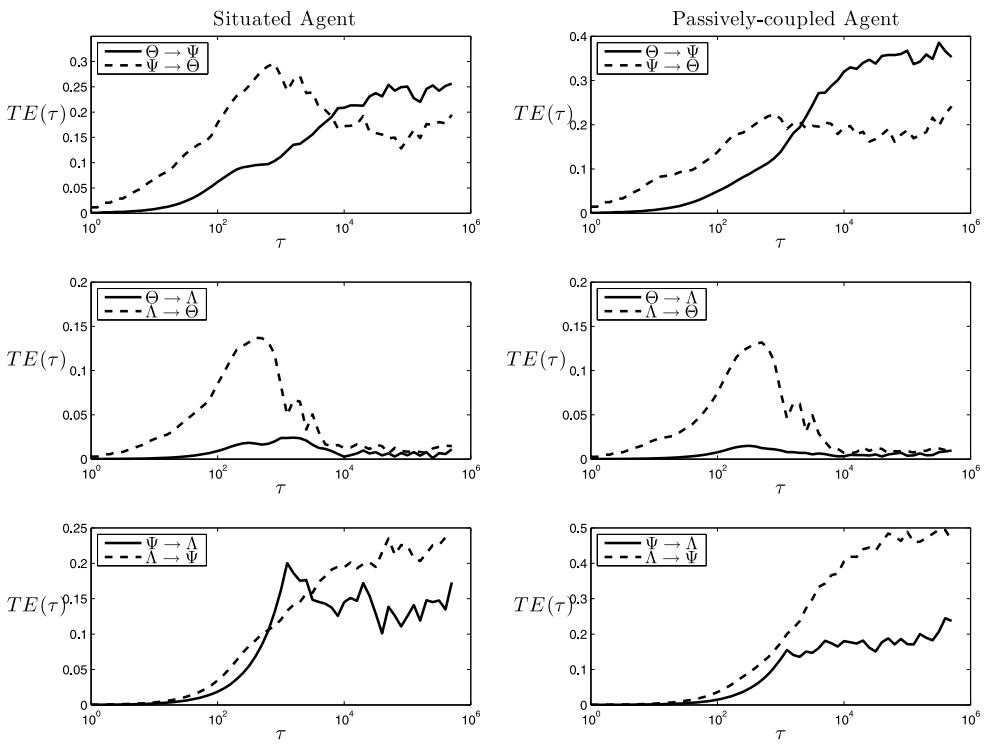


FIGURE 8 | Transfer Entropy. Values of transfer entropy at different timescales among cell assemblies (Θ), synaptic assemblies (Ψ) and behavioral patterns (Λ). Note that the duration of a trial corresponds to a value of τ of 1250 steps.

Putting together the transfer entropy flows above, we may summarize them in the schema at **Figure 9** (left). The behavior of the agent Λ generates an input that determines the emergence of cell assemblies Θ at fast timescales. A circular relation between the emergent cell assemblies and their underlying synapse ensembles Ψ generates a particular behavior Λ which is determined by the state of Ψ at longer timescales. We can observe how the resulting schema is similar to the one proposed by Varela (1997) and depicted in **Figure 1**, though adding interesting information about the timescales of each dependency.

Moreover, for the passively-coupled agent, the information flows from Ψ to Λ and from Ψ to Θ are disrupted and reduced in comparison to other flows, whereas information flows from Λ to Θ and Θ to Ψ[9] are maintained or even increased. This disruption of part of the transfer entropy flows dissolves the self-sustained neurodynamic structures that generate a coordinated behavior and reinforcing unidirectional influence from Λ to Θ and Θ to Ψ, as we depict at **Figure 9** (right).

These results strongly suggest that the generation of complex and integrated neurodynamic structures is a product of a double circular loop that strongly couples (1) neural oscillatory patterns with the plastic synaptic structures sustaining them and (2) these neurodynamic circular structures with the behavioral patterns generating them. This double loop is constituted in a way that creates a circular asymmetry between agent and environment, in which the oscillatory dynamics of the agent present a sensitivity of environmental parameters at shorter timescales, and its repeated activation generates the synaptic structures that are able to engage and disengage different behavioral patterns of the agent at longer timescales. The operational closure of the nervous system implies a special circular relationship between agent and environment, in which the agent individuates itself in front of its environment as it is capable of being sensitive to small fluctuations of its world while being able to act over it at longer timescales as a coherent dynamical unit.

S defined as a string of 4 bits (one for each sensor of the robot), in which each bit was equal to one if the activity of its corresponding sensor is higher than its average activation. The results of transfer entropy measures were strikingly similar to those displayed in **Figure 8**, indicating that the series of sensory activation are strongly correlated with particular behavioral macroscopic patterns of the agent. Since the use of Λ simplifies other analyses performed in this paper, we have chosen to use this variable instead of *S*.

[9]Flows from Θ to Ψ are almost equivalent to flows from Λ to Ψ, supporting our suspicion that they represent the same information flow.

4. DISCUSSION

In this paper we have presented a neurodynamical model of oscillatory activity with synaptic activity embedded in a robotic agent in a behavioral preference task. The model is based on a network of three Kuramoto oscillators with plastic homeostatic mechanisms designed to maintain constant phase relations among oscillators. Our goal was to explore metastability in behavior and neural assemblies in a context of embodied, adaptive activity, in which the agent is in continuous and bidirectional interaction with an environment, a dimension which is frequently neglected in the study of brain activity and organization. The model was carefully designed for exploring (1) the integration of transient assemblies underlying behavior through nonlinear coupling neural clusters generating specific conducts in the agent, and (2) the coordination between sensorimotor and plastic neurodynamic structures into a self-maintaining behavioral patterns. The integration of these two levels of activity gives rise to metastable sensorimotor integrated patterns which cannot be reduced to metastability of brain dynamics alone, as behavioral preferences of the agent emerge from the interaction between oscillator cell ensembles, ensembles of synaptic weights and the agent's sensorimotor coupling. We present a methodological framework to analyse the role of sensorimotor behavior in interaction with neural dynamics: we compare a situated agent, normally interacting with its

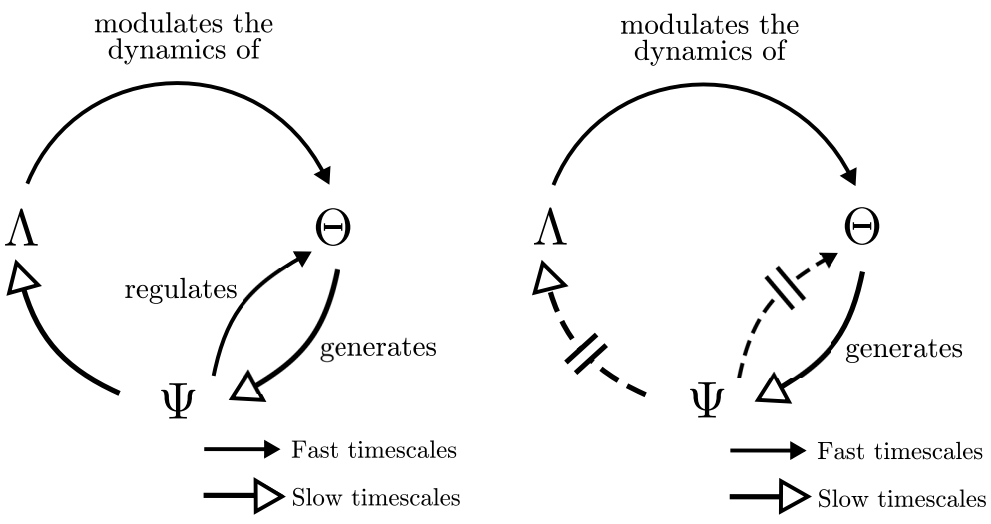


FIGURE 9 | Operational closure. Simplified information flows among cell assemblies (Θ), synaptic assemblies (Ψ) and the behavioral patterns (Λ) for the situated (left) and passively-coupled (right) agents.

environment, and a passively-coupled agent, receiving a sensory input recorded from the situated agent but unable to influence its environment in any way. Comparing a passively-coupled agent and a situated agent in an open environment we have found three different results shedding light on the relation between neural and behavioral metastability.

These results are obtained first, through a statistical description of the agent (in the form of a discrete characterization of the states of the agent neural configurations and behavioral patterns), depicting how the situated agent presents signatures of criticality and additional metastable states that are not present in the passively-coupled agent. Moreover, we show how those additional metastable states do not appear in neural variables alone, but in the combined space of neural and behavioral patterns, indicating that metastable states in behavior are not exclusively a direct mapping of neural metastable states. Instead, we find the existence of sensorimotor metastable states that extend the range of metastability of the agent's "brain."

Second, an analysis of the flow of mutual information between different groups of variables of the agent shows that in both the situated and passively-coupled models neural ensembles of oscillatory components contain a lot of information about the behavioral pattern being developed. However, this information is lost after the neural ensemble dissolves at the end of a trial, and only in the situated case this information is stored instead through changes in synaptic plasticity, generating the structures reinforcing future appearances of that behavioral pattern. This suggests that mechanisms of synaptic plasticity have a fundamental role in coordinating neural and behavioral processes, sculpting the sensorimotor structures sustaining extended metastable states. While synapse assemblies have been hypothesized to have a critical role for building up and dissolving metastable cell assemblies in the brain as well as linking together sequences of cell assemblies (Buzsáki, 2010, p.372), our results invite us to rethink the role of synapse assemblies in a broader sense, as fundamental elements that facilitate emergence and dissolution of metastable modes of engagement with the environment.

Finally, using transfer entropy we can depict the causal influences at different timescales between different components of the system. We observe how the situated agent generates a closed network of interaction, circularly organized with bidirectional interactions at different timescales. This network takes the form of a double coupling loop of (1) a circular causal dependence between the emergence of cell assemblies and the synaptic neural structure that generates them, and (2) a causal chain in which synaptic structures influence the behavior displayed by the agent, which in turn triggers the emergence of specific neural assemblies (**Figure 9**, left). If the sensorimotor loop is disrupted (e.g., when the agent is passively-coupled, and probably for more severe sensorimotor disruptions), this circular closure disappears and autonomous organization of the agent vanishes (**Figure 9**, right). This provides cues of what happens in real-life examples of disruption of sensorimotor coordination. For example, in Held and Hein's experiment on visually-guided behavior (Held and Hein, 1963), the "passive" kitten fails to develop perceptual abilities. Similar situations can take place in different physiological or pathophysiological conditions. Examples of those are the problems faced by deafferented subjects (i.e., without any proprioception) to develop behavioral automatisms exclusively in the absence of sensory feedback; needing to rely on visual feedback to perform simple tasks such as holding an egg without breaking it, or when are unable to maintain an upright posture in the dark (Cole and Paillard, 1995).

Interestingly, the schema of circular dependencies in the agent resembles the idea of operational closure of the nervous system proposed by Varela (1997) depicted in **Figure 1**. Moreover, it describes a novel characteristic of closure since this loop of closure creates a multiscale asymmetry between agent and environment: the agent is sensitive to changes in the environment at fast timescales, while it can influence the environment at slow timescales (**Figure 8**, left). Although the agent's oscillatory dynamics are mostly driven by inputs from the environment at fast timescales, it exerts an influence over the environment at the slower timescales of synaptic plasticity by generating structured behavioral habits (e.g., reaching repeatedly one of the lights) which will influence future stimuli received by the agent. In some sense, once the agent "sees" a light it is trapped in a behavioral field and has to reach it, but it still has a degree of autonomy in the sense that it can modulate its internal connectivity to influence which lights it is going to be sensitive to in the future. This allows us to identify the agent as a unit which is affected by bottom-up causal flows of sensorimotor stimuli and, at the same time, it is able to develop a downward causation modulating its sensorimotor interaction. Breaking the symmetry of the coupling between agent and environment has been proposed as one of the fundamental aspects which can constitute an agent as an autonomous entity able to regulate from within its exchange with the world, constituting its identity as a self-individuating system (Barandiaran et al., 2009). Previous characterizations of agent-environment asymmetry have referred only to the presence of a directionality in the flows of information from agent to environment depicting a causal influence from the former to the latter (see Seth, 2007; Bertschinger et al., 2008). These contributions do not take into account a self-referential operational closure of the system (Bertschinger et al., 2008, p.14) and only quantify the degree of self-determination of the system. In contrast, our approach captures agent-environment asymmetry as a circular relation of causal influences at different scales.

The type of analysis performed in our model has typically been unexplored by neuroscientists studying real life organisms, partly due to the difficulty of recording whole-brain activity of freely behaving animals. In general, recordings of neural activity have been limited to either small brain regions or to immobilized or anesthetized animals exhibiting limited behavior. Nonetheless, during the last few years some promising results point to the plausability of experiments involving the sensorimotor engagement of whole-brain or large brain areas. For example, an interesting technique for analysing brain and behavioral activity of a head-restrained mouse interacting with a virtual reality environment in a spherical treadmill (see Dombeck et al., 2007)

has been developed. Furthermore, the first report of whole-brain recording in freely behaving animals has been reported for the nematode Caenorhabditis elegans during free locomotion (Nguyen et al., 2016). These advances open-up an exciting path for neuroscience, allowing an exploration of how interesting properties of neural processes such as criticality and metastability are extended and amplified when they are embedded in ongoing embodied sensorimotor loops. In such scenarios, minimal models of brain-body-environment dynamical regulation in adaptive behavior, such as the one presented here, offer a conceptual basis for facing complex analysis in real animals due to the low dimensionality of their dynamics. Even models that have little connection with biological brains can provide insights into how neural and sensorimotor dynamics may interact (e.g., extending the range of metastable states, creating asymmetrical loops of causal interaction, etc.), as well as contribute to the development of conceptual methodological tools for understanding the role of the different scales of the system (e.g., the situated vs. passively-coupled comparison to address the role of sensorimotor regulation in the nervous system). Moreover, the availability of real data of brain-body-environment interaction will provide the opportunity to advance in the design of more accurate and realistic models, bringing us closer to capturing fundamental aspects of adaptive behavior.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: MA, MB, XB. Performed the experiments: MA. Analyzed the data: MA. Wrote the paper: MA, MB, XB.

FUNDING

This research has been partially supported by the project PSI2014-62092-EXP of the National Programme for Fostering Excellence in Scientific and Technical Research ("Explora Ciencia" call) from the Spanish Ministry of Economy and Competitiveness.

ACKNOWLEDGMENTS

Thanks to Ezequiel Di Paolo, Seth Bullock, and Ricard Solé for useful comments and discussion on some of the results presented in this paper.

REFERENCES

Abbott, L. F., Rajan, K., and Sompolinsky, H. (2009). Interactions between intrinsic and stimulus-evoked activity in recurrent neural networks. arXiv:0912.3832.

Acebrón, J. A., Bonilla, L. L., Pérez Vicente, C. J., Ritort, F., and Spigler, R. (2005). The Kuramoto model: a simple paradigm for synchronization phenomena. *Rev. Mod. Phys.* 77, 137–185. doi: 10.1103/RevModPhys.77.137

Aguilera, M., Barandiaran, X. E., Bedia, M. G., and Seron, F. (2015). Self-organized criticality, plasticity and sensorimotor coupling. Explorations with a neurorobotic model in a behavioural preference task. *PLoS ONE* 10:e0117465. doi: 10.1371/journal.pone.0117465

Aguilera, M., Bedia, M. G., Santos, B. A., and Barandiaran, X. E. (2013). The situated HKB model: how sensorimotor spatial coupling can alter oscillatory brain dynamics. *Front. Comput. Neurosci.* 7:117. doi: 10.3389/fncom.2013.00117

Barandiaran, X. E. (2016). Autonomy and enactivism: towards a theory of sensorimotor autonomous agency. *Topoi* 34, 1–22. doi: 10.1007/s11245-016-9365-4

Barandiaran, X. E., Di Paolo, E., and Rohde, M. (2009). Defining agency: individuality, normativity, asymmetry, and spatio-temporality in action. *Adapt. Behav.* 17, 367–386. doi: 10.1177/1059712309343819

Beer, R. D., and Williams, P. L. (2015). Information processing and dynamics in minimally cognitive agents. *Cogn. Sci.* 39, 1–38. doi: 10.1111/cogs.12142

Bertschinger, N., Olbrich, E., Ay, N., and Jost, J. (2008). Autonomy: an information theoretic perspective. *Biosystems* 91, 331–345. doi: 10.1016/j.biosystems.2007.05.018

Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles and readers. *Neuron* 68, 362–385. doi: 10.1016/j.neuron.2010.09.023

Cole, J., and Paillard, J. (1995). "Living without touch and peripheral information about body position and movement: studies with deafferented subjects," in *The Body and the Self*, eds J. L. Bermudez, A. J. Marcel, and N. M. Eilan (Cambridge, MA: MIT Press), 245–266.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.* 1, 123–132. doi: 10.1162/neco.1989.1.1.123

Di Paolo, E., Buhrmann, T., and Barandiaran, X. E. (2016). *Sensorimotor Life. An Enactive Proposal*. Oxford: Oxford University Press.

Di Paolo, E. A., and Iizuka, H. (2008). How (not) to model autonomous behaviour. *Biosystems* 91, 409–423. doi: 10.1016/j.biosystems.2007.05.016

Di Paolo, E. A., and Thompson, L. (2014). "The enactive approach," in *The Routledge Handbook of Embodied Cognition*, eds L. Shapiro (New York, NY: Routledge), 68–78.

Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L., and Tank, D. W. (2007). Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* 56, 43–57. doi: 10.1016/j.neuron.2007.08.003

Dotov, D. G. (2014). Putting reins on the brain. How the body and environment use it. *Front. Hum. Neurosci.* 8:795. doi: 10.3389/fnhum.2014.00795

Edelman, G. M., Gally, J. A., and Baars, B. J. (2011). Biology of consciousness. *Front. Psychol.* 2:4. doi: 10.3389/fpsyg.2011.00004

Engel, A. K., Maye, A., Kurthen, M., and König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends Cogn. Sci.* 17, 202–209. doi: 10.1016/j.tics.2013.03.006

Fuchs, T. (2011). The brain–a mediating organ. *J. Conscious. Stud.* 18, 196–221. Available online at: https://www.klinikum.uni-heidelberg.de/fileadmin/zpm/psychatrie/fuchs/The_Brain_-_A_Mediating_Organ.pdf

Harvey, I., Di Paolo, E., Wood, R., Quinn, M., Tuci, E., and Iridia, E. T. (2005). Evolutionary robotics: a new scientific tool for studying cognition. *Artif. Life* 11, 79–98. doi: 10.1162/1064546053278991

Hebb, D. O. (1952). *The Organisation of Behaviour: A Neuropsychological Theory*. New York, NY: Wiley.

Held, R., and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *J. Comp. Physiol. Psychol.* 56, 872–876. doi: 10.1037/h0040546

Ihlen, E. A. F., and Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: beyond 1/fluctuation. *J. Exp. Psychol. Gen.* 139, 436–463. doi: 10.1037/a0019098

Iizuka, H., and Di Paolo, E. A. (2007). Toward spinozist robotics: exploring the minimal dynamics of behavioral preference. *Adapt. Behav.* 15, 359–376. doi: 10.1177/1059712307084687

Iizuka, H., and Di Paolo, E. A. (2008). Extended homeostatic adaptation: improving the link between internal and behavioural stability. *Media* 5040, 1–10. doi: 10.1007/978-3-540-69134-1_1

Jensen, H. J. (1998). *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*. Cambridge: Cambridge University Press.

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: The MIT Press.

Kelso, J. A. S., Case, P., Holroyd, T., Horvath, E., Rczaszek, J., Tuller, B., et al. (1995). "Multistability and metastability in perceptual and brain dynamics," in *Ambiguity in Mind and Nature, number 64 in Springer Series in Synergetics*, eds D. P. Kruse and P. D. M. Stadler (Berlin; Heidelberg: Springer), 159–184.

Le Van Quyen, M. (2011). The brainweb of cross-scale interactions. *New Ideas Psychol.* 29, 57–63. doi: 10.1016/j.newideapsych.2010.11.001

Mezard, M., Parisi, G., and Virasoro, M. A. (1987). *Spin Glass Theory and Beyond*. Singapore: World Scientific.

Mora, T., and Bialek, W. (2011). Are biological systems poised at criticality? *J. Stat. Phys.* 144, 268–302. doi: 10.1007/s10955-011-0229-4

Nguyen, J. P., Shipley, F. B., Linder, A. N., Plummer, G. S., Liu, M., Setru, S. U., et al. (2016). Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* 113, E1074–E1081. doi: 10.1073/pnas.1507110112

Nolfi, S., and Floreano, D. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. Cambridge, MA: MIT Press.

O'Regan, J., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–1031. doi: 10.1017/S0140525X01000115

Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., and Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature* 397, 430–433. doi: 10.1038/17120

Seth, A. K. (2007). "Measuring autonomy by multivariate autoregressive modelling," in *Advances in Artificial Life, Number 4648 in Lecture Notes in Computer Science*, eds F. A. E. Costa L. M. Rocha, E. Costa, I. Harvey, and A. Coutinho (Berlin; Heidelberg: Springer), 475–484. doi: 10.1007/978-3-540-74913-4_48

Thompson, E., and Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends Cogn. Sci.* 5, 418–425. doi: 10.1016/S1364-6613(00)01750-2

Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *Science (New York, N.Y.)* 282, 1846–1851. doi: 10.1126/science.282.5395.1846

Van Orden, G. C., Holden, J. G., and Turvey, M. T. (2003). Self-organization of cognitive performance. *J. Exp. Psychol. Gen.* 132, 331–350. doi: 10.1037/0096-3445.132.3.331

Varela, F. (1992). "Autopoiesis and a biology of intentionality," in *Autopoiesis and Perception. Proceedings of a Workshop Held in Dublin City University, August 25th and 26th 1992. School of Electronic Engineering Technical Report, 1994*, eds B. McMullin and N. Murphy (Dublin), 4–14.

Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550

Varela, F., and Thompson, E. (2003). "Neural synchrony and the unity of mind: a neurophenomenological perspective," in *The Unity of Consciousness*, ed A. Cleeremans (New York, NY: Oxford University Press), 266–287.

Varela, F. J. (1995). Resonant cell assemblies: a new approach to cognitive functions and neuronal synchrony. *Biol. Res.* 28, 81–95.

Varela, F. J. (1997). Patterns of life: intertwining identity and cognition. *Brain Cogn.* 34, 72–87. doi: 10.1006/brcg.1997.0907

Ward, L. M. (2011). The thalamic dynamic core theory of conscious experience. *Conscious. Cogn.* 20, 464–486. doi: 10.1016/j.concog.2011.01.007

Werner, G. (2007a). Brain dynamics across levels of organization. *J. Physiol. Paris* 101, 273–279. doi: 10.1016/j.jphysparis.2007.12.001

Werner, G. (2007b). Metastability, criticality and phase transitions in brain and its models. *Biosystems* 90, 496–508. doi: 10.1016/j.biosystems.2006.12.001

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.

frontiers in Systems Neuroscience

ORIGINAL RESEARCH
published: 30 June 2015
doi: 10.3389/fnsys.2015.00097

Metastable dynamics in heterogeneous neural fields

Cordula Schwappach[1,2], *Axel Hutt*[3,4,5] *and Peter beim Graben*[1,6]*

[1] Department of German Studies and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany, [2] Department of Physics, Humboldt-Universität zu Berlin, Berlin, Germany, [3] Team Neurosys, Inria, Villers-les-Nancy, France, [4] Team Neurosys, Centre National de la Recherche Scientifique, UMR nō 7503, Loria, Villers-les-Nancy, France, [5] Team Neurosys, UMR nō 7503, Loria, Université de Lorraine, Villers-les-Nancy, France, [6] Bernstein Center for Computational Neuroscience, Humboldt-Universität zu Berlin, Berlin, Germany

We present numerical simulations of metastable states in heterogeneous neural fields that are connected along heteroclinic orbits. Such trajectories are possible representations of transient neural activity as observed, for example, in the electroencephalogram. Based on previous theoretical findings on learning algorithms for neural fields, we directly construct synaptic weight kernels from Lotka-Volterra neural population dynamics without supervised training approaches. We deliver a MATLAB neural field toolbox validated by two examples of one- and two-dimensional neural fields. We demonstrate trial-to-trial variability and distributed representations in our simulations which might therefore be regarded as a proof-of-concept for more advanced neural field models of metastable dynamics in neurophysiological data.

Keywords: neural fields, kernel construction, metastability, heteroclinic orbits, trial-to-trial variability, distributed representations, sub-networks, sparsity

OPEN ACCESS

Edited by:
Emili Balaguer-Ballester, Bournemouth University, UK

Reviewed by:
Pablo Varona, Universidad Autonoma de Madrid, Spain
Basabdatta Sen Bhattacharya, University of Lincoln, UK

***Correspondence:**
Peter beim Graben, Department of German Studies and Linguistics, Humboldt-Universität zu Berlin, Unter den Linden 6, D–10099 Berlin, Germany
peter.beim.graben@hu-berlin.de

Received: *26 March 2015*
Accepted: *15 June 2015*
Published: *30 June 2015*

Citation:
Schwappach C, Hutt A and beim Graben P (2015) Metastable dynamics in heterogeneous neural fields. Front. Syst. Neurosci. 9:97. doi: 10.3389/fnsys.2015.00097

1. Introduction

Metastable states and transient dynamics between metastable states have received increasing interest in the neuroscientific community in recent time. Beginning with Dietrich Lehmann's original idea to identify "atoms of thought" as metastable topographies, so-called *brain microstates*, in spontaneous and event-related electroencephalograms (EEG) (Lehmann et al., 1987; Lehmann, 1989; Lehmann et al., 2009), experimentalists found accumulating evidence that metastability is tentatively an important organization principle in neurodynamical systems. Mazor and Laurent (2005), e.g., reported metastable states in the locust odor system (cf. Rabinovich et al., 2001, 2008a), while Hudson et al. (2014) found metastability in the local field potentials of rats recovering from anesthesia. For the analysis of human EEG, several segmentation techniques into metastable states have recently been suggested by Hutt (2004), Allefeld et al. (2009), and beim Graben and Hutt (2015).

From a theoretical perspective, metastable EEG topographies or components of the event-related potential (ERP) have been identified with saddle-nodes in deterministic low-dimensional systems by Hutt et al. (2000) and Hutt and Riedel (2003). Particularly, the discoveries of winnerless competition (Rabinovich et al., 2001; Seliger et al., 2003) and heteroclinic orbits in neural population dynamics (Afraimovich et al., 2004a,b; Rabinovich et al., 2008b) led to better understanding of metastability and transient behavior in theoretical neuroscience. Winnerless competition is ubiquitous in complex excitation-inhibition networks with strong asymmetries. While symmetric connectivity usually leads to Hopfield-type attractor neural networks (Hopfield, 1982; Hertz et al., 1991) where transient dynamics is only observed for the motion from a basin

of attraction toward an asymptotically stable fixed point attractor, winnerless competition between neural Lotka-Volterra populations (Fukai and Tanaka, 1997; Cowan, 2014) allows for hierarchical transient computations, bifurcations, and the resolution of sequential decision problems, as applied for modeling speech processing (Kiebel et al., 2009), bird songs (Yildiz and Kiebel, 2011), syntactic parsing (beim Graben and Potthast, 2012), and, most recently, working memory (Rabinovich et al., 2014a,b).

However, these phenomena have been investigated on the rather abstract level of macroscopic neural populations so far, without reference to the mesoscopic and microscopic levels of spatially given nervous tissue and individual neurons. One important approach to characterize the former, nervous tissue at the mesoscopic scale, are *neural fields*, i.e., continuum approximations of infinitely large neural networks (Coombes et al., 2014). In a recent theoretical study, beim Graben and Hutt (2014) investigated stationary states and heteroclinic dynamics in neural fields with heterogeneous synaptic connectivity. The present work applies this previous study to describe experimentally observed transient neural activity as a proof-of-concept of our theoretical approach. We propose a novel hypothesis on the origin of trial-to-trial variability observed in most experimental data, on episodic cell assembly dynamics and on sparsely sampled neural representations.

Moreover, we disseminate our software implementation as a MATLAB *neural field toolbox* to facilitate further research on this intriguing field of computational neuroscience.

2. Materials and Methods

In this section we present some of the theoretical findings of beim Graben and Hutt (2014) and indicate how they have been implemented in our simulations.

2.1. Theoretical Background

An important representative of neural fields is given through the *Amari equation*

$$\frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int_\Omega w(x,y)\, f(u(y,t))\, \mathrm{d}y \tag{1}$$

describing the evolution of neural activity $u(x,t)$ at site $x \in \Omega \subset \mathbb{R}^d$ and time t (Amari, 1977). Here, Ω is a d-dimensional manifold, representing neural tissue. Moreover, $w(x,y)$ is the synaptic weight kernel, and f is a sigmoidal activation function, usually chosen as $f(u) = 1/(1+\exp(-\beta(u-\theta)))$, with gain $\beta > 0$, and threshold $\theta > 0$. The time scale of the dynamics, often characterized by a particular time constant is implicitly included in the kernel $w(x,y)$.

The neural field described by Equation (1) is called homogeneous when the kernel is translation invariant: $w(x,y) = w(x-y)$. If the field is not homogeneous it is called heterogeneous.

Stationary states, $v(x)$, of the Amari equation which are obtained from $\partial u/\partial t = 0$ obey the nonlinear Hammerstein integral equation

$$v(x) = \int_\Omega w(x,y) f(v(y))\, \mathrm{d}y\,. \tag{2}$$

By choosing a heterogeneous *Pincherle-Goursat kernel* (Veltz and Faugeras, 2010)

$$w(x,y) = v(x)v(y)\,, \tag{3}$$

and carrying out a linear stability analysis, beim Graben and Hutt (2014) were able to prove that the stationary state $v(x)$ is either an asymptotically stable fixed point attractor, or a saddle with a one-dimensional unstable manifold, i.e., a metastable state. Since such saddles could be connected along their stable and unstable directions, heterogeneous neural fields may exhibit stable heteroclinic sequences (SHS: Afraimovich et al., 2004b; Rabinovich et al., 2008b).

Let $\{v_k(x)\}$, $1 \le k \le n$ be such a collection of metastable states which we assume to be linearly independent. Then, this collection possesses a biorthogonal system of adjoints $\{v_k^+(x)\}$ obeying

$$\int_\Omega v_j^+(x) v_k(x)\, \mathrm{d}x = \delta_{jk}\,. \tag{4}$$

For the particular case of Lotka-Volterra neural populations, described by activities $\xi_k(t)$,

$$\frac{\mathrm{d}\xi_k}{\mathrm{d}t} = \xi_k \left(\sigma_k - \sum_{j=1}^n \rho_{kj}\xi_j \right) \tag{5}$$

with growth rates $\sigma_k > 0$, interaction weights $\rho_{kj} > 0$ and $\rho_{kk} = 1$ that are tuned according to the algorithm of Afraimovich et al. (2004b) and Rabinovich et al. (2008b), the population amplitude

$$\alpha_k(t) = \frac{\xi_k}{\sigma_k} \tag{6}$$

recruits its corresponding metastable state $v_k(x)$, leading to an order parameter expansion

$$u(x,t) = \sum_{k=1}^n \alpha_k(t) v_k(x) \tag{7}$$

of the neural field.

Under these assumptions, beim Graben and Potthast (2012) and beim Graben and Hutt (2014) have explicitly constructed the kernel $w(x,y)$ through a power series expansion of the right-hand-side of the Amari equation (Equation 1),

$$\begin{aligned} \frac{\partial u(x,t)}{\partial t} &= -u(x,t) + \int_\Omega w_1(x,y) u(y,t)\, \mathrm{d}y \\ &+ \int_\Omega \int_\Omega w_2(x,y,z) u(y,t) u(z,t)\, \mathrm{d}y\, \mathrm{d}z \end{aligned} \tag{8}$$

with Pincherle-Goursat kernels[1].

$$w_1(x, y) = \sum_k (\sigma_k + 1) v_k(x) v_k^+(y) \tag{9}$$

$$w_2(x, y, z) = -\sum_{kj} \sigma_j \rho_{kj} v_k(x) v_k^+(y) v_j^+(z)\,. \tag{10}$$

Interestingly, the kernel $w_1(x, y)$ describes a Hebbian synapse between sites y and x whereas the three-point kernel $w_2(x, y, z)$ further generalizes Hebbian learning to interactions between three sites x, y, z of neural tissue.

2.2. Numerical Studies

For a numerical implementation of the theoretical results above, we have to discretize time and space. Using MATLAB, temporal discretization on the one hand is achieved through the ordinary differential equation solver `ode15s` for stiff problems. On the other hand, spatial discretization converts the kernels w_1 and w_2 into tensors of rank two and three, respectively. Consequently, the integrals in Equation (8) become contractions over products of tensors and state vectors $u(t)$. In order to properly deal with tensor algebra, we use the Sandia Tensor Toolbox[2]. Our neural field toolbox, thus obtained is available as Supplementary Material. We evaluate our implementation in the next subsections by means of two examples.

2.2.1. One-dimensional Neural Field

In our first simulation, we use a $d = 1$ dimensional neural field where we choose $n = 3$ sine functions

$$v_k(x) = \sin kx \tag{11}$$

as metastable states on the domain $\Omega = [0, 2\pi]$ discretized with a spatial grid of $N_x = 100$ sites. According to the orthogonality relations

$$\int_\Omega \sin jx \, \sin kx \, \mathrm{d}x = \pi \delta_{jk} \tag{12}$$

we easily obtain the adjoint modes

$$v_k^+(x) = \frac{1}{\pi} \sin kx\,. \tag{13}$$

For the temporal dynamics we prepare the stable heteroclinic contour solving (Equation 5) used by beim Graben and Hutt (2015) with $\sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 3$. Metastable states $v_k(x)$ and their population activities $\xi_k(t)$ are shown in **Figure 1**.

We run simulations with one fixed initial condition and also from an ensemble of 60 initial conditions randomly distributed in the vicinity of the first saddle, where we add some small portion of Gaussian observational noise (noise level $\sigma = 0.005$) in order to demonstrate trial-to-trial variability and hence event-related phase decoherence (Jung et al., 2001; Makeig et al., 2002).

[1] There was a mistake in our previous reports (beim Graben and Potthast, 2012; beim Graben and Hutt, 2014). Although the kernel construction has been correctly derived, a minus sign was omitted in the final result for kernel $w_2(x, y, z)$. This is corrected now.

[2] http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.5.html

2.2.2. Two-dimensional Neural Field

For our second demonstration, we assume a spatially distributed response in a neural population to external stimuli triggering a sequence of neural activity patterns. It is well-established that sensory input features (Pasupathy and Connor, 2002) at earlier stages of the object's representation pathway and memory (Rissman and Wagner, 2012) is encoded by distributed cortical neural populations while objects are sparsely coded in later stages of the representation pathway (Connor, 2005). Here we consider a cortical neural population embedded in two-dimensional space involving interleaved patterns. These patterns are $d = 2$ dimensional gray scale bitmap images of the numbers[3] 1, 2, and 3 (see **Figure 4** in Section 3.2). In the implementation, these bitmaps are downsampled to a 20×20 grid and reshaped into vectors with $N_x = 400$ elements. Adjoint patterns are obtained as Moore-Penrose pseudoinverses (Hertz et al., 1991).

The temporal evolution of these patterns follows the same heteroclinic contour as above. Here, the underlying working assumption is the presence of interacting sub-networks, e.g., reflecting several distributed representations of signal features or of pieces of working memory. The study predicts what one expects to measure in single spatial locations while the neural system encodes information in a spatially distributed population.

3. Results

The results of our simulation studies are presented in this section.

3.1. One-dimensional Neural Field

For the one-dimensional neural field we compare in **Figure 2** the prescribed spatiotemporal dynamics as resulting from the order parameter expansion (Equation 7) with the solution of the Amari (Equation 8).

Figure 2A shows the prescribed dynamics on a spatiotemporal grid with time on the x-axis and space on the y-axis. The instantaneous activations are therefore given by vertical slices. Going from left to right, these slices first exhibit one wave crest (in red at the bottom) and one wave trough (in blue at the top), corresponding to metastable state $v_1(x)$. Around time $t = 15$ the frequency doubles and metastable state $v_2(x)$ can be observed for approximately seven ticks. The third metastable state met by the trajectory around time $t = 21$ is the mode with tripled frequency. It is only stable for five ticks and evolves thereafter into the first mode again.

In contrast, **Figure 2B** depicts the numerical solution of the Amari equation (Equation 8). Obviously, no deviation is visible.

In order to draw neurophysiologically relevant conclusions from our toy model, we consider the metastable states of the heteroclinic contour as "synthetic ERP components" (Barrès et al., 2013) measured with "electrodes" at the particular sampling points. Because ERPs are obtained from averaging spontaneous EEG over ensembles of several trials that are time-locked to the perception or processing of stimuli, we simulate 60 synthetic ERP trials by randomly preparing initial conditions of the Amari equation.

[3] Original images are taken from the webpage http://www.iconarchive.com/tag/number-3 before modifications with respect to color and resolution.

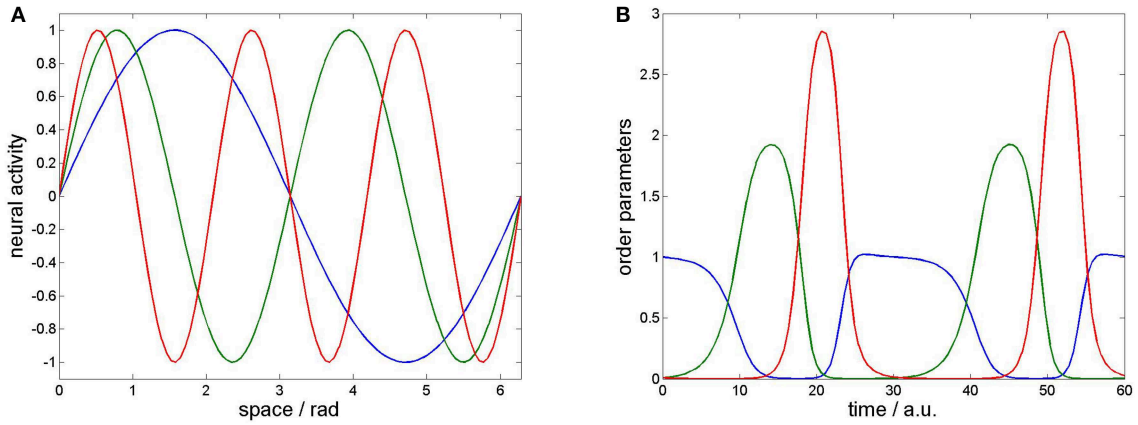


FIGURE 1 | Prescribed dynamics. (A) Three sinusoids as spatial patterns. **(B)** Stable heteroclinic contour resulting from winnerless competition in a Lotka-Volterra system (Equation 5). Blue: $k = 1$, green: $k = 2$, red: $k = 3$.

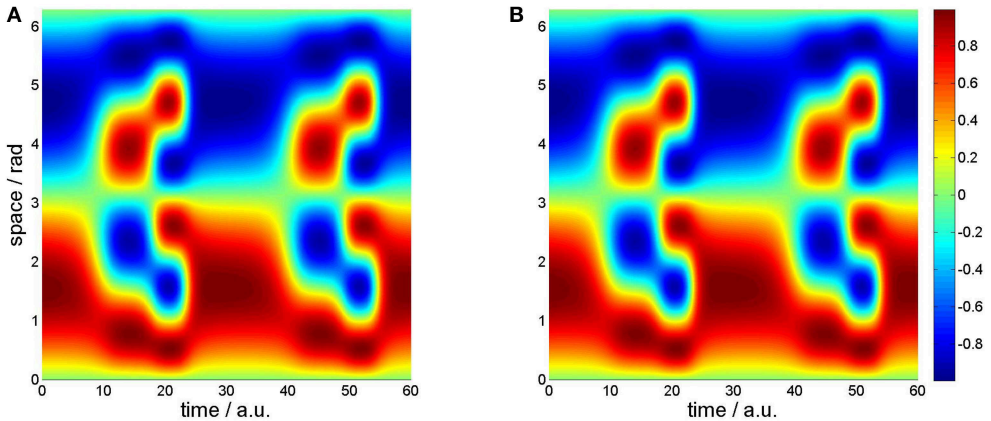


FIGURE 2 | One-dimensional spatiotemporal dynamics. (A) Prescribed trajectory from order parameter ansatz (Equation 7). **(B)** Solution of the Amari (Equation 8).

The results are displayed in **Figure 3** for four "measurement electrodes" at positions 3, 21, 47, and 88. Interestingly, our algorithm exhibited numerical instabilities in five runs which have been marked as "rejected" outliers and excluded from presentation. The resulting 55 trials are shown as colored traces in **Figure 3**. At simulation start all signals are nicely coherent, but later substantial phase dispersions take place (Jung et al., 2001; Makeig et al., 2002).

We also calculated the ERP averages from our simulation shown as bold black traces in **Figure 3**. On the one hand, the averaged ERP is much smoother than the noisy single realizations which justifies averaging in our simulation. However, the averaged ERP significantly decays in the course of time. This is obviously due to the increasing phase decoherence (Jung et al., 2001; Makeig et al., 2002).

3.2. Two-dimensional Neural Field

The numerical simulation of Equation (8) yields a sequence of two-dimensional transient patterns which is shown as a sampled sequence of snapshot maps in **Figure 4**.

According to the different growth rates σ_k of the populations, pattern "1" stays the longest period of time, pattern "2" is visible for a shorter period of time and pattern "3" can be seen for the shortest period of time. These modes represent interweaved spatial networks reflecting intrinsically stored activity patterns.

Now assuming that measurement of neural activity takes place at discrete spatial locations (color-coded points in **Figure 4**), one observes different transient dynamics dependent on the spatial location of the measurement point that is shown in **Figure 5**. Considering the red-coded spatial location, one observes strong activity in the time periods when pattern "1" is active, and well-reduced activity in the time windows of active patterns "2" and "3." Conversely, the activity at the blue-coded location defined in **Figure 4** raises only if pattern "3" is active, otherwise its activity is well-reduced. The green-coded spatial location shows negligible activity in time periods when pattern "1" is active while activity is increased during the emergence of patterns "2" and "3."

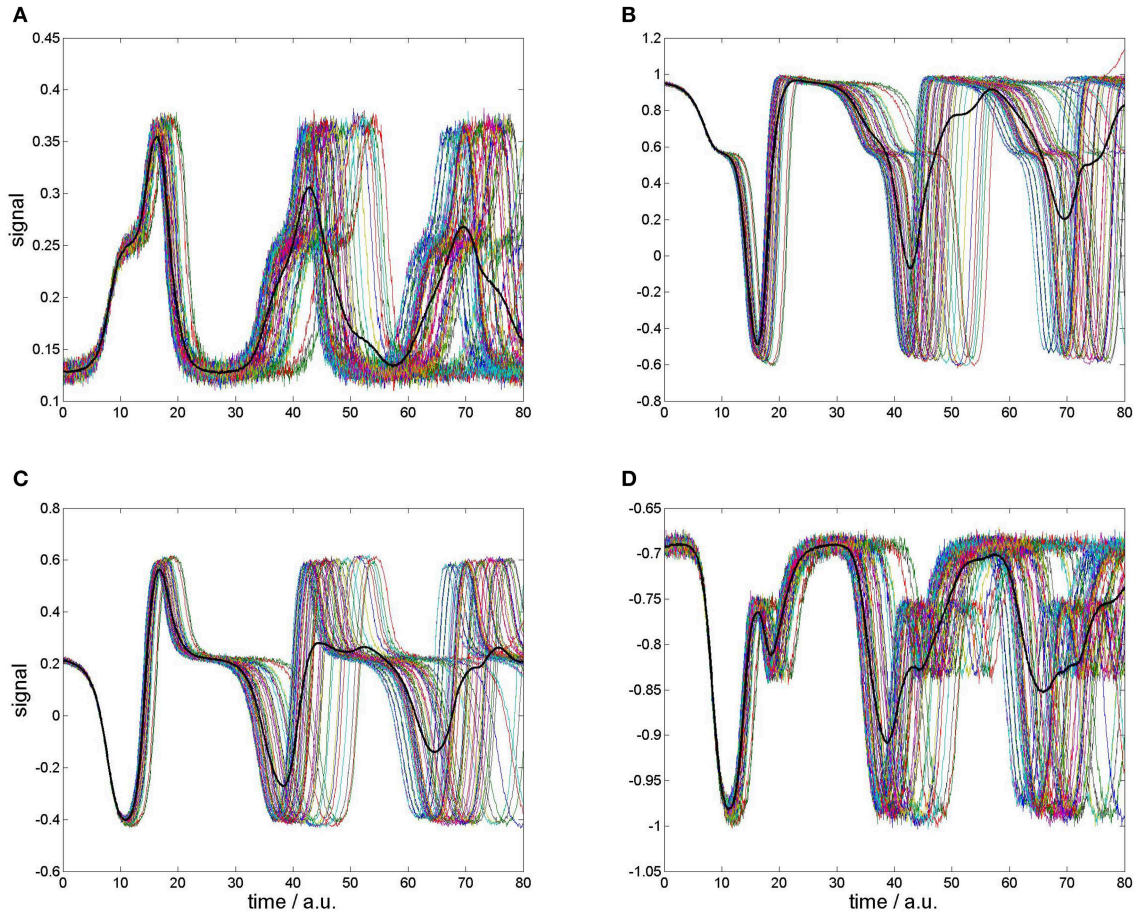


FIGURE 3 | Four selected "recording sites" for neural field simulation with 55 randomly prepared initial conditions (colored traces) and "grand average" (bold black trace). (A) At position: 3, **(B)** position 21, **(C)** position 47, **(D)** position 88.

4. Discussion

In this paper, we presented a software implementation (*neural field toolbox*) and numerical simulation results of previously reported theoretical findings on metastable states and heteroclinic dynamics in neural fields (beim Graben and Potthast, 2012; beim Graben and Hutt, 2014). For the particular case of Lotka-Volterra population dynamics and linearly independent spatial modes, the synaptic weight kernel of the Amari neural field equation (Amari, 1977) can be explicitly constructed from the prescribed metastable states and their evolution parameters as Pincherle-Goursat kernels. This is an important finding as our kernel construction method is not a standard training algorithm such as backpropagation (Igel et al., 2001; beim Graben and Potthast, 2009). Yet it implements a straightforward generalization of Hebbian learning algorithms (beim Graben and Potthast, 2009; Potthast and beim Graben, 2009).

We validated our algorithm by means of two examples, a one-dimensional neural field where metastable states are three sinusoidal excitations over a line, and a two-dimensional example where we have chosen three bitmap images as spatial modes. The temporal dynamics was prescribed as a heteroclinic contour connecting these three patterns in a closed loop. In both simulations, the results were in exact agreement with the prescribed trajectories.

Furthermore, we examined the issues of trial-to-trial variability and distributed representations. In the first example we created solutions for randomly prepared initial conditions, thereby emulating phase resetting in event-related brain potentials (ERP). We observed increasing phase decoherence in the resulting ERP averages. Our model presents a theoretically satisfying explanation for this ubiquitous experimental finding (Jung et al., 2001; Makeig et al., 2002). Assuming that ERP components are metastable states that are connected along heteroclinic orbits (Hutt and Riedel, 2003; beim Graben and Hutt, 2015), single ERP trials start from randomly distributed initial conditions, sometimes closer and sometimes farther from the respective metastable stable. These initial distances from a metastable state result in acceleration and hence in velocity differences in phase space, eventually leading to dispersion and decoherence. Moreover, such a dependence on initial conditions resembles previous experimental results by Pastalkova et al. (2008) showing that identical experimental initial conditions

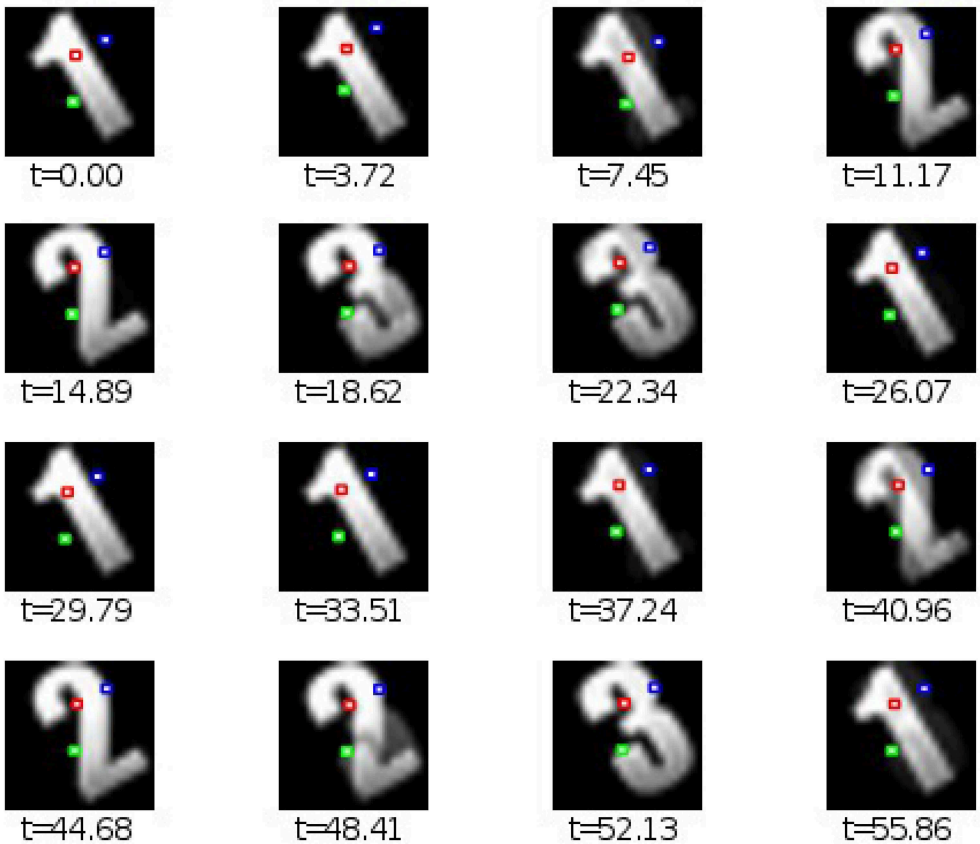


FIGURE 4 | Two-dimensional spatiotemporal solution of model (Equation 8) considering spatial patterns of the numbers "1," "2," and "3" as spatial modes $v_1(x)$, $v_2(x)$, and $v_3(x)$, respectively. The three color-coded points denote three spatial locations whose temporal evolution is shown in **Figure 5**.

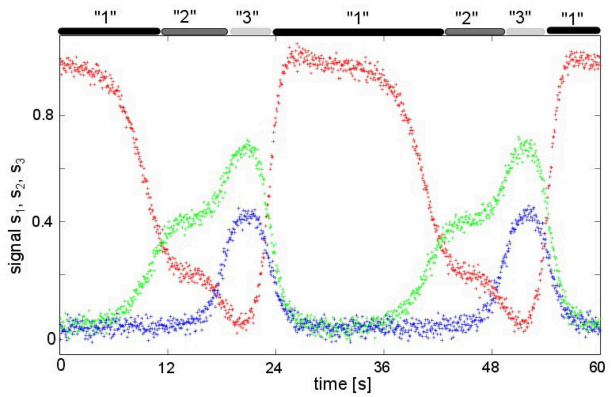


FIGURE 5 | The time-dependent activity $u(x_l, t)$ at three spatial locations x_l, $l = 1, 2, 3$ defined in Figure 4. The upper gray-colored bars denote the emergence time intervals of the corresponding patterns in **Figure 4**. The color codes of the time series correspond to the respective colors of the spatial locations in **Figure 4**.

in a motor task lead to identical sequences of cell assembly activations, while different initial conditions yield different sequences.

For the second example we considered the interaction of three two-dimensional populations, cf. **Figure 4**. The transient passage of the system at metastable attractors has been shown experimentally in previous studies, such as in middle-latent auditory evoked potentials (Hutt and Riedel, 2003) or in the population response of olfactory projection neurons to odor stimuli (Mazor and Laurent, 2005). For instance, the study of Mazor and Laurent (2005) also shows nicely the responses of single neurons in the population revealing different activity in different neurons: some neurons respond to the external stimulus, others remain silent. Such a distinction in response can easily be explained by an insufficient spatial sub-sampling in the measurement and the presence of spatially distributed patterns. However, just spatial sub-sampling does not explain the fully distinct activity of different neurons, such as different episode neurons found in the hippocampus (Pastalkova et al., 2008). Here, different neurons show distinct episodic temporal activities. The equivalent temporal evolution is shown in our simulations in **Figure 5**, where the units at different spatial locations exhibit different temporal sequences of activation that are highly correlated to the presence of the respective pattern representations. This difference results from interacting populations or cell assemblies.

The latter line of argumentation raises the question whether it may explain previous results on sparse neural representations or even may contribute to the question on the existence of "grandmother cells" (Connor, 2005; Quiroga et al., 2005). At a first glance, the present work assumes the existence of interacting spatially distributed sub-networks and supports their existence by a qualitative comparison to previous experimental results by Mazor and Laurent (2005) and Pastalkova et al. (2008). Our assumption of interacting sub-networks does not rule out sparse neural representations since our modeling approach does not stipulate contiguous spatial patterns but also allows for sparse patterns as well.

Metastable neural field dynamics as an ubiquitous organization principle of the brain is also consistent with findings from neuroanatomy and cognitive neuroscience. Anatomically, neural circuits comprise convergent and divergent pathways between populations (Kandel et al., 1991). Assuming that a particular sub-network gets activated by percolation along a convergent pathway and deactivated along a divergent pathway subsequently entails a saddle-node picture in its phase space description, hence a metastable attractor. In cognitive neuroscience, mental representations are regarded as intermediate results of cognitive computations in discrete time. In order to embed these into continuous physical time, they have to be considered connected through continuous trajectories along their stable and unstable directions, i.e., as metastable states, again (beim Graben and Potthast, 2009, 2012).

The present study is a first step toward metastability in neural fields. We hope that our work encourages further research on metastability in neural fields to describe transient neural dynamics by interacting populations and contribute to the description of neural information storage, being either distributed or sparse.

Author Contributions

This study reports results from CS's student internship at Department of German Studies and Linguistics, Humboldt-Universität zu Berlin. CS developed the program code and conducted the numerical simulations. AH included the sub-network study, PbG contributed the study on trial-to-trial variability and compiled the neural field toolbox. All authors wrote the manuscript together.

Funding

PbG acknowledges support by a Heisenberg Fellowship of the German Research Foundation DFG (GR 3711/1-2) and of the Bernstein Center for Computational Neuroscience, Berlin, hosting AH as visiting professor during October 2014. AH acknowledges funding from the European Research Council for support under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 257253.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnsys.2015.00097

Supplemental Data

A GIF animation of the two-dimensional neural field simulation is given as Supplementary Material. Moreover, we deliver a MATLAB software neural field toolbox. This package essentially comprises three routines: `amarikernels`, `amarieq`, `iniamari`, and a main program, `solveamari` to be evoked in the the following way: `amarikernels` is the training program for the synaptic weights. It takes four arguments: *V_patterns*, *sigmarange*, *compbias*, and *contourflag* and returns two kernel tensors $K1, K2$. *V_patterns* is a matrix whose columns are the metastable states in a spatial discretization, their order corresponds to the desired heteroclinic sequence. *sigmarange* is an interval of Lotka-Volterra grow rates σ characterizing the time scale of the dynamics, while *compbias* denotes the competition bias in the interaction matrix (ρ). The last parameter, *contourflag*, is a binary flag deciding whether the hereroclinic sequence is closed (1) or not (0). A closed hereroclinic sequence is called hereroclinic contour. `amarieq` defines the Amari equation (Equation 8) for the ODE solver. It has four input arguments, t, V, $K1, K2$, where t is the time span to be simulated, V is the actual field activity, and $K1, K2$ are the two synaptic weight kernels. `iniamari` prepares an initial condition at the surface of the simplex spanned by the metastable states *V_patterns*. The other arguments are *lead* and *remain*, denoting the leading direction toward the next saddle and its orthogonal projection on the remaining modes. Finally, `solveamari` presents the code for our one-dimensional neural field example, evoking the ODE solver in line 48.

References

Afraimovich, V. S., Rabinovich, M. I., and Varona, P. (2004a). Heteroclinic contours in neural ensembles and the winnerless competition principle. *Int. J. Bifurcat. Chaos* 14, 1195–1208. doi: 10.1142/S0218127404009806

Afraimovich, V. S., Zhigulin, V. P., and Rabinovich, M. I. (2004b). On the origin of reproducible sequential activity in neural circuits. *Chaos* 14, 1123–1129. doi: 10.1063/1.1819625

Allefeld, C., Atmanspacher, H., and Wackermann, J. (2009). Mental states as macrostates emerging from EEG dynamics. *Chaos* 19, 015102. doi: 10.1063/1.3072788

Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* 27, 77–87.

Barrès, V., Simons, A. III, and Arbib, M. (2013). Synthetic event-related potentials: a computational bridge between neurolinguistic models and experiments. *Neural Netw.* 37, 66–92. doi: 10.1016/j.neunet.2012.09.021

beim Graben, P., and Hutt, A. (2014). Attractor and saddle node dynamics in heterogeneous neural fields. *EPJ Nonlin. Biomed. Phys.* 2, 4. doi: 10.1140/epjnbp17

beim Graben, P., and Hutt, A. (2015). Detecting event-related recurrences by symbolic analysis: applications to human language processing. *Philos. Trans. A Math. Phys. Eng. Sci.* A373:20140089. doi: 10.1098/rsta.2014.0089

beim Graben, P., and Potthast, R. (2009). Inverse problems in dynamic cognitive modeling. *Chaos* 19, 015103. doi: 10.1063/1.3097067

beim Graben, P., and Potthast, R. (2012). "A dynamic field account to language-related brain potentials," in *Principles of Brain Dynamics: Global State Interactions, Chapter 5*, eds M. Rabinovich, K. Friston, and P. Varona (Cambridge, MA: MIT Press), 93–112.

Connor, C. E. (2005). Friends and grandmothers. *Nature* 435, 1036–1037. doi: 10.1038/4351036a

Coombes, S., beim Graben, P., Potthast, R., and Wright, J. (eds.). (2014). *Neural Fields: Theory and Applications*. (Berlin: Springer).

Cowan, J. (2014). "A personal account of the development of the field theory of large-scale brain activity from 1945 onward," in *Neural Fields: Theory and Applications, Chapter 2*, eds S. Coombes, P. beim Graben, R. Potthast, and J. Wright (Berlin: Springer), 47–96.

Fukai, T., and Tanaka, S. (1997). A simple neural network exhibiting selective activation of neuronal ensembles: from winner-take-all to winners-share-all. *Neural Comput.* 9, 77–97. doi: 10.1162/neco.1997.9.1.77

Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*, Vol. I *of Lecture Notes of the Santa Fe Institute Studies in the Science of Complexity* (Cambridge, MA: Perseus Books).

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558.

Hudson, A. E., Calderon, D. P., Pfaff, D. W., and Proekt, A. (2014). Recovery of consciousness is mediated by a network of discrete metastable activity states. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9283–9288. doi: 10.1073/pnas.1408296111

Hutt, A., and Riedel, H. (2003). Analysis and modeling of quasi-stationary multivariate time series and their application to middle latency auditory evoked potentials. *Physica D* 177, 203–232. doi: 10.1016/S0167-2789(02)00747-9

Hutt, A., Svensén, M., Kruggel, F., and Friedrich, R. (2000). Detection of fixed points in spatiotemporal signals by a clustering method. *Phys. Rev. E* 61, R4691–R4693. doi: 10.1103/physreve.61.r4691

Hutt, A. (2004). An analytical framework for modeling evoked and event-related potentials. *Int. J. Bifurcat. Chaos* 14, 653–666. doi: 10.1142/S0218127404009351

Igel, C., Erlhagen, W., and Jancke, D. (2001). Optimization of dynamic neural fields. *Neurocomputing* 36, 225–233. doi: 10.1016/S0925-2312(00)00328-3

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Hum. Brain Mapp.* 14, 166–185. doi: 10.1002/hbm.1050

Kandel, E. R., Schwartz, J. H., and Jessel, T. M. (eds.). (1991). *Principles of Neural Science, 3rd Edn*. (East Norwalk, CT: Appleton & Lange).

Kiebel, S. J., von Kriegstein, K., Daunizeau, J., and Friston, K. J. (2009). Recognizing sequences of sequences. *PLoS Comput. Biol.* 5:e1000464. doi: 10.1371/journal.pcbi.1000464

Lehmann, D., Ozaki, H., and Pal, I. (1987). EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. *Electroencephal. Clin. Neurophysiol.* 67, 271–288.

Lehmann, D., Pascual-Marqui, R. D., and Michel, C. (2009). EEG microstates. *Scholarpedia* 4:7632. doi: 10.4249/scholarpedia.7632

Lehmann, D. (1989). "Microstates of the brain in EEG and ERP mapping studies," in *Brain Dynamics*, Vol. 2 *of Springer Series in Brain Dynamics*, eds E. Başar and T. H. Bullock (Berlin: Springer), 72–83.

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., et al. (2002). Dynamic brain sources of visual evoked responses. *Science* 295, 690–694. doi: 10.1126/science.1066168

Mazor, O., and Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* 48, 661–673. doi: 10.1016/j.neuron.2005.09.032

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsaki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science* 321, 1322–1327. doi: 10.1126/science.1159775

Pasupathy, A., and Connor, C. (2002). Population coding of shape in area v4. *Nat. Neurosci.* 5, 1332–1338. doi: 10.1038/972

Potthast, R., and beim Graben, P. (2009). Inverse problems in neural field theory. *SIAM J. Appl. Dynam. Syst.* 8, 1405–1433. doi: 10.1137/080731220

Quiroga, Q. R., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107. doi: 10.1038/nature03687

Rabinovich, M., Volkovskii, A., Lecanda, P., Huerta, R., Abarbanel, H. D. I., and Laurent, G. (2001). Dynamical encoding by networks of competing neuron groups: winnerless competition. *Phys. Rev. Lett.* 87:068102. doi: 10.1103/PhysRevLett.87.068102

Rabinovich, M. I., Huerta, R., and Laurent, G. (2008a). Transient dynamics for neural processing. *Science* 321, 48–50. doi: 10.1126/science.1155564

Rabinovich, M. I., Huerta, R., Varona, P., and Afraimovich, V. S. (2008b). Transient cognitive dynamics, metastability, and decision making. *PLoS Comput. Biol.* 4:e1000072. doi: 10.1371/journal.pcbi.1000072

Rabinovich, M. I., Sokolov, Y., and Kozma, R. (2014a). Robust sequential working memory recall in heterogeneous cognitive networks. *Front. Syst. Neurosci.* 8:220. doi: 10.3389/fnsys.2014.00220

Rabinovich, M. I., Varona, P., Tristan, I., and Afraimovich, V. S. (2014b). Chunking dynamics: heteroclinics in mind. *Front. Comput. Neurosci.* 8:22. doi: 10.3389/fncom.2014.00022

Rissman, J., and Wagner, A. D. (2012). Distributed representations in memory: insights from functional brain imaging. *Ann. Rev. Psych.* 63, 101–128. doi: 10.1146/annurev-psych-120710-100344

Seliger, P., Tsimring, L. S., and Rabinovich, M. I. (2003). Dynamics-based sequential memory: winnerless competition of patterns. *Phys. Rev. E* 67:011905. doi: 10.1103/PhysRevE.67.011905

Veltz, R., and Faugeras, O. (2010). Local/global analysis of the stationary solutions of some neural field equations. *SIAM J. Appl. Dynam. Syst.* 9, 954–998. doi: 10.1137/090773611

Yildiz, I. B., and Kiebel, S. J. (2011). A hierarchical neuronal model for generation and online recognition of birdsongs. *PLoS Comput. Biol.* 7:e1002303. doi: 10.1371/journal.pcbi.1002303

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.

frontiers in Systems Neuroscience

ORIGINAL RESEARCH
published: 17 February 2016
doi: 10.3389/fnsys.2016.00011

Stimuli Reduce the Dimensionality of Cortical Activity

Luca Mazzucato[1], *Alfredo Fontanini*[1,2] *and Giancarlo La Camera*[1,2]*

[1] *Department of Neurobiology and Behavior, State University of New York at Stony Brook, Stony Brook, NY, USA,* [2] *Graduate Program in Neuroscience, State University of New York at Stony Brook, Stony Brook, NY, USA*

The activity of ensembles of simultaneously recorded neurons can be represented as a set of points in the space of firing rates. Even though the dimension of this space is equal to the ensemble size, neural activity can be effectively localized on smaller subspaces. The dimensionality of the neural space is an important determinant of the computational tasks supported by the neural activity. Here, we investigate the dimensionality of neural ensembles from the sensory cortex of alert rats during periods of ongoing (inter-trial) and stimulus-evoked activity. We find that dimensionality grows linearly with ensemble size, and grows significantly faster during ongoing activity compared to evoked activity. We explain these results using a spiking network model based on a clustered architecture. The model captures the difference in growth rate between ongoing and evoked activity and predicts a characteristic scaling with ensemble size that could be tested in high-density multi-electrode recordings. Moreover, we present a simple theory that predicts the existence of an upper bound on dimensionality. This upper bound is inversely proportional to the amount of pair-wise correlations and, compared to a homogeneous network without clusters, it is larger by a factor equal to the number of clusters. The empirical estimation of such bounds depends on the number and duration of trials and is well predicted by the theory. Together, these results provide a framework to analyze neural dimensionality in alert animals, its behavior under stimulus presentation, and its theoretical dependence on ensemble size, number of clusters, and correlations in spiking network models.

Keywords: gustatory cortex, dimensionality, hidden markov models, ongoing activity, mean field theory, spiking network model, metastable dynamics

OPEN ACCESS

Edited by:
Ruben Moreno-Bote,
Universitat Pompeu Fabra, Spain

Reviewed by:
Martin Paul Nawrot,
Universität zu Köln, Germany
Iñigo Arandia-Romero,
Universitat Pompeu Fabra, Spain

***Correspondence:**
Giancarlo La Camera
giancarlo.lacamera@stonybrook.edu

Received: *31 August 2015*
Accepted: *02 February 2016*
Published: *17 February 2016*

Citation:
Mazzucato L, Fontanini A and La Camera G (2016) Stimuli Reduce the Dimensionality of Cortical Activity. Front. Syst. Neurosci. 10:11. doi: 10.3389/fnsys.2016.00011

INTRODUCTION

Understanding the dynamics of neural activity and how it is generated in cortical circuits is a fundamental question in Neuroscience. The spiking activity of ensembles of simultaneously recorded neurons can be represented in terms of sequences of firing rate vectors, as shown e.g., in frontal (Abeles et al., 1995; Seidemann et al., 1996; Durstewitz et al., 2010), gustatory (Jones et al., 2007; Mazzucato et al., 2015), motor (Kemere et al., 2008), premotor and somatosensory cortex (Ponce-Alvarez et al., 2012). The dimension of each firing rate vector is equal to the number of ensemble neurons N and the collection of rate vectors across trials takes the form of a set of points in the N-dimensional space of firing rates. Such points may not fill the whole space, but be restricted to lie inside a lower-dimensional subspace (see Ganguli et al., 2008). Roughly, dimensionality is the minimal number of dimensions necessary to provide an accurate description

of the neural dynamics. If ensemble neurons are independent of each other, neural activities at different times will scatter around in the space of firing rate, filling a large portion of the space. In this case, dimensionality will be maximal and equal to the size of the ensemble N. At the other extreme, if all neurons are strongly correlated, ensemble activity localizes along a line. In this case, dimensionality is minimal and equal to one. These simple examples suggest that dimensionality captures information about the structure of a cortical circuit and the functional relations among the simultaneously recorded neurons, such as their firing rates correlation computed over timescales of hundreds of milliseconds.

Different definitions of dimensionality have been introduced for different tasks and across neural systems (Ganguli et al., 2008; Churchland et al., 2010a; Abbott et al., 2011; Ganguli and Sompolinsky, 2012; Cadieu et al., 2013; Rigotti et al., 2013; Gao and Ganguli, 2015). Such measures of dimensionality can shed light on the underlying neural computation; for example, they can predict the onset of an error trial in a recall task (Rigotti et al., 2013), or can allow the comparison of classification accuracy between different brain areas (e.g., IT vs. V4) and synthetic algorithms (Cadieu et al., 2013). Here, we investigate a measure of dimensionality closely related to the firing rate correlations of simultaneously recorded neurons (Abbott et al., 2011); such correlations may provide a signature of feature-based attention (Cohen and Maunsell, 2009) and other top-down cognitive factors (Nienborg et al., 2012). We elucidate the dependence of dimensionality on experimental parameters, such as ensemble size and interval length, and we show that it varies across experimental conditions. We address these issues by comparing recordings of ensembles of neurons from the gustatory cortex (GC) of alerts rats to a biologically plausible network model based on neural clusters with recurrent connectivity. This model captures neural activity in GC during periods of ongoing and stimulus-evoked activity, explaining how the spatiotemporal dynamics of ensemble activity is organized in sequences of metastable states and how single-neuron firing rate distributions are modulated by stimulus presentation (Mazzucato et al., 2015). Here, we show that the same model expounds the observed dependence of dimensionality on ensemble size and how such dependence is reduced by the presentation of a stimulus. By comparing the clustered network model with a homogeneous network without clusters, we find that the clustered network has a larger dimensionality that depends on the number of clusters and the firing rate correlations among ensemble neurons. A simple theory explains these results and allows extrapolating the scaling of dimensionality to very large ensembles. Our theory shows that recurrent networks with clustered connectivity provide a substrate for high-dimensional neural representations, which may lead to computational advantages.

METHODS

Experimental Procedures

Adult female Long Evans rats were used for this study (Samuelsen et al., 2012; Mazzucato et al., 2015). Animals received *ad lib.* access to food and water, unless otherwise mentioned. Movable bundles of 16 microwires attached to a "mini-microdrive" (Fontanini and Katz, 2006; Samuelsen et al., 2012) were implanted in GC (AP 1.4, ML $\pm$ 5 from bregma, DV –4.5 from dura). After electrode implantation, intra-oral cannulae (IOC) were inserted bilaterally (Phillips and Norgren, 1970; Fontanini and Katz, 2005). At the end of the surgery a positioning bolt for restraint was cemented in the acrylic cap. Rats were given at least 7 days for recovery before starting the behavioral procedures outlined below. All experimental procedures were approved by the Institutional Animal Care and Use Committee of Stony Brook University and complied with University, state, and federal regulations on the care and use of laboratory animals. More details can be found in Samuelsen et al. (2012).

Rats were habituated to being restrained and receiving fluids through IOCs, and then trained to self-deliver water by pressing a lever following a 75 dB auditory cue at a frequency of 4 KHz. The interval at which lever-pressing delivered water was progressively increased to 40 $\pm$ 3 s (ITI). During experimental sessions additional tastants were automatically delivered at random times near the middle of the ITI, at random trials and in the absence of the anticipatory cue. A computer-controlled, pressurized, solenoid-based system delivered $\sim$40 μl of fluids (opening time $\sim$40 ms) directly into the mouth through a manifold of 4 polymide tubes slid into the IOC. The following four tastants were delivered: 100 mM NaCl, 100 mM sucrose, 100 mM citric acid, and 1 mM quinine HCl. Water ($\sim$50 μl) was delivered to rinse the mouth clean through a second IOC 5 s after the delivery of each tastant. Each tastant was delivered for at least 6 trials in each condition. Upon termination of each recording session the electrodes were lowered by at least 150 μm so that a new ensemble could be recorded.

Evoked activity periods were defined as the interval after tastant delivery (time $t = 0$ in our figures) and before water rinse (time $t = 5$ s). Only trials in which the tastants were automatically delivered were considered for the analysis of evoked activity, to minimize the effects of cue-related expectations (Samuelsen et al., 2012). Ongoing activity periods were defined as the 5 s-long intervals at the end of each inter-trial period.

The behavioral state of the rat was monitored during the experiment for signs of disengagement. Erratic lever pressing, inconstant mouth movements and fluids dripping from the mouth indicated disengagement and led to the termination of the experiment. In addition, since disengagement from the task is also reflected in the emergence of high power μ oscillations in local field potentials, occurrences of such periods were removed offline and not analyzed further (Fontanini and Katz, 2008).

Data Analysis

Single neuron action potentials were amplified, bandpass filtered (at 300–8 KHz), digitized and recorded to a computer (Plexon, Dallas, TX). Single units of at least 3:1 signal-to-noise ratio were isolated using a template-matching algorithm, cluster cutting techniques and examination of inter-spike interval plots (Offline Sorter, Plexon, Dallas, TX). All data analyses and model simulations were performed using custom software

written in Matlab (Mathworks, Natick, MA, USA), Mathematica (Wolfram Research, Champaign, IL), and C. Starting from a pool of 299 single neurons in 37 sessions, neurons with peak firing rate lower than 1 Hz (defined as silent) were excluded from further analysis, as well as neurons with a large peak around the 6–10 Hz in the spike power spectrum, which were considered somatosensory (Katz et al., 2001; Samuelsen et al., 2012; Horst and Laubach, 2013). Only ensembles with 3 or more simultaneously recorded neurons were further analyzed (167 non-silent, non-somatosensory neurons from 27 ensembles). We analyzed ongoing activity in the 5 s interval preceding either the auditory cue or taste delivery, and evoked activity in the 5 s interval following taste delivery in trials without anticipatory cue, wherein significant taste-related information is present (Jezzini et al., 2013).

Hidden Markov Model (HMM) Analysis

Here we briefly outline the procedure used in Mazzucato et al. (2015), see this reference and (Jones et al., 2007; Escola et al., 2011; Ponce-Alvarez et al., 2012) for further details. Under the HMM, a system of N recorded neurons is assumed to be in one of a predetermined number of hidden (or latent) states (Rabiner, 1989; Zucchini and MacDonald, 2009). Each state m is defined as a vector of N firing rates $\nu_i(m)$, $i = 1, \ldots, N$, one for each simultaneously recorded neuron. In each state, the neurons were assumed to discharge as stationary Poisson processes (Poisson-HMM). We matched the model to the data segmented in 1-ms bins (see below). In such short bins, we found that typically at most one spike was emitted across all simultaneously recorded neurons. If more than one neuron fired an action potential in a given bin, only one (randomly chosen) was kept for further analysis (this only occurred in a handful of bins per trial; Escola et al., 2011). We denote by $y_i(t)$ the spiking activity of the i-th neuron in the interval $[t, t+dt]$, $y_i(t) = 1$ if the neuron emitted a spike and $y_i(t) = 0$ otherwise. Denoting with S_t the hidden state of the ensemble at time t, the probability of having a spikes from neuron i in a given state m in the interval $[t, t+dt]$ is given by $p\left(y_i(t) = 1 \mid S_t = m\right) = 1 - e^{\nu_i(m)dt}$.

The firing rates $\nu_i(m)$ completely define the states and are also called "emission probabilities" in HMM parlance. The emission and transition probabilities were found by maximization of the log-likelihood of the data given the model via the expectation-maximization (EM), or Baum-Welch, algorithm (Rabiner, 1989), a procedure known as "training the HMM." For each session and type of activity (ongoing vs. evoked), ensemble spiking activity from all trials was binned at 1 ms intervals prior to training assuming a fixed number of hidden states M (Jones et al., 2007; Escola et al., 2011). For each given number of states M, the Baum-Welch algorithm was run 5 times, each time with random initial conditions for the transition and emission probabilities. The range of hidden states M for the HMM analyses were $M_{min} = 10$ and $M_{max} = 20$ for spontaneous activity, and $M_{\min} = 10$ and $M_{\max} = 40$ for evoked activity. Such numbers were based on extensive exploration of the parameter space and previous studies (Jones et al., 2007; Miller and Katz, 2010; Escola et al., 2011; Ponce-Alvarez et al., 2012; Mazzucato et al., 2015). For evoked activity, each HMM was trained on all four tastes simultaneously. Of the models thus obtained, the one with largest total likelihood M^* was taken as the best HMM match to the data, and then used to estimate the probability of the states given the model and the observations in each bin of each trial (a procedure known as "decoding"). During decoding, only those hidden states with probability exceeding 80% in at least 50 consecutive bins were retained (henceforth denoted simply as "states"). State durations were approximately exponentially distributed with median duration 0.60 s (95% CIs: 0.07–4.70) during ongoing activity and 0.30 s (0.06–2.80) during evoked activity (Mazzucato et al., 2015).

The firing rate fits $\nu_i(m)$ in each trial were obtained from the analytical solution of the maximization step of the Baum-Welch algorithm,

$$\nu_i(m) = -\frac{1}{dt} ln\left(1 - \frac{\sum_{t=1}^{T} r_m(t)\, y_i(t)}{\sum_{t=1}^{T} r_m(t)}\right). \tag{1}$$

Here, $[y_i(1), \ldots, y_i(T)]$ is the spike train of the i-th neuron in the current trial, and T is the total duration of the trial. $r_m(t) = P(S_t = m | y(1), \ldots, y(T))$ is the probability that the hidden state S_t at time t is m, given the observations.

Dimensionality Measure

We defined the dimensionality of the neural activity as

$$d = \frac{1}{\sum_{i=1}^{N} \widetilde{\lambda}_i^2}, \tag{2}$$

where the $\widetilde{\lambda}_i$ are the principal eigenvalues expressed as fractions of the total amount of variance explained, i.e., $\widetilde{\lambda}_i = \lambda_i / \left(\sum_j \lambda_j\right)$, where λ_j are the eigenvalues of the covariance matrix of the firing rates (see below).

The dimensionality can be computed exactly in some relevant special cases. The calculation is simplified by the observation that Equation (2) is equivalent to

$$d = \frac{\left[Tr\left(C_f\right)\right]^2}{Tr(C_f^2)},$$

where C_f is the true covariance matrix of the firing rate vectors, $Tr(A) \equiv \sum_{i=1}^{N} A_{ii}$ is the trace of matrix A, and $Tr\left(A^2\right) = \sum_{i,j=1}^{N} A_{ij}A_{ji}$. We consider in the following only the case of firing rates in equal bins, hence we can replace C_f with the covariance matrix of the spike counts C in the definition of d:

$$d = \frac{[Tr(C)]^2}{Tr(C^2)} = \frac{b_N^2}{c_N + a_N}, \tag{3}$$

where for later convenience we have introduced the notation

$$a_N = \sum_{i=1}^{N} C_{ii}^2,\ b_N = \sum_{i=1}^{N} C_{ii},\ c_N = \sum_{i \neq j}^{N} C_{ij}C_{ji}. \tag{4}$$

Note that d does not depend on the distribution of firing rates, but only on their covariance, up to a common scaling factor.

Dimensionality in the case of uniform pair-wise correlations

When all the pair-wise correlations $r_{ij}$are identical, $r_{ij} = \rho$ for all $i \neq j$,

$$r_{ij} = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix} \tag{5}$$

we have $C_{ij} = \rho\sqrt{\sigma_i^2\sigma_j^2}$ for $i \neq j$, where $\sigma_i^2 = C_{ii}$ is the spike count variance. In this case, we find from Equation (4) that

$$a_N = \sum_{i=1}^{N} \sigma_i^4,\ b_N = \sum_{i=1}^{N} \sigma_i^2,\ c_N = \rho^2(b_N^2 - a_N). \tag{6}$$

and the dimensionality, Equation (3), is given by

$$d = \frac{1}{\rho^2 + \left(1-\rho^2\right) g(N)}, \tag{7}$$

where

$$g_N = \frac{a_N}{b_N^2} = \frac{\sum_{i=1}^{N} \sigma_i^4}{\left(\sum_{i=1}^{N} \sigma_i^2\right)^2},$$

Note that since both a_N and b_N scale as N when N is large, in general $g_N \sim \frac{1}{N}$ for large N.

If all spike counts have equal variance, $\sigma_i = \sigma$, we find exactly $g_N = \frac{1}{N}$:

$$d = \frac{1}{\rho^2 + \frac{(1-\rho^2)}{N}} = \frac{N}{N\rho^2 + (1-\rho^2)}, \tag{8}$$

and the dependence of d on the variance drops out. Note that for uncorrelated spike counts ($\rho = 0$) this formula gives $d = N$, whereas for any finite correlation we find the upper bound $d = 1/\rho^2$. For $N > 1$, the dimensionality is inversely related to the amount of pair-wise correlation ρ.

Consider the case where spike counts have variances σ_i^2 drawn from a probability distribution with mean $E\left[\sigma_i^2\right] = \sigma^2$ and variance $Var\left[\sigma_i^2\right] = \delta\sigma^4$, and the pair-wise correlation coefficients r_{ij}, for $i \neq j$, are drawn from a distribution with mean $E\left[r_{ij}\right] = \rho$ and variance $Var\left[r_{ij}\right] = \delta\rho^2$. In such a case one can evaluate Equation (3) approximately by its Taylor expansion around the mean values of the quantities in Equation (4). At leading order in N one finds

$$E\left[d\right] \approx \frac{E\left[b_N^2\right]}{E\left[c_N\right] + E\left[a_N\right]} = \frac{N\sigma^4 + \delta\sigma^4}{(N-1)\,\sigma^4(\rho^2 + \delta\rho^2) + \sigma^4 + \delta\sigma^4}, \tag{9}$$

where $E[.]$ denotes expectation. To obtain this result we have used the definitions in Equation (4), from which

$$\begin{aligned} E[\mathrm{a_N}] &= N(\sigma^4 + \delta\sigma^4),\ E[b_N^2] = N^2\sigma^4 + N\delta\sigma^4, \\ E\left[c_N\right] &= (N^2 - N)\sigma^4(\rho^2 + \delta\rho^2), \end{aligned} \tag{10}$$

and the fact that, given a random vector X_i with mean μ_i and covariance C_{ij}, and a constant symmetric matrix A_{ij}, the expectation value of the quadratic form $\sum_{i,j} X_i A_{ij} X_j$ is

$$E[\sum_{i,j} X_i A_{ij} X_j] = \sum_{ij} (A_{ij} C_{ji} + \mu_i A_{ij} \mu_j). \tag{11}$$

In the case of uncorrelated spike counts ($\rho = 0$, $\delta\rho = 0$), dimensionality still depends linearly on the ensemble size N, but with a smaller slope $\frac{\sigma^4}{\sigma^4 + \delta\sigma^4} < 1$ compared to the case of equal variances (Equation 8 with $\rho = 0$).

Dimensionality in the Case of Neural Clusters

Given an ensemble of N neurons arranged in Q clusters (motivated by the model network described later in section "Spiking neuron model"), we created ensembles of uncorrelated spike trains for $N \leq Q$ and correlated within each cluster for $N > Q$. Thus, if $N \leq Q$ the correlation matrix is the $N \times N$ identity matrix. If $N > Q$, the (Q+1)th neuron was added to the first cluster, with correlation ρ with the other neuron of the cluster, and uncorrelated to the neurons in the remaining clusters. The (Q+2)th neuron was added to the second cluster, with correlation ρ with the other neuron of the second cluster, and uncorrelated to the neurons in the remaining clusters, and so on. Similarly, the (2Q+p)th neuron ($p \leq Q$) was added to the p-th cluster, with pair-wise correlation ρ with the other neurons of the same cluster, but no correlation with the neurons in the remaining clusters; and so on. In general, for $N = mQ + p$ neurons (where $m = \left[\frac{N}{Q}\right]_- \geq 1$ is the largest integer smaller than $\frac{N}{Q}$), the procedure picked $m + 1$ neurons per cluster for the first p cluster and m neurons per cluster for the remaining $Q - p$ clusters, with uniform pair-wise correlations ρ in the same cluster while neurons from different clusters were uncorrelated. The resulting correlation matrix r was block diagonal

$$r = diag\left(R_1, \dots, R_Q\right),$$

where each of the Q blocks contains the correlations of neurons from the same cluster. Inside each block R_i, the off-diagonal terms are equal to the uniform within-cluster correlation ρ:

$$R_i = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

The first p blocks have size $(m+1) \times (m+1)$ and the last $Q - p$ blocks have size $m \times m$, so that $(m+1)\,p + m\left(Q - p\right) = N$. The remaining elements of matrix r (representing pair-wise correlations of neurons belonging to different clusters) were all zero. Recalling that $C_{ij} = r_{ij}\sigma_i\sigma_j$, one finds $Tr\,(C) = pb_{m+1} + \left(Q - p\right) b_m$ and$Tr\left(C^2\right) = \rho^2[pb_{m+1}^2 + \left(Q - p\right) b_m^2] +$

$(1-\rho^2)[pa_{m+1}+(Q-p)a_m]$, where a_n and b_n are defined in Equation (6), from which one obtains

$$d=\begin{cases} b_N^2/a_N & ,\ N\le Q \\ \frac{[pb_{m+1}+(Q-p)b_m]^2}{\rho^2[pb_{m+1}^2+(Q-p)b_m^2]+(1-\rho^2)[pa_{m+1}+(Q-p)a_m]} & ,\ N>Q \end{cases} \quad (12)$$

In the approximation where all neurons have the same variance this simplifies to

$$d=\begin{cases} N & ,\ N\le Q \\ \frac{N}{1+m\rho^2[1-(Q-p)/N]} & ,\ N>Q. \end{cases} \quad (13)$$

Recall that in the formulae above m and p depend on N. For finite ρ, Equation (13) predicts the bound $d \le Q/\rho^2$ for any $N > 1$, with this value reached asymptotically for large N. When single neuron variances σ_i^2 are drawn from a distribution with mean $E\left[\sigma_i^2\right]=\sigma^2$ and variance $Var\left[\sigma_i^2\right]=\delta\sigma^4$, an expression for the dimensionality can be obtained from Equation (12) at leading order in the expectation values of the quantities in Equation (4) (not shown), with a procedure similar to that used to obtain Equation (9).

Pair-Wise Correlations

Given neuron i and neuron j's spike trains, we computed the spike count correlation coefficient r_{ij}

$$r_{ij}=\frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}},$$

where S is the sample covariance matrix of the spike counts estimated as

$$S_{ij}=\frac{1}{N_bN_T-1}\sum_{b,s=1}^{N_b,N_T}\left(n_i\left(b,s\right)-<n_i>\right)\left(n_j\left(b,s\right)-<n_j>\right), \quad (14)$$

where $n_i(b, s)$ is the spike count of neuron i in bin b and trial s. The sum goes over all N_b bins and over all N_T trials in a session, whereas $< n_i >$ is the average across trials and bins for neuron i. In the main text and figures we present results obtained with a bin size of 200 ms, but have performed the same analyses with bin sizes varying from 10 ms to 5 s (see Results for details).

Significance of the correlation was estimated as follows (Renart et al., 2010): $N_{shuffle} = 200$ trial-shuffled correlation coefficients $r_{ij}^{'}$ were computed, then a p-value was determined as the fraction of shuffled coefficients $r_{ij}^{'}$ whose absolute value exceeded the absolute value of the experimental correlation, $p=\frac{\#\left(\left|r_{ij}^{'}\right|>|r_{ij}|\right)}{N_{shuffle}}$. For example, a correlation r was significant at $p = 0.05$ confidence level if no more than 10 shuffled correlation coefficients out of 200 exceeded r.

The pair-wise correlations of firing rates vectors computed in bins of fixed duration T were given by Equation (14) with $n_i\left(b,s\right)$ replaced by $n_i\left(b,s\right)/T$. Instead, correlations of firing rates vectors inside hidden states (which have variable duration) were estimated after replacing $n_i\left(b,s\right)$ in Equation (14) with $\nu_i\,(m,s)$, the firing rate of neuron i in state m in trial s. For each trial s, this quantity was computed according to Equation (1).

Estimation of Dimensionality

The eigenvalues λ_j in Equation (2) were found with a standard Principal Component Analysis (PCA) of the set of all firing rate vectors (Chapin and Nicolelis, 1999). The firing rate vectors were obtained via the HMM analysis (see Equation 1); all data from either ongoing or evoked activity were used. For the analysis of **Figure 3E**, where the duration and number of trials were varied, only the firing rate vectors of the HMM states present in the given trial snippet were used (even if present for only a few ms). When firing rate vectors in hidden states were not available (mainly, in "shuffled" datasets and in asynchronous homogeneous networks, see below for details), the firing rates were computed as spike counts in $T = 200$ ms bins divided by T, $n_i\left(b,s\right)/T$, where $n_i\left(b,s\right)$ is as defined in Equation (14) (**Figures 3F,G**, **6E**, **7D**, **9A**). Dimensionality values were averaged across 20 simulated sessions for each ensemble size N; in each session, 40 trials of 5 s duration, resulting in $N_T = 1{,}000$ bins, were used (using bin widths of 50–500 ms did not change the results). Note that for the purpose of computing the dimensionality (Equation 3), it is equivalent to use either the binned firing rate $n_i\left(b,s\right)/T$ or the spike count $n_i\left(b,s\right)$.

In our data, d roughly corresponded to the number of principal components explaining between 80 and 90% of the variance. However, note that all eigenvalues are retained in our definition of dimensionality given in Equation (2) above.

Shuffled Datasets

The dimensionality of the data as a function of ensemble size N was validated against surrogate datasets constructed by shuffling neurons across different sessions while matching the empirical distribution of ensemble sizes. Comparison analyses between empirical and shuffled ensembles were trial-matched using the minimal number of trials per condition across ensembles, and then tested for significant difference with the Mann-Whitney test on samples obtained from 20 bootstrapped ensembles. Neurons whose firing rate variance exceeded the population average by two standard deviations were excluded (8/167 of non-silent, non-somatosensory neurons).

Dependence on the Number of Trials: Simulations (Figures 7E, 8A)

The estimate of d from data depends on the number and duration of the trials (**Figure 3E** and Equation 16 below). To investigate this phenomenon in a simple numerical setting we generated $N \times N_T$ "nominal" firing rates, thought of as originating from N neurons, each sampled N_T times (trials). The single firing rates were sampled according to a log-normal distribution with equal means and covariance leading to Equation (7), i.e., $C_{ij} = \rho\sigma_i\sigma_j(1-\delta_{ij}) + \sigma_i^2\delta_{ij}$, with $\delta_{ij} = 1$ if $i = j$, and zero otherwise (note that the actual distribution used is immaterial since the dimensionality only depends on the covariance matrix, see Equation 3). We considered the two cases of equal variance for all ensemble neurons, $\sigma_i = \sigma$ for all i (**Figure 8A**) or variances σ_i sampled from a log-normal distribution (**Figure 8A** and "+" in **Figure 7E**). The same N and N_T as used for the analysis of the model simulations in **Figure 7D** were used (where the "trials" were N_T bins of 200 ms in 40 intervals of 5 second

duration for each ensemble size N). The covariance of the data thus generated was estimated according to Equation (14), based on which the dimensionality Equation (3) was computed. The estimated dimensionality depends on N and N_T and was averaged across 100 values of d, each obtained as explained above. Note that in this simplified setting increasing the duration of each trial is equivalent to adding more trials, i.e., the effect of having a trial 400 ms long producing 2 firing rates (one for each 200 ms bin) is equivalent to having two trials of 200 ms duration. In the general case, the effect of trial duration on d will depend on how trial duration affects the variance and correlations of the firing rates.

Dependence on the Number of Trials: Theory

The dependence of dimensionality on the number of trials can be computed analytically under the assumption that N ensemble neurons generate spike counts n_i, for $i = 1, \ldots, N$, distributed according to a multivariate Gaussian. Since we are interested in the spike-count covariance Equation (14), we can assume the spike-count distribution to have zero mean and true covariance C_{ij}. The matrix $M^{(N_T)} = (N_T - 1) \cdot S^{(N_T)}$, where $S^{(N_T)}$ is the covariance matrix Equation (14) sampled from N_T trials, is distributed according to a Wishart distribution $W_N\left(C_{ij}, N_T - 1\right)$ with $N_T - 1$ degrees of freedom (Mardia et al., 1979). Since the variance of the Wishart distribution,

$$Var\left(M_{ij}\right) = (N_T - 1)(C_{ij}^2 + C_{ii}C_{jj}),$$

is proportional to N_T, we obtain the variance of the entries of the sample covariance as

$$Var\left(S_{ij}^{(N_T)}\right) = \frac{C_{ij}^2 + C_{ii}C_{jj}}{N_T - 1}, \tag{15}$$

to be used in the estimator of d (from Equation 3)

$$\hat{d} = \frac{[Tr\,(S)]^2}{Tr(S^2)} = \frac{\widehat{b_N^2}}{\widehat{c_N} + \widehat{a_N}},$$

where $\widehat{a_N}$, $\widehat{c_N}$, $\widehat{b_N^2}$ are given by Equation (4) with C replaced by S. With a calculation similar to that used to obtain Equation (9), to leading order in N and N_T one finds

$$E[\hat{d}] \approx \frac{E[\widehat{b_N^2}]}{E[\widehat{c_N}] + E[\widehat{a_N}]},$$

with

$$E[\widehat{a_N}] = N\left(\sigma^4 + \delta\sigma^4\right) + \frac{2N\sigma^4}{N_T - 1},$$

$$E[\widehat{b_N^2}] = N^2\sigma^4 + N\delta\sigma^4 + \frac{2N\sigma^4}{N_T - 1},$$

$$E\left[\widehat{c_N}\right] = \left(N^2 - N\right)\left(\rho^2 + \delta\rho^2\right)\sigma^4 + \left(N^2 - N\right)\frac{1 + \rho^2 + \delta\rho^2}{N_T - 1}\sigma^4,$$

where we also used Equations (10) and (11), with $Var\left[\sigma_i^2\right] = \delta\sigma^4$ and $Var\left[r_{ij}\right] = \delta\rho^2$, for $i \neq j$. In conclusion, one finds

$$E[\hat{d}] = \frac{\left(N^2 + \frac{2N}{N_T - 1}\right)\sigma^4 + N\delta\sigma^4}{\left(N^2 - N\right)\left(\rho^2 + \delta\rho^2 + \frac{1 + \rho^2 + \delta\rho^2}{N_T - 1}\right)\sigma^4 + N\left(1 + \frac{2}{N_T - 1}\right)\sigma^4 + N\delta\sigma^4}. \tag{16}$$

Model Fitting

The dependence of the data's dimensionality on ensemble size N was fitted by a straight line via standard least-squares,

$$d = \beta_1 \cdot N + \beta_0,$$

separately for ongoing and evoked activity (**Figures 3B–D**, **6B–D**). Comparison between the dimensionality of evoked and ongoing activity was carried out with a 2-way ANOVA with condition (evoked vs. ongoing) and ensemble size (N) as factors. Since d depends on the number and duration of the trials used to estimate the covariance matrix (**Figure 3E** and Equation 16), we matched both the number of trials and trial length in comparisons of ongoing and evoked dimensionality. If multiple tastes were used, the evoked trials were each matched to a random subset of an equal number of ongoing trials.

The dependence of dimensionality d on ensemble size N in a surrogate dataset of Poisson spike trains with mean pairwise correlation ρ (generated according to the algorithm described in the next section) was modeled as Equation (16) with $\delta\rho^2 = \alpha\rho^2$ and $\delta\sigma^4 = \sigma^4 = \beta$ (**Figure 7D**, dashed lines); N_T was fixed to 1000 (40 trials of 5 s each, segmented in 200 ms bins). The parameters α, β were tuned to fit all Poisson trains simultaneously on datasets with $N = 5, 10, \ldots, 100$ and $\rho = 0, 0.01, 0.05, 0.1, 0.2$, with 20 ensembles for each value (**Figure 7D**; only the fits for $\rho = 0, 0.1, 0.2$ are shown). A standard non-linear least-squares procedure was used (Holland and Welsch, 1977).

Generation of Correlated Poisson Spike Trains

Ensembles of independent and correlated Poisson spike trains were generated for the analysis of **Figure 7**. Ensembles of independent stationary Poisson spike trains with given firing rates ν_i were generated by producing their interspike intervals according to an exponential distribution with parameter ν_i. Stationary Poisson spike trains with fixed pairwise correlations (but no temporal correlations) were generated according to the method reported in Macke et al. (2009), that we briefly outline below.

We split each trial into 1 ms bins and consider the associated binary random variable $X_i(t) = 1$ if the i-th neuron emitted a spike in the t-th bin, and $X_i(t) = 0$ if no spike was emitted. These samples were obtained by first drawing a sample from an auxiliary N-dimensional Gaussian random variable $\mathcal{U} \sim \mathcal{N}(\gamma, \Lambda)$ and then thresholding it into 0 and 1: $X_i = 1$ if $\mathcal{U}_i > 0$, and $X_i = 0$ otherwise. Here, $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_N\}$ is the mean vector and $\Lambda = \left\{\Lambda_{ij}\right\}$ is the covariance matrix of the N-dimensional Gaussian variable $\mathcal{U}$. For appropriately chosen

parameters γ_i and Λ_{ij} the method generates correlated spike trains with the desired firing rates ν_i and pairwise spike count correlation coefficients r_{ij}.

The prescription for γ_i and Λ_{ij} is most easily expressed as a function of the desired probabilities μ_i of having a spike in a bin of width dt, $\mu_i = P(X_i(t) = 1)$, and the pairwise covariance c_{ij} of the random binary vectors $X_i(t)$ and $X_j(t)$, from which γ_i and Λ_{ij} can be obtained by inverting the following relationships:

$$\begin{aligned} \mu_i &= \Phi(\gamma_i), \\ c_{ii} &= \Phi(\gamma_i)\,\Phi(-\gamma_i), \\ c_{ij} &= \Phi_2\left(\gamma_i, \gamma_j, \Lambda_{ij}\right) - \Phi(\gamma_i)\,\Phi\left(\gamma_j\right), \; i \neq j. \end{aligned}$$

Here, $\Phi(x)$ is the cumulative distribution of a univariate Gaussian with mean 0 and variance 1 evaluated at x, and $\Phi_2\left(x, y, \Lambda\right)$ is the cumulative distribution of a bivariate Gaussian with means 0, variances 1 and covariance Λ evaluated at (x, y) (note that the distributions Φ and Φ_2 are unrelated to the N-dimensional Gaussian $\mathcal{U} \sim \mathcal{N}(\gamma, \Lambda)$). Without loss of generality we imposed unit variances for $\mathcal{U}_i$, i.e., $\Lambda_{ii} = 1$.

We related the spike probabilities μ_i to the firing rates ν_i as $\mu_i = 1 - e^{-\nu_i dt}$, with $(1 - \mu_i)$ being the probability of no spikes in the same bin. When dt approaches zero, $\mu_i \approx \nu_i dt$ and the spike trains generated as vectors of binary random variables by sampling $\mathcal{U} \sim \mathcal{N}(\gamma, \Lambda)$ will approximate Poisson spike trains ($dt = 1$ ms bins were used). In order to have a fair comparison with the data generated by the spiking network model (described in the next section), the mean firing rates of the Poisson spike trains were matched to the average firing rates obtained from the simulated data.

Since γ *and* Λ were the same in all bins, values of $X_i(t)$ and $X_i(s)$ were independent for $t \neq s$ (i.e., the spike trains had no temporal correlations). As a consequence, the random binary vectors have the same pair-wise correlations as the spike counts, and the c_{ij} are related to the desired r_{ij} by $c_{ij} = r_{ij}\sqrt{\mu_i(1-\mu_i)\,\mu_j\left(1-\mu_j\right)}$, where $\mu_i(1-\mu_i)$ is the variance of X_i. See Macke et al. (2009) for further details.

Spiking Network Model

We modeled the data with a recurrent spiking network of $N = 5000$ randomly connected leaky integrate-and-fire (LIF) neurons, of which 4000 excitatory (E) and 1000 inhibitory (I). Connection probability $p_{\beta\alpha}$ from neurons in population $\alpha \in E, I$ to neurons in population $\beta \in E, I$ were $p_{EE} = 0.2$ and $p_{EI} = p_{IE} = p_{II} = 0.5$; a fraction $f = 0.9$ of excitatory neurons were arranged into Q different clusters, with the remaining neurons belonging to an unstructured ("background") population (Amit and Brunel, 1997). Synaptic weights $J_{\beta\alpha}$ from neurons in population $\alpha \in E, I$ to neurons in population $\beta \in E, I$ scaled with N as $J_{\beta\alpha} = j_{\beta\alpha}/\sqrt{N}$, with $j_{\beta\alpha}$ constants having the following values (units of mV): $j_{EI} = 3.18$, $j_{IE} = 1.06$, $j_{II} = 4.24$, $j_{EE} = 1.77$. Within an excitatory cluster synaptic weights were potentiated, i.e., they took average values of $\langle J\rangle_+ = J_+ j_{EE}$ with $J_+ > 1$, while synaptic weights between units belonging to different clusters were depressed to average values $\langle J\rangle_- = J_- j_{EE}$, with $J_- = 1 - \gamma f(J_+ - 1) < 1$, with $\gamma = 0.5$. The latter relationship between J_+ and J_- helps to maintain balance between overall potentiation and depression in the network (Amit and Brunel, 1997).

Below spike threshold, the membrane potential V of each LIF neuron evolved according to

$$\tau_m \frac{dV}{dt} = -V + \tau_m\left(I_{rec} + I_{ext} + I_{stim}\right)$$

with a membrane time constant $\tau_m = 20$ ms for excitatory and 10 ms for inhibitory units. The input current was the sum of a recurrent input I_{rec}, an external current I_{ext} representing an ongoing afferent input from other areas, and an external stimulus I_{stim} representing e.g., a delivered taste during evoked activity only. In our units, a membrane capacitance of 1nF is set to 1. A spike was said to be emitted when V crossed a threshold V_{thr}, after which V was reset to a potential $V_{reset} = 0$ for a refractory period of $\tau_{ref} = 5$ ms. Spike thresholds were chosen so that, in the unstructured network (i.e., with $J_+ = J_- = 1$), the E and I populations had average firing rates of 3 and 5 spikes/s, respectively (Amit and Brunel, 1997). The recurrent synaptic input I^i_{rec} to unit i evolved according to the dynamical equation

$$\tau_s \frac{dI^i_{rec}}{dt} = -I^i_{rec} + \sum_{j=1}^{N} J_{ij} \sum_k \delta\left(t - t^j_k\right),$$

where t^j_k was the arrival time of k-th spike from the j-th pre-synaptic unit, and τ_s was the synaptic time constant (3 and 2 ms for E and I units, respectively), resulting in an exponential post-synaptic current in response to a single spike, $\frac{J_{ij}}{\tau_s}\exp(-t/\tau_s)\Theta(t)$, where $\Theta(t) = 1$ for $t \geq 0$, and $\Theta(t) = 0$ otherwise. The ongoing external current to a neuron in population α was constant and given by $I_{ext} = N_{ext} p_{\alpha 0} J_{\alpha 0} \nu_{ext}$, where $N_{ext} = n_E N$, $p_{\alpha 0} = p_{EE}$, $J_{\alpha 0} = \frac{j_{\alpha 0}}{\sqrt{N}}$ with $j_{E0} = 0.3$, $j_{I0} = 0.1$, and $\nu_{ext} = 7$ spikes/s. During evoked activity, stimulus-selective units received an additional input representing one of the four incoming stimuli. The stimuli targeted combinations of neurons as observed in the data. Specifically, the fractions of neurons responsive to $n = 1, 2, 3$ or all 4 stimuli were 17% (27/162), 22% (36/162), 26% (42/162), and 35% (57/162) (Jezzini et al., 2013; Mazzucato et al., 2015). Each stimulus had constant amplitude ν_{stim} ranging from 0 to 0.5 ν_{ext}. In the following we measure the stimulus amplitude as percentage of ν_{ext} (e.g., "10%" corresponds to $\nu_{stim} = 0.1\ \nu_{ext}$). The onset of each stimulus was always $t = 0$, the time of taste delivery. The stimulus current to a unit in population α was constant and given by $I_{stim} = N_{ext} p_{\alpha 0} J_{\alpha 0} \nu_{stim}$.

Mean Field Analysis of the Model

The stationary states of the spiking network model in the limit of large N were found with a mean field analysis (Amit and Brunel, 1997; Brunel and Hakim, 1999; Fusi and Mattia, 1999; Curti et al., 2004; Mazzucato et al., 2015). Under typical conditions, each neuron of the network receives a large number of small post-synaptic currents (PSCs) per integration time constant. In such a case, the dynamics of the network can be analyzed under the diffusion approximation within the population density approach. The network has $\alpha = 1, \ldots, Q + 2$ sub-populations, where the

first Q indices label the Q excitatory clusters, $\alpha = Q+1$ labels the "background" units, and $\alpha = Q+2$ labels the homogeneous inhibitory population. In the diffusion approximation (Tuckwell, 1988; Lánský and Sato, 1999; Richardson, 2004), the input to each neuron is completely characterized by the infinitesimal mean μ_α and variance σ_α^2 of the post-synaptic potential (see Mazzucato et al., 2015 for the expressions of the infinitesimal mean and variance for all subpopulations).

Parameters were chosen so that the network with $J_+ = J_- = 1$ (where all $E \to E$ synaptic weights are equal) would operate in the balanced asynchronous regime (van Vreeswijk and Sompolinsky, 1996, 1998; Renart et al., 2010), where incoming contributions from excitatory and inhibitory inputs balance out, neurons fire irregular spike trains with weak pair-wise correlations.

The unstructured network has only one dynamical state, i.e., a stationary point of activity where all E and I neurons have constant firing rate ν_E and ν_I, respectively. In the structured network (where $J_+ > 1$), the network undergoes continuous transitions among a repertoire of states, as shown in the main text. To avoid confusion between network activity states and HMM states, we refer to the former as network "configurations" instead of states. Admissible networks configurations must satisfy the $Q+2$ self-consistent mean field equations (Amit and Brunel, 1997)

$$\nu_\alpha = F_\alpha\left(\mu_\alpha(\vec{\nu}), \sigma_\alpha^2(\vec{\nu})\right),$$

where $\vec{\nu} = \left[\nu_1, \ldots, \nu_Q, \nu_E^{bg}, \nu_I\right]$ is the firing rate vector and $F_\alpha\left(\mu_\alpha, \sigma_\alpha^2\right)$ is the current-to-rate response function of the LIF neurons. For fast synaptic times, i.e., $\frac{\tau_s}{\tau_m} \ll 1$, $F_\alpha\left(\mu_\alpha, \sigma_\alpha^2\right)$ is well approximated by (Brunel and Sergi, 1998; Fourcaud and Brunel, 2002)

$$F_\alpha(\mu_\alpha, \sigma_\alpha) = \left(\tau_{ref} + \tau_{m,\alpha}\sqrt{\pi}\int_{H_{eff,\alpha}}^{\Theta_{eff,\alpha}} e^{u^2}\left[1+\mathrm{erf}(u)\right]\right)^{-1},$$

where

$$\Theta_{eff,\alpha} = \frac{V_{thr,\alpha} - \mu_\alpha}{\sigma_\alpha} + a k_\alpha,$$

$$H_{eff,\alpha} = \frac{V_{reset,\alpha} - \mu_\alpha}{\sigma_\alpha} + a k_\alpha,$$

where $k_\alpha = \sqrt{\tau_{s,\alpha}/\tau_{m,\alpha}}$ is the square root of the ratio of synaptic time constant to membrane time constant, and $a = \frac{|\zeta(1/2)|}{\sqrt{2}} \sim 1.03$. This theoretical response function has been fitted successfully to the firing rate of neocortical neurons in the presence of *in vivo*-like fluctuations (Rauch et al., 2003; Giugliano et al., 2004; La Camera et al., 2006, 2008).

The fixed points $\vec{\nu}^*$ of the mean field equations were found with Newton's method (Press et al., 2007). The fixed points can be either stable (attractors) or unstable depending on the eigenvalues λ_α of the stability matrix

$$S_{\alpha\beta} = \frac{1}{\tau_{s,\alpha}}\left(\frac{\partial F_\alpha\left(\mu_\alpha(\vec{\nu}), \sigma_\alpha^2(\vec{\nu})\right)}{\partial \nu_\beta}\right) - \left(\frac{\partial F_\alpha\left(\mu_\alpha(\vec{\nu}), \sigma_\alpha^2(\vec{\nu})\right)}{\partial \sigma_\alpha^2}\frac{\partial \sigma_\alpha^2}{\partial \nu_\beta} - \delta_{\alpha\beta}\right),$$

evaluated at the fixed point $\vec{\nu}^*$ (Mascaro and Amit, 1999). If all eigenvalues have negative real part, the fixed point is stable (attractor). If at least one eigenvalue has positive real part, the fixed point is unstable. Stability is meant with respect to an approximate linearized dynamics of the mean and variance of the input current:

$$\tau_{s,\alpha}\frac{dm_\alpha}{dt} = -m_\alpha + \mu_\alpha(\vec{\nu})$$

$$\frac{\tau_{s,\alpha}}{2}\frac{ds_\alpha^2}{dt} = -s_\alpha^2 + \sigma_\alpha^2(\vec{\nu})$$

$$\nu_\alpha(t) = F_\alpha\left(m_\alpha(\vec{\nu}), s_\alpha^2(\vec{\nu})\right),$$

where μ_α and σ_α^2 are the stationary values for fixed $\vec{\nu}$ given earlier. For fast synaptic dynamics in the asynchronous balanced regime, these rate dynamics are in very good agreement with simulations (La Camera et al., 2004—see Renart et al., 2004; Giugliano et al., 2008 for more detailed discussions).

Metastable Configurations in the Network Model

The stable configurations of a network with an infinite number of neurons were obtained in the mean field approximation of the previous section and are shown in **Figure 4B** for $Q = 30$ and a range of values of the relative potentiation parameter J_+. Above the critical point $J_+ = 4.2$, stable configurations characterized by a finite number of active clusters emerge (gray lines; the number of active clusters is reported next to each line). For a given J_+, the firing rate is the same in all active clusters and is inversely proportional to the total number of active clusters. Stable patterns of firing rates are also found in the inhibitory population (red lines), in the inactive clusters (having low firing rates; gray dashed lines), and in the unstructured excitatory population (dashed blue lines). For a fixed value of J_+, multiple stable configurations coexist with different numbers of active clusters. For example, for $J_+ = 5.3$, configurations with up to 7 active clusters are stable, each configuration with different firing rates. This generates multistable firing rates in single neurons, i.e., the property, also observed in the data, that single neurons can attain more than 2 firing rates across states (Mazzucato et al., 2015). Note that if $J_+ \leq 5.15$ an alternative stable configuration of the network with all clusters inactive (firing rates < 10 spikes/s) is also possible (single brown line).

Strictly speaking, the configurations in **Figure 4B** are stable only in a network containing an infinite number of uncorrelated neurons. In a finite network (or when neurons are strongly correlated) these configurations can lose stability due to strong fluctuations, which ignite transitions among the different configurations. Full details are reported in Mazzucato et al. (2015).

Model Simulations and Analysis of Simulated Data

The dynamical equations of the LIF neurons were integrated with the Euler algorithm with a time step of $dt = 0.1$ ms. We simulated 20 different networks (referred to as "sessions" in the following) during both ongoing and evoked activity. We chose four different stimuli per session during evoked activity (to mimic taste delivery). Trials were 5 s long. The HMM analyses for **Figures 2**, **5** were performed on ensembles of randomly selected excitatory neurons with the same procedure used for the data (see previous section "Hidden Markov Model (HMM) analysis"). The ensemble sizes were chosen so as to match the empirical ensemble sizes (3–9 randomly selected neurons). For the analysis of **Figure 9A**, ensembles of increasing size (from 5 to 100 neurons) were used from simulations with $Q = 30$ clusters. When the ensemble size was less than the number of clusters ($N \leq Q$), each neuron was selected randomly from a different cluster; when ensemble size was larger than the number of clusters, one neuron was added to each cluster until all clusters were represented, and so on until all N neurons had been chosen. To allow comparison with surrogate Poisson spike trains, the dimensionality of the simulated data was computed from the firing rate vectors in $T = 200$ ms bins as explained in section "Dimensionality measure." For control, the dimensionality was also computed from the firing rate vectors in hidden states obtained from an HMM analysis, obtaining qualitatively similar results.

RESULTS

Dimensionality of the Neural Activity

We investigate the dimensionality of sequences of firing rate vectors generated in the GC of alert rats during periods of ongoing or evoked activity (see Methods). To provide an intuitive picture of the meaning of dimensionality adopted in this paper, consider the firing rate vectors from N simultaneously recorded neurons. These vectors can occupy, a priori, the entire N-dimensional vector space minimally required to describe the population activity of N independent neurons. However, the sequence of firing rate vectors generated by the neural dynamics may occupy a subspace that is spanned by a smaller number $m < N$ of coordinate axes. For example, the data obtained by the ensemble of three simulated spike counts in **Figure 1** mostly lie on a 2D space, the plane shaded in gray. Although 3 coordinates are still required to specify all data points, a reduced representation of the data, such as that obtained from PCA, would quantify the dimension of the relevant subspace as being close to 2. To quantify this fact we use the following definition of dimensionality (Abbott et al., 2011)

$$d = \left(\sum_{i=1}^{N} \widetilde{\lambda}_i^2 \right)^{-1},$$

where N is the ensemble size and $\widetilde{\lambda}_i$ are the normalized eigenvalues of the covariance matrix, each expressing the fraction of the variance explained by the corresponding principal component (see Methods for details). According to this formula, if the first n eigenvalues express each a fraction $1/n$ of the variance while the remaining eigenvalues vanish, the dimensionality is $d = n$. In less symmetric situations, d reflects roughly the dimension of the linear subspace explaining most variance about all data points. In the example of the data on the gray plane of **Figure 1**, $d = 1.8$, which is close to 2, as expected. Similarly, data points lying mostly along the blue and red straight lines in **Figure 1** have a dimensionality of 0.9, close to 1. In all cases, $d > 0$ and $d \leq N$, where N is the ensemble size.

The blue and red data points in **Figure 1** were obtained from a fictitious scenario where neuron 1 and neuron 2 were selective to surrogate stimuli A and B, respectively, and are meant to mimic two possible evoked responses. The subspace containing responses to both stimuli A and B would have a dimensionality $d_{A+B} = 1.7$, similar to the dimensionality of the data points distributed on the gray plane (meant instead to represent spike counts during ongoing activity in the same fictitious scenario). Thus, a dimensionality close to 2 could originate from different patterns of activity, such as occupying a plane or two straight lines. Other and more complex scenarios are, of course, possible. In general, the dimensionality

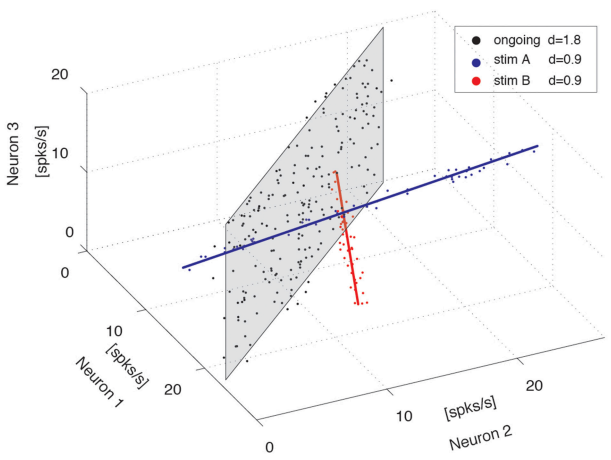


FIGURE 1 | Dimensionality of the neural representation. Pictorial representation of the firing rate activity of an ensemble of $N = 3$ neurons. Each dot represents a three-dimensional vector of ensemble firing rates in one trial. Ensemble ongoing activity localizes around a plane (black dots cloud surrounding the shaded black plane), yielding a dimensionality of $d = 1.8$. Activity evoked by each of two different stimuli localizes around a line (red and blue dots clouds and lines), yielding a dimensionality of $d = 0.9$ in both cases.

will reflect existing functional relationships among ensemble neurons (such as pair-wise correlations) as well as the response properties of the same neurons to external stimuli. The pictorial example of **Figure 1** caricatures a stimulus-induced reduction of dimensionality, as found in the activity of simultaneously recorded neurons from the GC of alert rats, as we show next.

Dimensionality is Proportional to Ensemble Size

We computed the dimensionality of the neural activity of ensembles of 3–9 simultaneously recorded neurons in the gustatory cortex of alert rats during the 5 s inter-trial period preceding (ongoing activity) and following (evoked activity) the delivery of a taste stimulus (said to occur at time $t = 0$; see Methods). Ensemble activity in single trials during both ongoing (**Figure 2A**) and evoked activity (**Figure 2B**) could be characterized in terms of sequences of metastable states, where each state is defined as a collection of firing rates across simultaneously recorded neurons (Jones et al., 2007; Mazzucato et al., 2015). Transitions between consecutive states were detected via a Hidden Markov Model (HMM) analysis, which provides the probability that the network is in a certain state at every 1 ms bin (**Figure 2**, color-coded lines superimposed to raster plots). The ensemble of spike trains was considered to be in a given state if the posterior probability of being in that state exceeded 80% in at least 50 consecutive 1-ms bins (**Figure 2**, color-coded shaded areas). Transitions among states were triggered by the co-modulation of a variable number of ensemble neurons and occurred at seemingly random times (Mazzucato et al., 2015). For this reason, the dimensionality of the neural activity was computed based on the firing rate vectors in each HMM state (one firing rate vector per state per trial; see Methods for details).

The average dimensionality of ongoing activity across sessions was $d_{ongoing} = 2.6 \pm 1.2$ (mean±SD; range: [1.2, 5.0]; 27 sessions). An example of the eigenvalues for a representative ensemble of eight neurons is shown in **Figure 3A**, where $d = 4.42$. The dimensionality of ongoing activity was approximately linearly related to ensemble size (**Figure 3B**, linear regression, $r = 0.4$, slope $b_{ongoing} = 0.26 \pm 0.12$, $p = 0.04$). During evoked activity dimensionality did not differ across stimuli (one-way ANOVA, no significant difference across tastants, $p > 0.8$), hence all evoked data points were combined for further analysis. An example of the eigenvalue distribution of the ensemble in **Figure 2B** is shown in **Figure 3C**, where $d_{evoked} = 1.3 \sim 1.7$ across 4 different taste stimuli. Across all sessions, dimensionality was overall smaller ($d_{evoked} = 2.0 \pm 0.6$, mean±SD, range: [1.1, 3.9]) and had a reduced slope as a function of N compared to ongoing activity (**Figure 3D**, linear regression, $r = 0.39$, slope $b_{evoked} = 0.13 \pm 0.03$, $p < 10^{-4}$). However, since dimensionality depends on the number and duration of the trials used for its estimation (**Figure 3E**), a proper comparison requires matching trial number and duration for each data point, as described next.

Stimulus-Induced Reduction of Dimensionality

We matched the number and duration of the trials for each data point and ran a two-way ANOVA with condition (ongoing vs. evoked) and ensemble size as factors. Both the main dimensionality [$F_{(1, 202)} = 11.93$, $p < 0.001$] and the slope were significantly smaller during evoked activity [test of interaction, $F_{(6, 202)} = 5.09$, $p < 10^{-4}$]. There was also a significant effect of ensemble size [$F_{(6, 202)} = 18.72$, $p < 10^{-14}$], confirming the results obtained with the separate regression analyses. These results suggest that stimuli induce a reduction of the effective space visited by the firing rate vector during evoked activity. This was confirmed by a paired sample analysis of the individual dimensionalities across all $27 \times 4 = 108$ ensembles (27 ensemble times 4 gustatory stimuli; $p < 0.002$, Wilcoxon signed-rank test).

Dimensionality is larger in Ensembles of Independent Neurons

The dimensionality depends on the pair-wise correlations of simultaneously recorded neurons. Shuffling neurons across ensembles would destroy the correlations (beyond those expected by chance), and would give a measure of how different the dimensionality of our datasets would be compared to sets of independent neurons. We measured the dimensionality of surrogate datasets obtained by shuffling neurons across sessions; because shuffling destroys the structure of the hidden states, firing rates in bins of fixed duration (200 ms) were used to estimate the dimensionality (see Methods for details). As expected, the slope of d vs. N was larger in the shuffled datasets compared to the simultaneously recorded ensembles (not shown) during both ongoing activity ($b_{shuff} = 0.67 \pm 0.06$ vs. $b_{data} = 0.60 \pm 0.01$; mean ± SD, Mann-Whitney test, $p < 0.001$, 20 bootstraps), and evoked activity ($b_{shuff} = 0.36 \pm 0.07$ vs. $b_{data} = 0.29 \pm 0.01$; $p < 0.001$). Especially during ongoing activity, this result was accompanied by a narrower distribution of pair-wise correlations in the shuffled datasets compared to the simultaneously recorded datasets (**Figure 3G**), and is consistent with an inverse relationship between dimensionality and pair-wise correlations (see Equation 9).

Time Course of Dimensionality as a Function of Ensemble Size

Unlike ongoing activity, the dependence of dimensionality on ensemble size (the slope of the linear regression of d vs. N) was modulated during different epochs of the post-stimulus period [**Figure 3F**, full lines; two-way ANOVA; main effect of time $F_{(4, 495)} = 3.80$, $p < 0.005$; interaction time x condition: $F_{(4, 495)} = 4.76$, $p < 0.001$]. In particular, the dependence of d on the ensemble size N almost disappeared immediately after stimulus presentation in the simultaneously recorded, but not in the shuffled ensembles (trial-matched slope in the first evoked second: $b_{evoked} = 0.07 \pm 0.01$ vs $b_{shuff} = 0.19 \pm 0.07$) and converged to a stable value after approximately 1 second (slope after the first second $b_{evoked} = 0.38 \pm 0.01$; compare with a stable average slope during ongoing activity of $b_{ongoing} = 0.57 \pm 0.01$, **Figure 3F**).

Note that the dimensionality is larger when the firing rate is computed in bins (as in **Figure 3F**) rather than in HMM states (as in **Figures 3B–D**, where the slopes are about half than in **Figure 3F**). The reason is that firing rates and correlations are

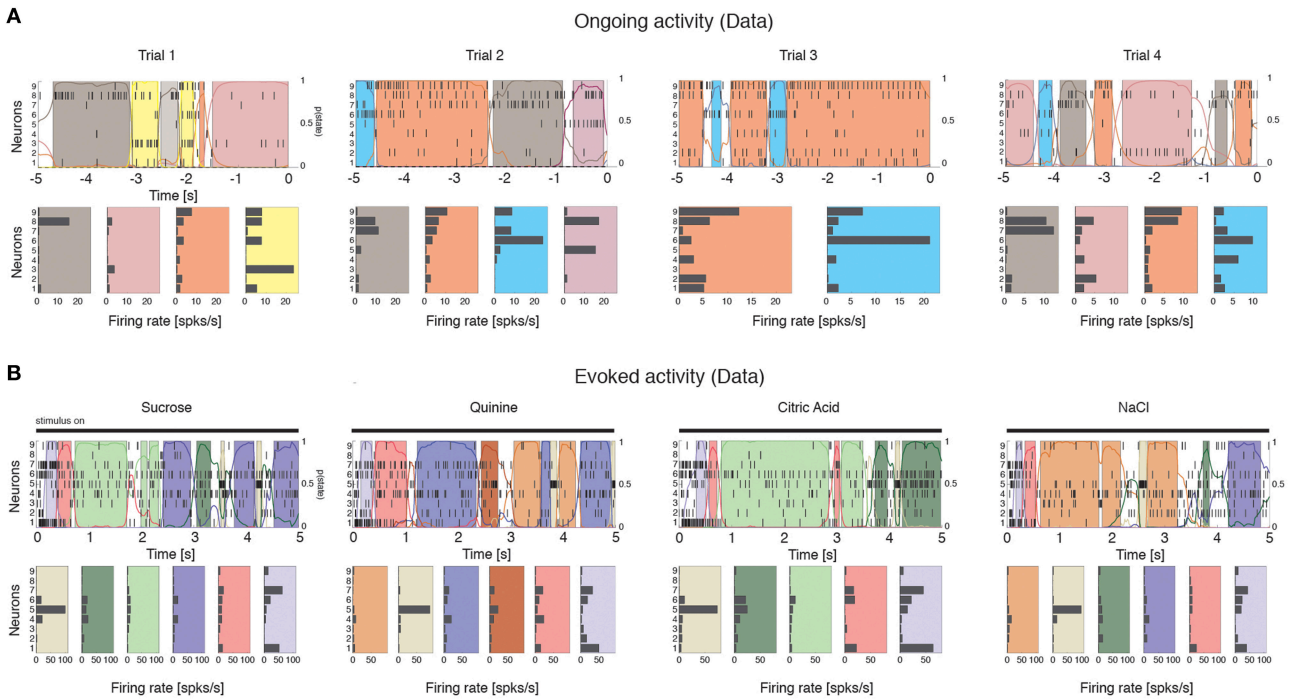


FIGURE 2 | Ensemble neural activity is characterized by sequences of states. (A) Upper panels: Representative trials from one ensemble of nine simultaneously recorded neurons during ongoing activity, segmented according to their ensemble states (HMM analysis, thin black vertical lines are action potentials; states are color-coded; smooth colored lines represent the probability for each state; shaded colored areas indicate intervals where the probability of a state exceeds 80%). Lower panels: Average firing rates across simultaneously recorded neurons (states are color-coded as in the upper panels). In total, 6 hidden states were found in this example session (only 5 states shown). X-axis for population rasters: time preceding the next event at (0 = stimulus delivery); Y-axis for population rasters: left, ensemble neuron index, right, probability of HMM states; X-axis for average firing rates panels: firing rates (spks/s); Y-axis for firing rate panels: ensemble neuron index. **(B)** Ensemble rasters and firing rates during evoked activity for four different tastes delivered at $t = 0$: sucrose, sodium chloride, citric acid, and quinine (notations as in panel **A**). In total, eight hidden states were found in this session during evoked activity.

approximately constant during the same HMM state, whereas they may change when estimated in bins of fixed duration that include transitions among hidden states. These changes tend to dilute the correlations resulting in higher dimensionality as predicted e.g., by Equation (9). A comparison of the pair-wise correlations of binned firing rates (**Figure 3G**) vs. those of firing rates in HMM states (**Figure 3H**) confirmed this hypothesis. Also, if the argument above is correct, one would expect a dependence of dimensionality on (fixed) bin duration. We computed the correlations and dimensionality of binned firing rates for various bin durations and found that r increases and d decreases for increasing bin durations (not shown). However, the slope of d vs. N is always larger in ongoing than in evoked activity regardless of bin size (ranging from 10 ms to 5 s; not shown). This confirms the generality of the results of **Figures 3B–D**, which were obtained using firing rate vectors in hidden states.

To summarize our main results so far, we found that dimensionality depends on ensemble size during both ongoing and evoked activity, and such dependence is significantly reduced in the post-stimulus period. This suggests that while state sequences during ongoing activity explore a large portion of the available firing rate space, the presentation of a stimulus initially collapses the state sequence along a more stereotyped and lower-dimensional response (Katz et al., 2001; Jezzini et al., 2013). During both ongoing and evoked activity, the dimensionality is also different than expected by chance in a set of independent neurons (shuffled datasets).

Clustered Spiking Network Model of Dimensionality

To gain a mechanistic understanding of the different dimensionality of ongoing and evoked activity, we have analyzed a spiking network model with clustered connectivity which has been shown to capture many essential features of the data (Mazzucato et al., 2015). In particular, the model reproduces the transitions among latent states in both ongoing and evoked activity. The network (see Methods for details) comprises Q clusters of excitatory neurons characterized by stronger synaptic connections within each cluster and weaker connections between neurons in different clusters. All neurons receive recurrent input from a pool of inhibitory neurons that keeps the network in a balanced regime of excitation and inhibition in the absence of external stimulation (**Figure 4A**). In very large networks (technically, in networks with an infinite number of neurons), the stable configurations of the neural activity are characterized by a finite number of active clusters whose firing rates depend on the number clusters active at any given moment, as shown

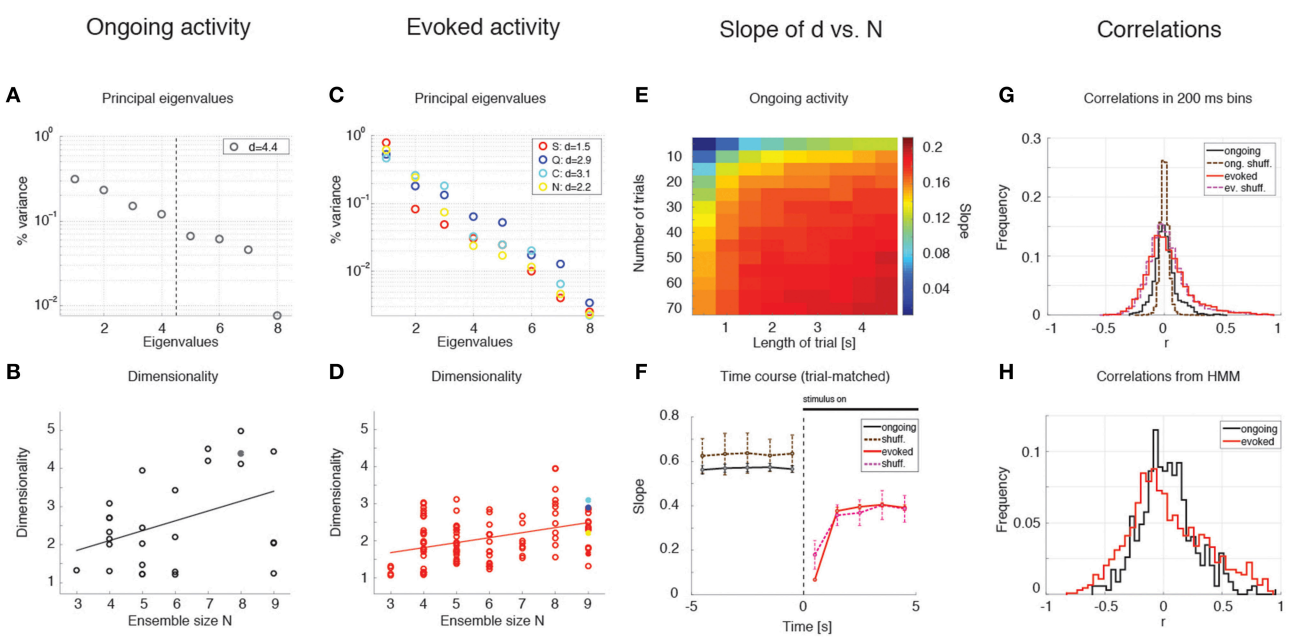


FIGURE 3 | Dependence of dimensionality on ensemble size (data). (A) Fraction of variance explained by each principal eigenvalue for an ensemble of 8 neurons during ongoing activity (corresponding to the filled dot in panel **B**) in the empirical dataset. The dashed vertical line represents the value of the dimensionality for this ensemble ($d = 4.4$). X-axis, eigenvalue number; Y-axis, fraction of variance explained by each eigenvalue. **(B)** Dimensionality of neural activity across all ensembles in the empirical dataset during ongoing activity (circles, linear regression fit, $d = b \cdot N + a,\ b = 0.26 \pm 0.12,\ a = 1.07 \pm 0.74,\ r = 0.4$), estimated from HMM firing rate fits on all ongoing trials in each session (varying from 73 to 129). X-axis: ensemble size; Y-axis: dimensionality. **(C)** Fraction of variance explained by each principal eigenvalue for the ensemble in **(A)** during evoked activity. Principal eigenvalues for sucrose (S, orange), sodium chloride (N, yellow), citric acid (C, cyan), and quinine (Q, blue) are presented (corresponding to the color-coded dots in panel **D**). X-axis, eigenvalue number; Y-axis, percentage of variance explained by each eigenvalue. **(D)** Dimensionality of neural activity across all ensembles in the empirical dataset during evoked activity (notations as in panel **B**, linear regression: $d = b \cdot N + a,\ b = 0.13 \pm 0.03,\ a = 1.27 \pm 0.19,\ r = 0.39$), estimated from HMM firing rate fits on evoked trials in each condition (varying from 7 to 11 trials across sessions for each tastant). **(E)** The slope of the linear regression of dimensionality (d) vs. ensemble size (N) as a function of the length of the trial interval and the number of trials used to estimate the dimensionality. X-axis, length of trial interval [s]; Y-axis, number of trials. **(F)** Time course of the trial-matched slopes of d vs. N, evaluated with 200 ms bins in consecutive 1 s intervals during ongoing (black curve, $t < 0$) and evoked periods (red curve, $t > 0$; error bars represent SD). A significant time course is triggered by stimulus presentation (see Results for details). The slopes of the empirical dataset (thick curves) were smaller than the slope of the shuffled dataset (dashed curves) during ongoing activity. X-axis, time from stimulus onset at $t = 0$ [s]; Y-axis, slope of d vs. N. **(G)** Distribution of pair-wise correlations in simultaneously recorded ensembles (black and red histograms for ongoing and evoked activity, respectively) and shuffled ensembles (brown and pink dashed histograms for ongoing and evoked activity, respectively) from 200 ms bins. X-axis, correlation; Y-axis, frequency. **(H)** Distribution of pair-wise correlations from HMM states during ongoing (black) and evoked activity (red) for all simultaneously recorded pairs of neurons. X-axis, correlation; Y-axis, frequency.

in **Figure 4B** (where $Q = 30$). In a finite network, however, finite size effects ignite transitions among these configurations, inducing network states (firing rate vectors) on randomly chosen subsets of neurons that resemble the HMM states found in the data (**Figure 5**; see Mazzucato et al., 2015 for details).

The dimensionality of the simulated sequences during ongoing and evoked activity was computed as done for the data, finding similar results. For the examples in **Figure 5**, we found $d_{ongoing} = 4.0$ for ongoing activity (**Figure 6A**) between $d_{evoked} = 2.2$ and $d_{evoked} = 3.2$ across tastes during evoked activity (**Figure 6C**). Across all simulated sessions, we found an average $d_{ongoing} = 2.9 \pm 0.9$ (mean $\pm$ SD) for ongoing activity and $d_{evoked} = 2.4 \pm 0.7$ for evoked activity. The model captured the essential properties of dimensionality observed in the data: the dimensionality did not differ across different tastes (one-way ANOVA, $p > 0.2$) and depended on ensemble size during both ongoing (**Figure 6B**; slope $= 0.36 \pm 0.07,\ r = 0.77,\ p < 10^{-4}$) and evoked periods (**Figure 6D**; slope $= 0.12 \pm 0.04,\ r = 0.29,\ p = 0.01$). As for the data, the dependency on ensemble size was smaller for evoked compared to ongoing activity. We performed a trial-matched two-way ANOVA as done on the data and found, also in the model, a main effect of condition [ongoing vs. evoked: $F_{(1,\ 146)} = 22.1, p < 10^{-5}$], a main effect of ensemble size [$F_{(6,\ 146)} = 14.1, p < 10^{-11}$], and a significant interaction [$F_{(6,\ 146)} = 3.8, p = 0.001$]. These results were accompanied by patterns of correlations among the model neurons (**Figures 6E,F**) very similar to those found in the data (**Figures 3G,H**; see section "Dimensionality is larger in the presence of clusters" for statistics of correlation values). As in the data, narrower distributions of correlations were found for binned firing rates (**Figure 6E**) compared to firing rates in hidden states (**Figure 6F**; compare with **Figures 3G,H**, respectively). Moreover, shuffling neurons

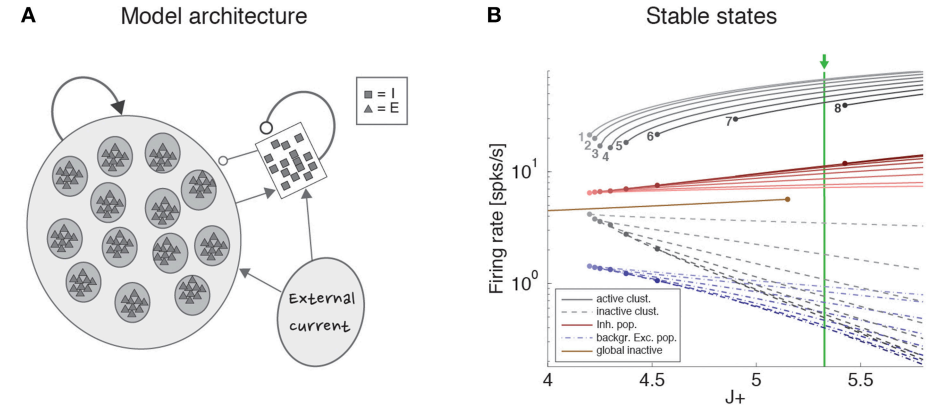


FIGURE 4 | Recurrent network model. (A) Schematic recurrent network architecture. Triangles and squares represent excitatory and inhibitory LIF neurons respectively. Darker disks indicate excitatory clusters with potentiated intra-cluster synaptic weights. **(B)** Mean field solution of the recurrent network. Firing rates of the stable states for each subpopulation are shown as function of the intra-cluster synaptic potentiation parameter J_+: firing rate activity in the active clusters (solid gray lines), firing rate in the inactive clusters (dashed gray lines), activity of the background excitatory population (dashed blue lines), activity of the inhibitory population (solid red lines). In each case, darker colors represent configurations with larger number of active clusters. Numbers denote how many clusters are active in each stable configuration. Configurations with 1–8 active clusters are stable in the limit of infinite network size. A global configuration where all clusters are inactive (brown line) becomes unstable at the value $J_+ = 5.15$. The vertical green line represents the value of $J_+ = 5.3$ chosen for the simulations. X-axis, intra-cluster potentiation parameter J_+ in units of J_{EE}; Y-axis, Firing rate (spks/s).

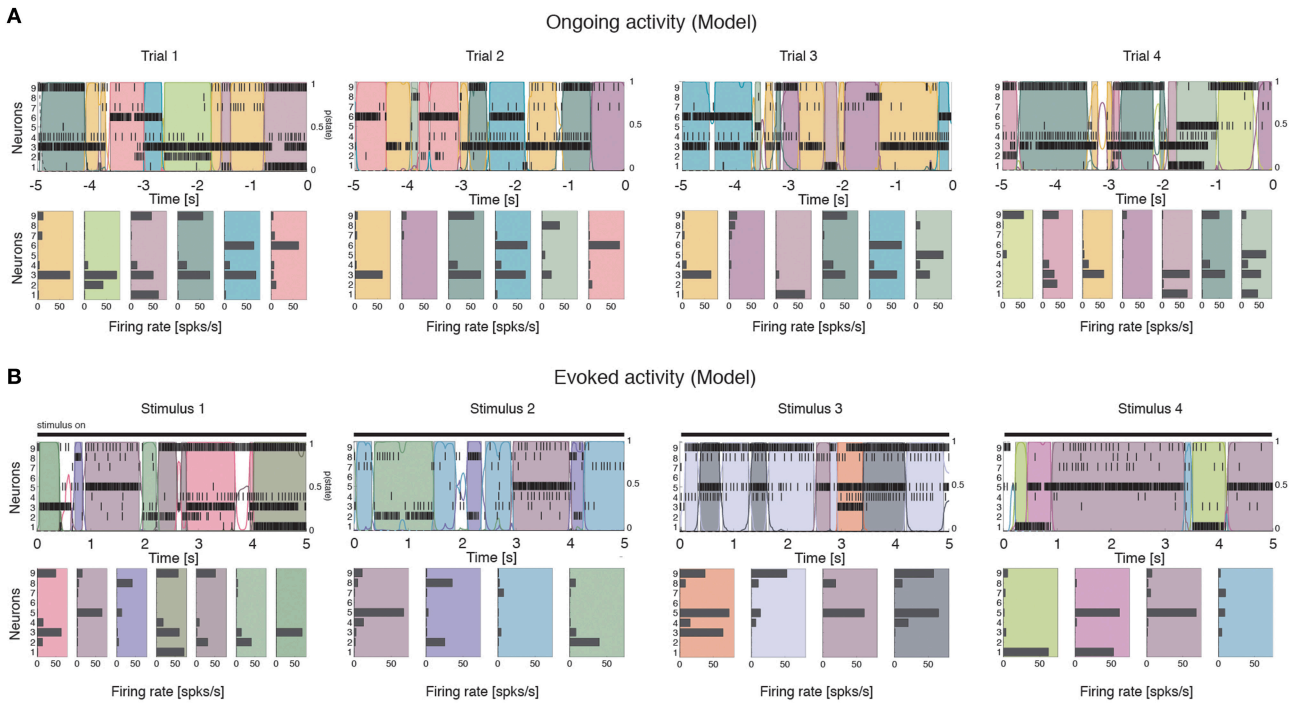


FIGURE 5 | Ensemble activity in the recurrent network model is characterized by sequences of states. Representative trials from one ensemble of nine simultaneously recorded neurons sampled from the recurrent network, segmented according to their ensemble states (notations as in **Figure 1**). **(A)** ongoing activity. **(B)** Ensemble activity evoked by four different stimuli, modeled as an increase in the external current to stimulus-selective clusters (see Methods for details).

across datasets reduced the correlations (**Figure 6E**, dashed), resulting in a larger slope of d vs. N (not shown). Finally, d during ongoing activity was always larger than during evoked activity also when computed on binned firing rates (not shown), as found in the data (see section "Dependence of dimensionality on bin size").

Since the model was not fine-tuned to find these results, the different dimensionalities of ongoing and evoked activity, and

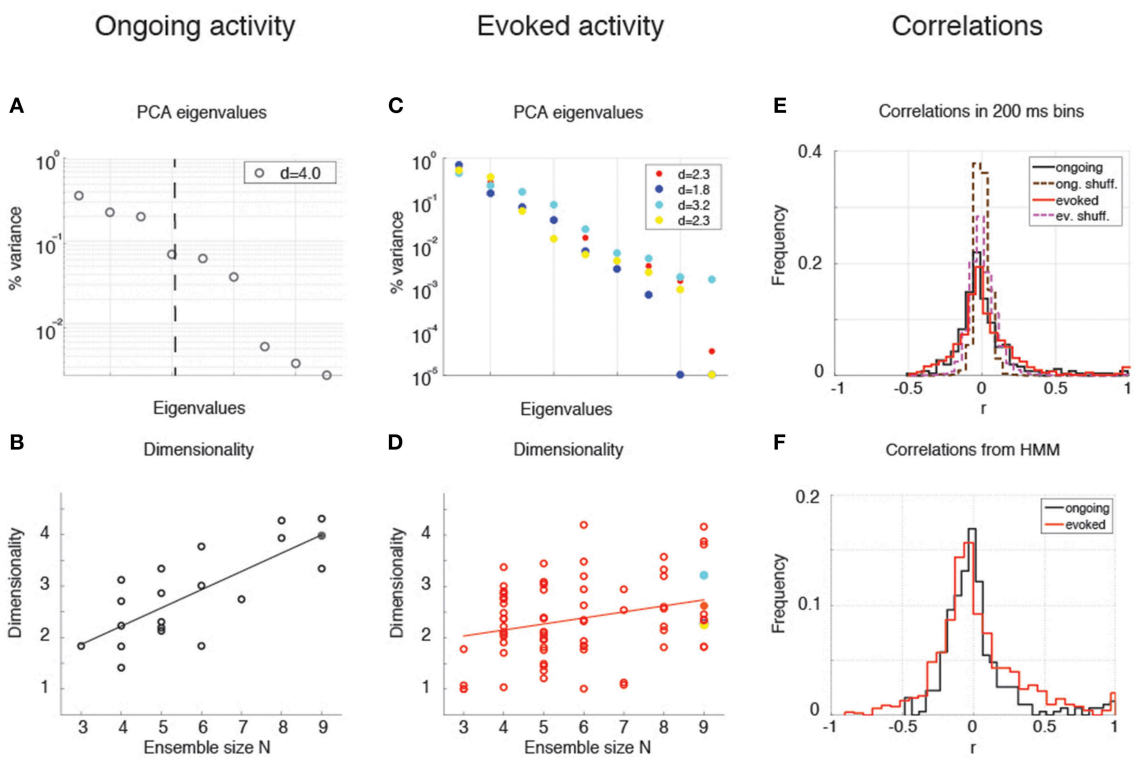


FIGURE 6 | Dependence of dimensionality on ensemble size (model). (A) Fraction of variance explained by each principal eigenvalue for an ensemble of 9 neurons during ongoing activity (corresponding to the filled dot in panel **B**) in the model network of **Figure 5** (notations as in **Figure 3A**). **(B)** Dimensionality of neural activity across all ensembles in the model during ongoing activity (linear regression fit, $d = b \cdot N + a,\ b = 0.36 \pm 0.07,\ a = 0.80 \pm 0.43,\ r = 0.77$), estimated from HMM firing rate fits. X-axis, ensemble size; Y-axis, dimensionality. **(C)** Fraction of variance explained by each principal eigenvalue for the ensemble in panel **A** during evoked activity. Principal eigenvalues for four tastes are presented (corresponding to the color-coded dots in panel **D**). X-axis, eigenvalue number; Y-axis, percentage of variance explained by each eigenvalue. **(D)** Dimensionality of neural activity across all ensembles in the model during evoked activity (notations as in panel **B**, linear regression: $d = b \cdot N + a,\ b = 0.12 \pm 0.04,\ a = 1.70 \pm 0.26,\ r = 0.29$). **(E)** Distribution of pair-wise correlations in simultaneously recorded ensembles from the clustered network model (black and red histograms for ongoing and evoked activity, respectively) and in shuffled ensembles (brown and pink dashed histograms for ongoing and evoked activity, respectively) from 200 ms bins. X-axis, correlation; Y-axis, frequency. **(F)** Distribution of pair-wise correlations from HMM states during ongoing (black) and evoked activity (red) for all simultaneously recorded pairs of neurons. X-axis, correlation; Y-axis, frequency.

their associated patterns of pair-wise correlations, are likely the consequence of the organization in clusters and of the ensuing dynamics during ongoing and evoked activity.

Scaling of Dimensionality with Ensemble Size and Pair-Wise Correlations

The dependence of dimensionality on ensemble size observed in the data (**Figure 3B**) and in the model (**Figure 6B**) raises the question of whether or not the dimensionality would converge to an upper bound as one increases the number of simultaneously recorded neurons. In general, this question is important in a number of settings, related e.g., to coding in motor cortex (Ganguli et al., 2008; Gao and Ganguli, 2015), performance in a discrimination task (Rigotti et al., 2013), or coding of visual stimuli (Cadieu et al., 2013). We can attack this question aided by the model of **Figure 4**, where we can study the effect of large numbers of neurons, but also the impact on dimensionality of a clustered network architecture compared to a homogeneous one, at parity of correlations and ensemble size.

We consider first the case of a homogeneous network of neurons having no clusters and low pair-wise correlations, but having the same firing rates distributions (which were approximately log-normal, **Figure 7A**) and the same mean pair-wise correlations as found in the data ($\rho \sim 0.01 - 0.2$). This would require solving a homogeneous recurrent network self-consistently for the desired firing rates and correlations. As a proxy for this scenario, we generated 20 sessions of 40 Poisson spike trains having exactly the desired properties (including the case of independent neurons for which $\rho = 0$). Two examples with $\rho = 0$ and $\rho = 0.1$, respectively, are shown in **Figures 7B,C**. Since in the asynchronous homogeneous network there are no transitions and hence no hidden states, the dimensionality was estimated based on the rate vectors in bins of 200 ms duration (using bin widths of 50–500 ms did not change the results; see Methods for details).

We found that the dimensionality grows linearly with ensemble size in the absence of correlations, but is a concave function of N in the presence of pair-wise correlations (circles in **Figure 7D**). Thus, as expected, the presence of correlations reduces the dimensionality and suggests the possibility of an upper bound. A simple theoretical calculation mimicking this scenario shows that d in this case converges indeed to an upper bound that depends on the inverse of the square of the pair-wise correlations. For example, in the case of uniform correlations (ρ) and equal variances of the spike counts, Equation (8) of Methods, $d\,(N,\rho) = \frac{1}{\rho^2+(1-\rho^2)/N}$, shows that $d = N$ in the absence of correlations, but $d < 1/\rho^2$ in the presence of correlations. These properties remain approximately true if the variances σ_i^2 of the firing rates are drawn from a distribution with mean $E\left[\sigma_i^2\right] = \sigma^2$ and variance $Var\left[\sigma_i^2\right] = \delta\sigma^4$. As Equation (9) shows, in such a case dimensionality is reduced compared to the case of equal variances, for example $d \approx \frac{\sigma^4}{\sigma^4+\delta\sigma^4}N < N$ for large N when $\rho = 0,\ \delta\rho = 0$.

The analytical results are shown in **Figure 7E** (full lines correspond to Equation 8), together with their estimates ("+") based on 1000 data points (same number as trials in **Figure 7D**; see Methods). The estimates are based on surrogate datasets

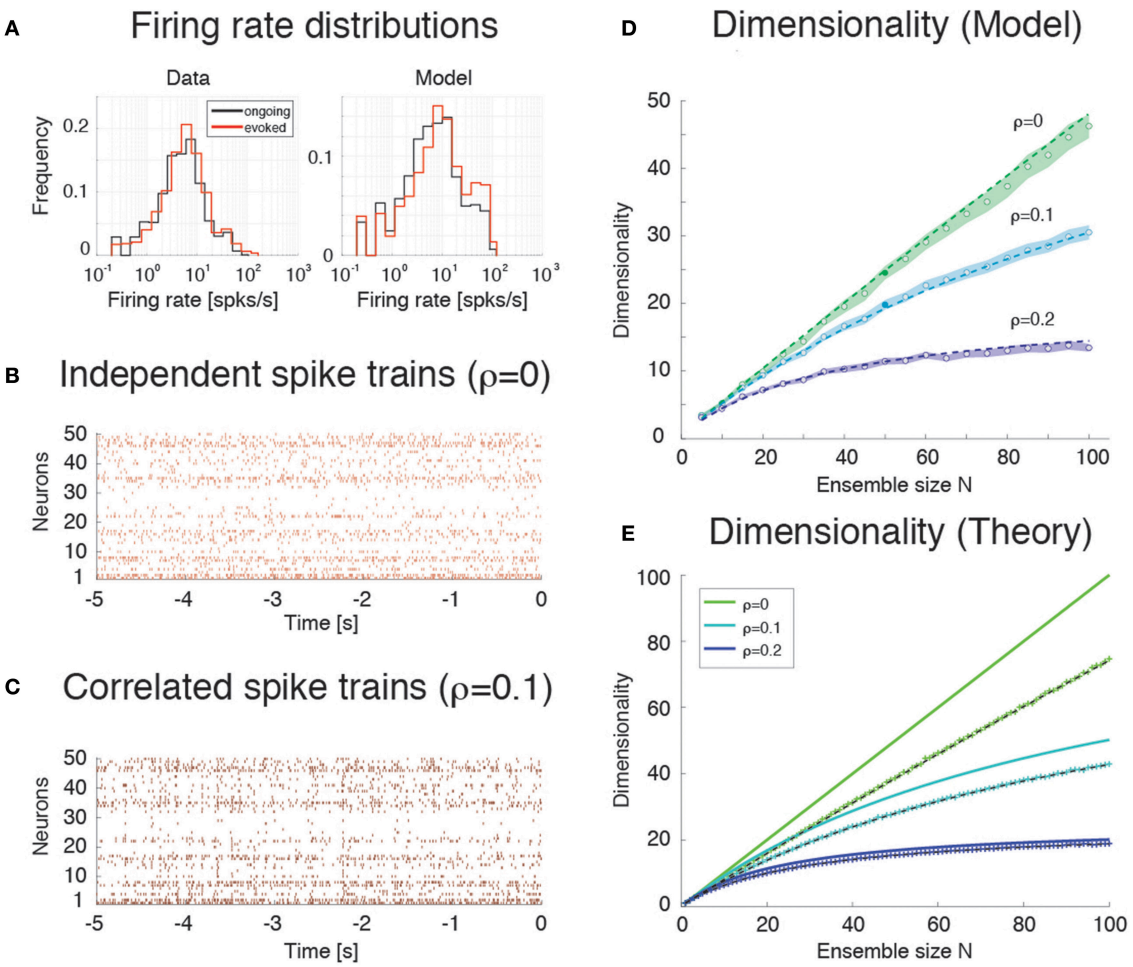


FIGURE 7 | Dimensionality and correlation. (A) Empirical single neuron firing rate distributions in the data (left) and in the model (right), for ongoing (black), and evoked activity (red). The distributions are approximately lognormal. X-axis, Firing rate (spks/s); Y-axis, frequency. **(B)** Example of independent Poisson spike trains with firing rates matched to the firing rates obtained in simulations of the spiking network model. **(C)** Example of correlated Poisson spike trains with firing rates matched to the firing rates obtained in simulations of the spiking network model. Pair-wise correlations of $\rho = 0.1$ were used (see Methods). X-axis, time [s]; Y-axis, neuron index. **(D)** Dimensionality as a function of ensemble size N in an ensemble of Poisson spike trains with spike count correlations $\rho = 0,\ 0.1,\ 0.2$ and firing rates matched to the model simulations of **Figure 6**. Dashed lines represent the fit of Equation (16) to the data (with $\delta\rho^2 = \alpha\rho^2,\ \sigma^4 = \delta\sigma^4 = \beta$), with best-fit parameters (mean ± s.e.m.) $\alpha = 0.22 \pm 10^{-5},\ \beta = 340 \pm 8$. Filled circles (from top to bottom): dimensionality of the data (raster plots) shown in **(B,C)** (shaded areas represent SD). X-axis, ensemble size; Y-axis, dimensionality. **(E)** Theoretical prediction for the dependence of dimensionality on ensemble size N and firing rate correlation ρ for the case of uniform correlation, Equation (8) (thick lines; green to cyan to blue shades represent increasing correlations). "+" are dimensionality estimates from $N_T = 1,000$ trials for each N (same N_T as in panel **D**, each trial providing a firing rate value sampled from a log-normal distribution), in the case of log-normally distributed firing rate variances σ_i^2 with mean $\sigma^2 = 40$ $(spk/s)\hat{}2$ and standard deviation 0.5 σ^2. Theoretical predictions from Equation (16) match the estimated values in all cases (dashed black lines). X-axis, ensemble size N; Y-axis, dimensionality.

with lognormal-distributed variances σ_i^2 to mimic the empirical distribution of variances found in GC (not shown).

Estimation Bias

Comparison of **Figures 7D,E** shows that the dimensionality of the homogeneous network is underestimated compared to the theoretical value given by Equation (8). This is due to a finite number of trials and the presence of unequal variances with spread $\delta\sigma^4$ ("+" in **Figure 7E**). As **Figure 7E** shows, taking this into account will reduce the dimensionality to values comparable to those of the homogeneous network of **Figure 7D**. The dimensionality in that case is well predicted by Equation (16) (broken lines in **Figure 7E**). The same Equation (16) was fitted successfully to the data in **Figure 7D** (dashed) by tuning 2 parameters to account for the unknown variance and correlation width of the firing rates (see Methods for details).

Empirically, estimates of the dimensionality Equation (2) based on a finite number N_T of trials tend to underestimate d (**Figure 3E**). The approximate estimator Equation (16) confirms that, for any ensemble size N, d is a monotonically increasing function of the number of trials (**Figure 8A**). Note that this holds for the mean value of the estimator (Equation 16) over many datasets, not for single estimates, which could overestimate the true d (not shown). Equation (16) also provides an excellent description of dimensionality as a function of firing rates' variance $\delta\sigma^4$ (**Figure 8B**) and pair-wise correlations width $\delta\rho^2$ (**Figure 8C**). In particular, the mean and the variance of the pair-wise correlations have an interchangeable effect on d (see Equation 16); they both decrease the dimensionality and so does the firing rate variance $\delta\sigma^4$ (**Figure 8B**).

Scaling of Dimensionality in the Presence of Clusters

We next compared the dimensionality of the homogeneous network's activity to that predicted by the clustered network model of **Figure 4**. To allow comparison with the homogeneous network, dimensionality was computed based on the spike counts in 200 ms bins rather than the HMM's firing rate vectors as in **Figure 6** (see Methods for details).

We found that the dependence of d on N in the clustered network depends on how the neurons are sampled. If the sampling is completely random, so that any neuron has the same probability of being added to the ensemble regardless of cluster membership, a concave dependence on N will appear, much like the case of the homogeneous network (**Figure 9A**, dashed lines). However, if neurons are selected one from each cluster until all clusters have been sampled once, then one neuron from each cluster until all clusters have been sampled twice, and so on, until all the neurons in the network have been sampled, then the dependence of d on N shows an abrupt transition when $N = Q$, i.e., when the number of sampled neurons reaches the number of clusters in the network (**Figure 9A**, full lines; see **Figure 9B** for raster plots with $Q = 30$ and $N = 50$). In the following, we refer to this sampling procedure as "ordered sampling," as a reminder that neurons are selected randomly from each cluster, but the clusters are selected in serial order. For $N \leq Q$, the dimensionality grows linearly with ensemble size in both ongoing (slope 0.24 ± 0.01, $r = 0.79$, $p < 10^{-10}$, black line) and evoked periods (slope 0.19 ± 0.01, $r = 0.84$, $p < 10^{-10}$; red line), and was larger during ongoing than evoked activity [trial-matched two-way ANOVA, main effect:

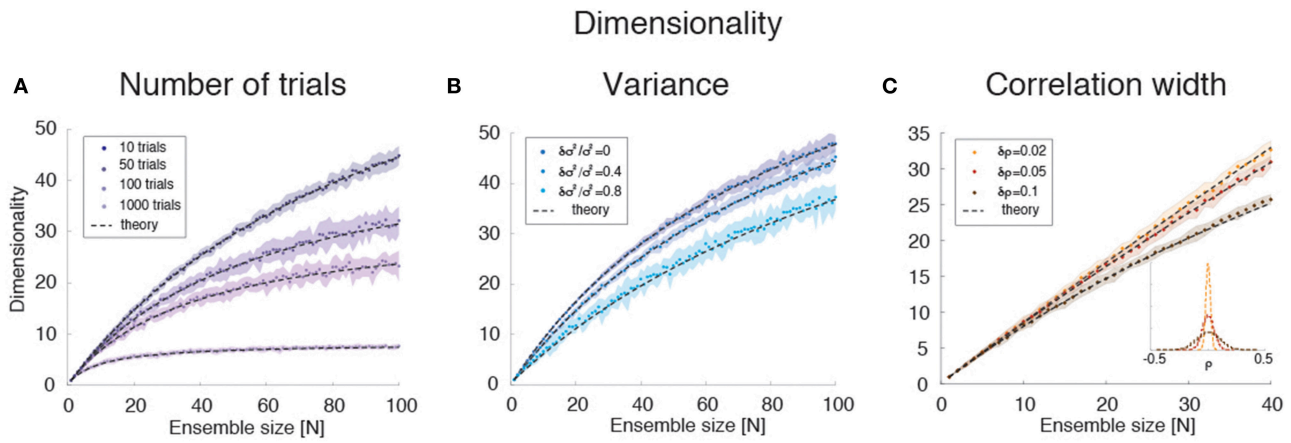


FIGURE 8 | Dimensionality estimation. (A) Dependence of dimensionality on the number of trials for variable ensemble size N, for fixed correlations $\rho = 0.1$ and firing rates variances σ_i^2 with mean σ^2 and standard deviation $\delta\sigma^2 = 0.4\,\sigma^2$. Dashed lines: theoretical prediction, Equation (16); dots: mean values from simulations of 20 surrogate datasets containing 10–1000 trials each (shaded areas: SD), with darker shades representing increasing number of trials. X-axis: ensemble size; Y-axis, dimensionality. **(B)** Dependence of dimensionality on the spread $\delta\sigma^2$ of the firing rates variances for fixed correlations $\rho = 0.1$ and firing rate variance with mean σ^2. Dashed lines: theoretical prediction, Equation (16); dots: mean values from simulations of 20 surrogate datasets containing 1000 trial each (shaded areas: SD), with lighter shades representing increasing values of $\delta\sigma^2/\sigma^2$). X-axis, ensemble size; Y-axis, dimensionality. **(C)** Dependence of dimensionality on the width $\delta\rho = \sqrt{Var(\rho)}$ of pair-wise firing rate correlations (with zero mean, $\rho = 0$), for firing rates variances σ_i^2 with mean σ^2 and standard deviation $\delta\sigma^2 = 0.4\,\sigma^2$. Dashed lines: theoretical prediction, Equation (16); dots: mean values from simulations of 20 surrogate datasets containing 1000 trials each (shaded areas: SD), with darker shades representing increasing values of $\delta\rho$. *Inset*: distribution of correlation coefficients used in the main figure. X-axis, ensemble size; Y-axis, dimensionality. In all panels, $\sigma^2 = 40$ (*spk*/s)^2.

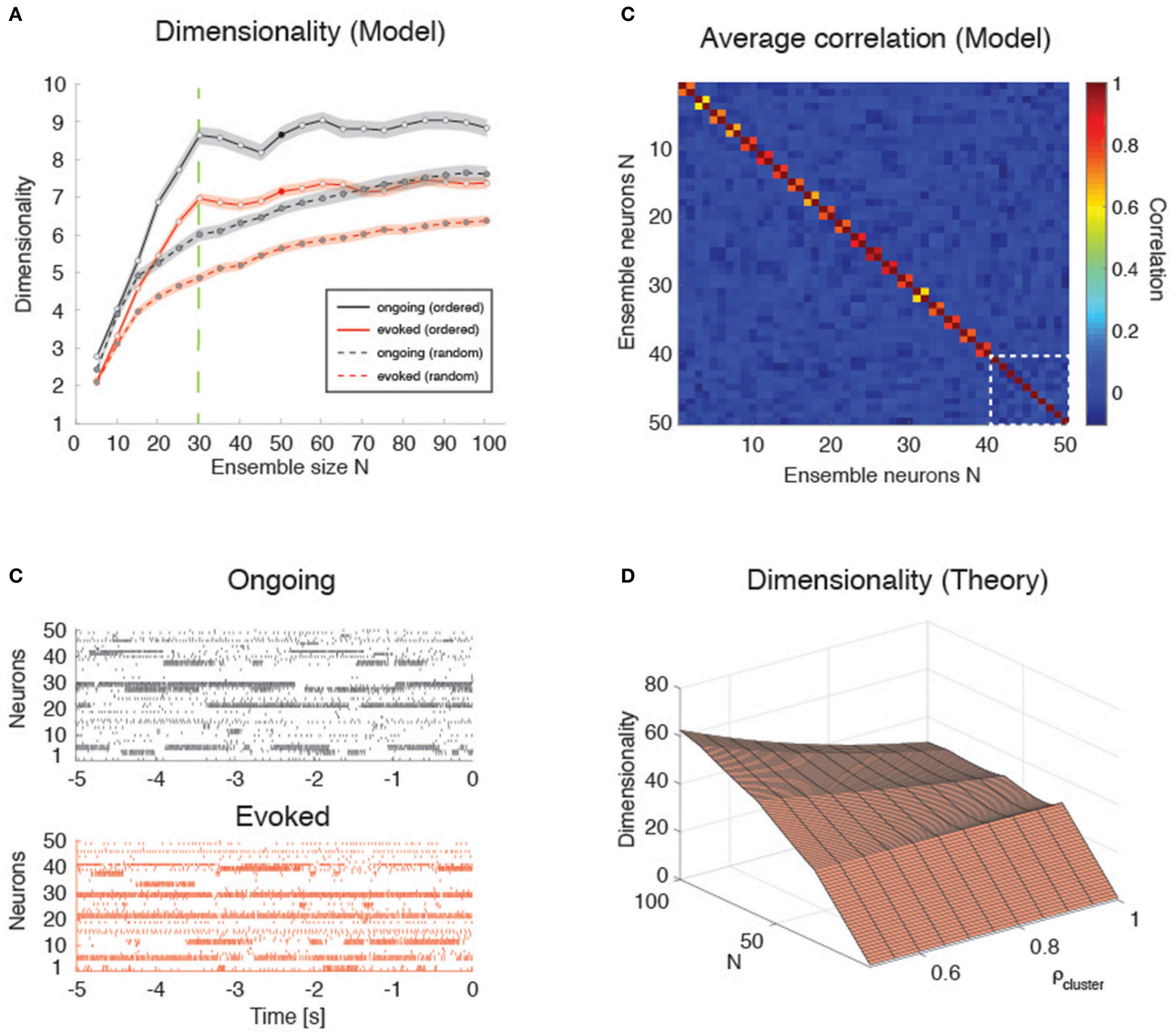


FIGURE 9 | Dimensionality in a clustered network. (A) Trial-matched dimensionality as a function of ensemble size in the recurrent network model (ongoing and evoked activity in black and red, respectively, with shaded areas representing s.e.m.). Filled lines represent ordered sampling, where ensembles to the left of the green vertical line ($N = Q = 30$) contain at most one neuron per cluster, while to the right of the line they contain one or more neurons from all clusters (filled circles indicate representative trials in panel **B**). Dashed lines represent random sampling of neurons, regardless of cluster membership. X-axis, ensemble size; Y-axis, dimensionality. **(B)** Representative trial of an ensemble of 50 neurons sampled from the recurrent network in **Figure 4** during ongoing activity (upper plot, in black) or evoked activity (lower plot, in red) for the case of "ordered sampling" (full lines in panel **A**). Neurons are sorted according to their cluster membership (adjacent neuron pairs with similar activity belong to the same cluster, for neurons #1 up to #40; the last 10 neurons are sampled from the remaining clusters). X-axis, time to stimulus presentation at $t = 0$ (s); Y-axis, neuron index. **(C)** Average correlation matrix for 20 ensembles of $N = 50$ neurons from the clustered network model with $Q = 30$ clusters. For the first 40 neurons, adjacent pairs belong to the same cluster; the last 10 neurons (delimited by a dashed white square) belong to the remaining clusters (neurons are ordered as in panel **B**). Thus, neurons 1, 3, 5, …, 39 (20 neurons) belong to the first 20 clusters; neurons 2, 4, 6, …, 40 (20 neurons) belong also the first 20 clusters; and neurons 41, 42, 43, …, 50 (10 neurons) belong to the remaining 10 clusters. X-axis, Y-axis: neuron index. **(D)** Plot of Equation (12) giving d vs. N and ρ (uniform within-cluster correlations) for the sampling procedure of panel **(B)**. X-axis, ensemble size N; Y-axis, dimensionality.

$F_{(1,\,948)} = 168$, $p < 10^{-30}$; interaction: $F_{(5,\,948)} = 4.1$, $p < 0.001$].

These results are in keeping with the empirical and model results based on the HMM analysis (**Figures 3**, **6**). However, in the case of ordered sampling, the dependence of dimensionality on ensemble size tends to disappear for $N \geq Q$ both during ongoing (slope 0.010 ± 0.003, $r = 0.1$, $p < 0.001$) and evoked periods (slope 0.009 ± 0.002, $r = 0.13$, $p < 10^{-4}$; **Figure 9A**, full lines). The average dimensionality over the range $30 \leq N \leq 100$ was significantly larger for ongoing, $d_{ongoing} = 8.74 \pm 0.06$, than for evoked activity, $d_{evoked} = 7.15 \pm 0.04$ [trial-matched two-way ANOVA, main effect: $F_{(1,2212)} = 488$, $p < 10^{-30}$], confirming

that dimensionality during ongoing is larger than during evoked activity also in this case. The difference in dimensionality between ongoing and evoked activity also holds in the case of random sampling on the entire range of N values (**Figure 9A**, dashed lines), confirming the generality of this finding.

Dimensionality is Larger in the Presence of Clusters

Intuitively, the dimensionality saturates at $N = Q$ in the clustered network because additional neurons will be highly correlated with already sampled ones. For $N \leq Q$, each new neuron's activity adds an independent degree of freedom to the neural dynamics and thus increases its dimensionality. For $Q > N$, additional neurons are highly correlated with an existing neuron, adding little or no additional contribution to d. Indeed, compared to the low overall correlations found across all neuron pairs in the data (and used as *desiderata* for the homogeneous network), neurons belonging to the same model cluster had a much higher correlation of $\rho = 0.92$ [0.56, 0.96] (median and [25, 75]-percentile), while neurons belonging to different clusters had negligible correlation ($\rho \approx 0$, [−0.10, 0.06]). A negligible median correlation was typical: for example, negligible was the overall median correlation regardless of cluster membership ($\rho \approx 0$ [−0.109, 0.083]); and the empirical correlation both during ongoing ([−0.047, 0.051] , with rare maximal values of $\rho \sim 0.5$), and evoked activity ([−0.085, 0.113], with rare maximal values of $\rho \sim 0.9$). While we note the qualitative agreement of model and empirical correlations, we emphasize that these numbers were obtained using 200 ms bins and that they were quite sensitive to bin duration. In particular, the maximal correlations (regardless of sign) were substantially reduced for smaller bin durations (not shown).

Plugging these values into a correlation matrix reflecting the clustered architecture and the "ordered" sampling procedure used in **Figure 9B**, we obtained the matrix shown in **Figure 9C**, where pairwise correlations depend on whether or not the neurons belong to the same cluster (for the first 40 neurons, adjacent pairs belong to the same cluster; the last 10 neurons belong to the remaining clusters). It is natural to interpret such correlation matrix as the noisy observation of a block-diagonal matrix such that neurons in the same cluster have uniform correlation while neurons from different clusters are uncorrelated. For such a correlation matrix the dimensionality can be evaluated exactly (see Equation 12 of Methods). In the approximation where all neurons have the same variance, this reduces to Equation (13), i.e.,

$$d(N,\rho) = \begin{cases} N & , \ N \leq Q \\ \frac{N}{1+m\rho^2[1-(Q-p)/N]} & , \ N > Q \end{cases},$$

where $N = mQ + p$. This formula is plotted in **Figure 9D** for relevant values of ρ and N and it explains the origin of the abrupt transition in dimensionality at $Q = \mathrm{N}$. (The reasons for a dimensionality lower than N for $N \leq Q$ in the data–see **Figure 9A**–are, also in this case, the finite number of data points (250) used for its estimation and the non-uniform distributions of firing rate variances and correlations).

Note that the formula also predicts cusps in dimensionality (which become local maxima for large ρ) whenever the ensemble size is an exact multiple of the number of clusters. This is also visible in the simulated data of **Figure 9A**, where local maxima seem to appear at $N = 30, 60, 90$ with $Q = 30$ clusters. It is also worth mentioning that, for low intra-cluster correlations, the dependence on N predicted by Equation (13) becomes smoother and the cusps harder to detect (not shown), suggesting that the behavior of a clustered network with weak clusters tends to converge to the behavior of a homogeneous asynchronous network—therefore lacking sequences of hidden states. Thus, the complexity of the network dynamics is reflected in how its dimensionality scales with N, assuming that one may sample one neuron per cluster (i.e., via "ordered sampling").

Even though it is not clear how to perform ordered sampling empirically (see Discussion), this result is nevertheless useful since it represents an upper bound also in the case of random sampling (see **Figure 9A**, dashed lines). Equation (13) predicts that $d \leq Q/\rho^2$, with this value reached asymptotically for large N. In the case of random sampling, growth to this bound is even slower (**Figure 9A**). For comparison, in a homogeneous network $d \leq 1/\rho^2$ from Equation (8), a bound that is smaller by a factor of Q. Finally, homogeneous dimensionality is dominated by clustered dimensionality also in the more realistic case of non-uniform variances and correlations, where similar bounds are found in both cases (see Methods for details).

DISCUSSION

In this paper we have investigated the dimensionality of the neural activity in the gustatory cortex of alert rats. Dimensionality was defined as a collective property of ensembles of simultaneously recorded neurons that reflects the effective space occupied by the ensemble activity during either ongoing or evoked activity. If one represents ensemble activity in terms of firing rate vectors, whose dimension is the number of ensemble neurons N, then the collection of rate vectors across trials takes the form of a set of points in the N-dimensional space of firing rates. Roughly, dimensionality is the minimal number of dimensions necessary to provide an accurate description of such set of points, which may be localized on a lower-dimensional subspace inside the whole firing rate space.

One of the main results of this paper is that the dimensionality of evoked activity is smaller than that of ongoing activity, i.e., stimulus presentation quenches dimensionality. More specifically, the dimensionality is linearly related to the ensemble size, with a significantly larger slope during ongoing activity compared to evoked activity (compare **Figures 3B,D**). We explained this phenomenon using a biologically plausible, mechanistic spiking network model based on recurrent connectivity with clustered architecture. The model was recently introduced in Mazzucato et al. (2015) to account for the observed dynamics of ensembles of GC neurons as sequences of metastable states, where each state is defined as a vector of firing rates across simultaneously recorded neurons. The model captures the reduction in trial-to-trial variability and the multiple firing rates attained by single neurons across different

states observed in GC upon stimulus presentation. Here, the same model was found to capture also the stimulus-induced reduction of dimensionality. While the set of active clusters during ongoing activity varies randomly, allowing the ensemble dynamics to explore a large portion of firing rate space, the evoked set of active clusters is limited mostly to the stimulus-selective clusters only (see Mazzucato et al., 2015 for a detailed analysis). The dynamics of cluster activation in the model thus explains the more pronounced dependence of dimensionality on ensemble size found during ongoing compared to evoked activity.

We presented a simple theory of how dimensionality depends on the number of simultaneously recorded neurons N, their firing rate correlations, their variance, and the number and duration of recording trials. We found that dimensionality increases with N and decreases with the amount of pair-wise correlations among the neurons (e.g., **Figure 8C**). At parity of correlations, dimensionality is maximal when all neurons have the same firing rate variance, and it decreases as the distribution of count variances becomes more heterogeneous (e.g., **Figure 8B**). The estimation of dimensionality based on a finite dataset is an increasing function of the number of trials (**Figure 8A**). Finally, introducing clustered correlations in the theory, and sampling one neuron per cluster as in **Figure 9B**, results in cusps at values of N that are multiples of the number of clusters (**Figure 9D**), in agreement with the predictions of the spiking network model (**Figure 9A**, full lines).

Dimensionality Scaling with Ensemble Size

The increased dimensionality with sample size, especially during ongoing activity, was found empirically in datasets with 3–9 neurons per ensemble, but could be extrapolated for larger N in a spiking network model with homogeneous or clustered architecture. In homogeneous networks with finite correlations the dimensionality is predicted to increase sub-linearly with N (Equation 8), whereas in the clustered network it may exhibit cusps at multiple values of the number of clusters (**Figure 9A**), and would saturate quickly to a value that depends on the ratio of the number of clusters Q and the amount of pair-wise correlations, $d \leq Q/\rho^2$. Testing this prediction requires the ability to sample neurons one from each cluster, until all clusters are sampled, and seems beyond the current recording techniques. However, looking for natural groupings of neurons based on response similarities could uncover spatial segregation of clusters (Kiani et al., 2015) and could perhaps allow sampling neurons according to this procedure. Moreover, the model predicts a slower approach to a similar bound also in the case of random sampling.

Dimensionality in a homogeneous network is instead bounded by $1/\rho^2$, and hence it is a factor Q smaller than in the clustered network. Dimensionality is maximal in a population of independent neurons ($\rho = 0$), where it grows linearly with N; however, neurons of recurrent networks have wide-ranging correlations (see e.g., **Figures 6E,F** and its empirical counterpart, **Figures 3G,H**). Since the presence of even low correlations can dramatically reduce the dimensionality (see **Figure 7D**), the neural activity in a clustered architecture can reach much higher values at parity of correlations, representing an intermediate case between a homogeneous network and a population of independent neurons.

Evidence for the presence of spatial clusters has been recently reported in the prefrontal cortex based on correlations analyses (Kiani et al., 2015). An alternative possibility is that neural clusters are not spatially but functionally arranged, and cluster memberships vary with time and task complexity (Rickert et al., 2009). Can our model provide indirect tools to help uncover the presence of clusters? A closer look at **Figures 6E,F** reveal a small peak at large correlations due to the contribution of highly correlated neurons belonging to the same cluster. This peak would be absent in a homogenous network and thus is the signature of a clustered architecture. However, such peak is populated by only small fraction ($1/Q$) of the total number of neuron pairs, which hinders its empirical detection (no peak at large correlations is clearly visible in our data, see **Figures 3G,H**).

Dimensionality and Trial-to-Trial Variability

Cortical recordings from alert animals show that neurons produce irregular spike trains with variable spike counts across trials (Shadlen and Newsome, 1994; Fontanini and Katz, 2008; Moreno-Bote, 2014). Despite many efforts, it remains a key issue to establish whether variability is detrimental (Gur et al., 1997; White et al., 2012) or useful (McDonnell and Ward, 2011) for neural computation.

Trial-to-trial variability is reduced during preparatory activity (Churchland et al., 2006), during the presentation of a stimulus (Churchland, 2010b), or when stimuli are expected (Samuelsen et al., 2012), a phenomenon that would not occur in a population of independent or homogeneously connected neurons (Litwin-Kumar and Doiron, 2012). Recent work has shown that the stimulus-induced reduction of trial-to-trial variability can be due to spike-frequency adaptation in balanced networks (Farkhooi et al., 2013) or to slow dynamic fluctuations generated in a recurrent spiking networks with clustered connectivity (Deco and Hugues, 2012; Litwin-Kumar and Doiron, 2012; Mazzucato et al., 2015). In clustered network models, slow fluctuations in firing rates across neurons can ignite metastable sequences of neural activity, closely resembling metastable sequences observed experimentally (Abeles et al., 1995; Seidemann et al., 1996; Jones et al., 2007; Kemere et al., 2008; Durstewitz et al., 2010; Ponce-Alvarez et al., 2012; Mazzucato et al., 2015). The slow, metastable dynamics of cluster activation produces high variability in the spike count during ongoing activity. While cluster activations occur at random times during ongoing activity periods, stimulus presentation locks cluster activation at its onset, leading to a decrease in trial-to-trial variability.

Similarly, a stimulus-induced reduction of dimensionality is obtained in the same model. In this case, preferred cluster activation due to stimulus onset generates an increase in pair-wise correlations that reduce dimensionality. Note that the two properties (trial-to-trial variability and dimensionality) are conceptually distinct. An ensemble of Poisson spike trains can be highly correlated (hence have low dimensionality), yet the Fano Factor of each spike train will still be 1 (hence high),

independently of the correlations among neurons. In a recurrent network, however, dimensionality and trial-to-trial variability may become intertwined and exhibit similar properties, such as the stimulus-induced reduction observed in a model with clustered connectivity. A deeper investigation of the link between dimensionality and trial-to-trial variability in recurrent networks is left for future studies.

Alternative Definitions of Dimensionality

Following (Abbott et al., 2011) we have defined dimensionality (Equation 2) as the dimension of an effective linear subspace of firing rate vectors containing the most variance of the neural activity. It differs somewhat from the typical dimensionality reduction based on PCA that retains only the number of eigenvectors explaining a predefined amount of variance (see Broome et al., 2006; Geffen et al., 2009), because Equation (2) includes contribution from all eigenvalues. Moreover, we have computed the firing rate correlations in bins of variable width that match the duration of the HMM states. Although, our main results do no depend on bin size (see Results' section "Time course of dimensionality as a function of ensemble size"), the actual value of dimensionality decreases with increasing bin duration. Thus, any choice of bin size (e.g., 200 ms in **Figures 3F,G**) remains somewhat arbitrary. A better method is to use a variable bin size as dictated by the HMM analysis, as done in **Figures 3B–D**. This method also prevents diluting correlations among firing rates that would occur if one neuron were to change state inside the current bin, because during a hidden state the firing rates of the neurons are constant (by definition). Thus, this provides a principled adaptive procedure for selecting the bin size and eliminates the dependence of dimensionality on the bin width used for the analysis.

Other definitions of neural dimensionality have been proposed in the literature, which aim at capturing different properties of the neural activity, typically during stimulus-evoked activity. A measure of dimensionality related to ours, and referred to as "complexity," was introduced in Cadieu et al. (2013). According to their definition, population firing rate vectors from all evoked conditions were first decomposed along their kernel Principal Components (Montavon et al., 2011). A linear classifier was then trained on an increasing number of leading PCs in order to perform a discrimination task, where the number of PCs used was defined as the complexity of the representation. In general, the classification accuracy improves with increasing complexity, and it may saturate when all PCs containing relevant features are used—with the remaining PCs representing noise or information irrelevant to the task. Reaching high accuracy at low complexity implies good generalization performance, i.e., the ability to classify novel variations of a stimulus in the correct category. Neural representations in monkey inferotemporal cortex (IT) were found to require lower complexity than in area V4, confirming IT's premier role in classifying visual objects despite large variations in shape, orientation and background (Cadieu et al., 2013). Complexity relies on a supervised algorithm and is an efficient tool to capture the generalization properties of evoked representations (see DiCarlo et al., 2012) for its relevance to visual object recognition).

A second definition of dimensionality, sometimes referred to as "shattering dimensionality" in the Machine Learning literature, has been used to assess the discrimination properties of the neural representation (Rigotti et al., 2013). Given a set of p firing rate vectors, one can split them into two classes (e.g., white and black colorings) in 2^p different ways, and train a classifier to learn as many of those binary classification labels as possible. The shattering dimensionality is then defined as (the logarithm of) the largest number of binary classifications that can be implemented. This measure of dimensionality was found to drop significantly in monkey prefrontal cortex during the error trials of a recall task, and thus predicts the ability of the monkey to correctly perform the task (Rigotti et al., 2013).

A flexible and informative neural representation is one that achieves a large shattering dimensionality (good discrimination) while keeping a low complexity (good generalization). Note that both complexity and shattering dimensionality represent measures of classification performance in task-related paradigms, and their definition requires a set of evoked conditions to be classified via a supervised learning algorithm. While both definitions could be applied to neural activity in our stimulus-evoked data, their interpretation is not readily extended to periods of ongoing activity, as the latter is not associated to desired targets in a way that can be learned by a classification algorithm. Since our main aim was to compare the dimensionality of ongoing and evoked activity, the unsupervised approach of Abbott et al. (2011) and their notion of "effective" dimensionality was better suited for our analysis. A related definition of dimensionality has been used by Gao and Ganguli (2015) to investigate neural representations of movements in motor cortex.

Many measures of dimensionality used in the literature (including ours and some of those discussed above) are based on pair-wise correlations. However, neural activity is known to give rise also to higher-order correlations (Martignon et al., 2000). Given that the extent and relevance of higher-order correlations is actively debated (Schneidman et al., 2006; Staude et al., 2010), it would be useful to include them in measures of dimensionality. This is left for a future study.

Ongoing Activity and Task Complexity

The relationship between ongoing and stimulus-evoked activity has been linked to the functional connectivity of local cortical circuits, and their mutual relationship has been the object of both theoretical and experimental investigations, often with contrasting conclusions (e.g., Arieli et al., 1996; Tsodyks et al., 1999; Kenet et al., 2003; Luczak et al., 2009; Tkacik et al., 2010; Berkes et al., 2011; Mazzucato et al., 2015). Here, we have focused on the dimensionality of ongoing and evoked activity and have shown that neural activity during ongoing periods occupies a space of larger dimensionality compared to evoked activity. Although, based on a different measure of dimensionality, recent results on the relation between the dimensionality of evoked activity and task complexity suggest that evoked dimensionality is roughly equal to the number of task conditions (Rigotti et al., 2013). It is natural to ask whether the dimensionality of ongoing activity provides an estimate of the complexity of the hardest task

that can be supported by the neural activity. Moreover, based on the clustered network model, the presence of clusters imposes an upper value $d \leq Q/\rho^2$ during ongoing activity, suggesting that a discrimination task with up to $\propto Q$ different conditions may be supported. The experience of taste consumption is by itself multidimensional, including chemo- and oro-sensory aspects (i.e., taste identity Jezzini et al., 2013, and concentration Sadacca et al., 2012, texture, temperature, Yamamoto et al., 1981, 1988) as well as psychological aspects (hedonic value Katz et al., 2001; Grossman et al., 2008, anticipation Samuelsen et al., 2012; Gardner and Fontanini, 2014, novelty Inberg et al., 2013; Bermudez-Rattoni, 2014, and satiety effects de Araujo et al., 2006). It is tempting to speculate that neural activity during ongoing periods explores all these different dimensions, while evoked activity is confined to the features of the particular taste being delivered or attended in a specific context.

Establishing a precise experimental and theoretical link between the number of clusters and task complexity is an important question left for future studies.

ACKNOWLEDGMENTS

This work was supported by a National Institute of Deafness and Other Communication Disorders Grant K25-DC013557 (LM), by the Swartz Foundation Award 66438 (LM), by National Institute of Deafness and Other Communication Disorders Grant R01-DC010389 (AF), by a Klingenstein Foundation Fellowship (AF), and by a National Science Foundation Grant IIS-1161852 (GL). We thank Drs. Stefano Fusi and Memming Park for useful discussions and David Ecker at the Research Technologies DoIT of Stony Brook University for access to its computational resources.

REFERENCES

Abbott, L. F., Rajan, K., and Sompolinsky, H. (2011). "Interactions between intrinsic and stimulus-evoked activity in recurrent neural networks," in *The Dynamic Brain: An Exploration of Neuronal Variability and its Functional Significance,* eds D. L. Glanzman and M. Ding (New York, NY: Oxford University Press), 65–82.

Abeles, M., Bergman, H., Gat, I., Meilijson, I., Seidemann, E., Tishby, N., et al. (1995). Cortical activity flips among quasi-stationary states. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8616–8620. doi: 10.1073/pnas.92.19.8616

Amit, D. J., and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* 7, 237–252. doi: 10.1093/cercor/7.3.237

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. doi: 10.1126/science.273.5283.1868

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331, 83–87. doi: 10.1126/science.1195870

Bermudez-Rattoni, F. (2014). The forgotten insular cortex: its role on recognition memory formation. *Neurobiol. Learn. Mem.* 109, 207–216. doi: 10.1016/j.nlm.2014.01.001

Broome, B. M., Jayaraman, V., and Laurent, G. (2006). Encoding and decoding of overlapping odor sequences. *Neuron* 51, 467–482. doi: 10.1016/j.neuron.2006.07.018

Brunel, N., and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* 11, 1621–1671. doi: 10.1162/089976699300016179

Brunel, N., and Sergi, S. (1998). Firing frequency of leaky intergrate-and-fire neurons with synaptic current dynamics. *J. Theor. Biol.* 195, 87–95. doi: 10.1006/jtbi.1998.0782

Cadieu, C. F., Hong, H., Yamins, D., Pinto, N., Majaj, N. J., and DiCarlo, J. J. (2013). *The Neural Representation Benchmark and its Evaluation on Brain and Machine. arXiv:13013530 [csNE].*

Chapin, J. K., and Nicolelis, M. A. (1999). Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *J. Neurosci. Methods* 94, 121–140. doi: 10.1016/S0165-0270(99)00130-2

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I., and Shenoy, K. V. (2010a). Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron* 68, 387–400. doi: 10.1016/j.neuron.2010.09.015

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., et al. (2010b). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378. doi: 10.1038/nn.2501

Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G., and Shenoy, K. V. (2006). Neural variability in premotor cortex provides a signature of motor preparation. *J. Neurosci.* 26, 3697–3712. doi: 10.1523/JNEUROSCI.3762-05.2006

Cohen, M. R., and Maunsell, J. H. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 12, 1594–1600. doi: 10.1038/nn.2439

Curti, E., Mongillo, G., La Camera, G., and Amit, D. J. (2004). Mean field and capacity in realistic networks of spiking neurons storing sparsely coded random memories. *Neural Comput.* 16, 2597–2637. doi: 10.1162/0899766042321805

de Araujo, I. E., Gutierrez, R., Oliveira-Maia, A. J., Pereira, A., Nicolelis, M. A., and Simon, S. A. (2006). Neural ensemble coding of satiety states. *Neuron* 51, 483–494. doi: 10.1016/j.neuron.2006.07.009

Deco, G., and Hugues, E. (2012). Neural network mechanisms underlying stimulus driven variability reduction. *PLoS Comput. Biol.* 8:e1002395. doi: 10.1371/journal.pcbi.1002395

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

Durstewitz, D., Vittoz, N. M., Floresco, S. B., and Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* 66, 438–448. doi: 10.1016/j.neuron.2010.03.029

Escola, S., Fontanini, A., Katz, D., and Paninski, L. (2011). Hidden markov models for the stimulus-response relationships of multistate neural systems. *Neural Comput.* 23, 1071–1132. doi: 10.1162/NECO_a_00118

Farkhooi, F., Froese, A., Muller, E., Menzel, R., and Nawrot, M. P. (2013). Cellular adaptation facilitates sparse and reliable coding in sensory pathways. *PLoS Comput. Biol.* 9:e1003251. doi: 10.1371/journal.pcbi.1003251

Fontanini, A., and Katz, D. B. (2005). 7 to 12 Hz activity in rat gustatory cortex reflects disengagement from a fluid self-administration task. *J. Neurophysiol.* 93, 2832–2840. doi: 10.1152/jn.01035.2004

Fontanini, A., and Katz, D. B. (2006). State-dependent modulation of time-varying gustatory responses. *J. Neurophysiol.* 96, 3183–3193. doi: 10.1152/jn.00804.2006

Fontanini, A., and Katz, D. B. (2008). Behavioral states, network states, and sensory response variability. *J. Neurophysiol.* 100, 1160–1168. doi: 10.1152/jn.90592.2008

Fourcaud, N., and Brunel, N. (2002). Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput.* 14, 2057–2110. doi: 10.1162/089976602320264015

Fusi, S., and Mattia, M. (1999). Collective behavior of networks with linear (VLSI) integrate-and-fire neurons. *Neural Comput.* 11, 633–652. doi: 10.1162/089976699300016601

Ganguli, S., Bisley, J. W., Roitman, J. D., Shadlen, M. N., Goldberg, M. E., and Miller, K. D. (2008). One-dimensional dynamics of attention and decision making in LIP. *Neuron* 58, 15–25. doi: 10.1016/j.neuron.2008.01.038

Ganguli, S., and Sompolinsky, H. (2012). Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Ann. Rev. Neurosci.* 35, 485–508. doi: 10.1146/annurev-neuro-062111-150410

Gao, P., and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* 32, 148–155. doi: 10.1016/j.conb.2015.04.003

Gardner, M. P., and Fontanini, A. (2014). Encoding and tracking of outcome-specific expectancy in the gustatory cortex of alert rats. *J. Neurosci.* 34, 13000–13017. doi: 10.1523/JNEUROSCI.1820-14.2014

Geffen, M. N., Broome, B. M., Laurent, G., and Meister, M. (2009). Neural encoding of rapidly fluctuating odors. *Neuron* 61, 570–586. doi: 10.1016/j.neuron.2009.01.021

Giugliano, M., Darbon, P., Arsiero, M., Luscher, H. R., and Streit, J. (2004). Single-neuron discharge properties and network activity in dissociated cultures of neocortex. *J. Neurophysiol.* 92, 977–996. doi: 10.1152/jn.00067.2004

Giugliano, M., La Camera, G., Fusi, S., and Senn, W. (2008). The response of cortical neurons to *in vivo*-like input current: theory and experiment: II. Time-varying and spatially distributed inputs. *Biol. Cybern.* 99, 303–318. doi: 10.1007/s00422-008-0270-9

Grossman, S. E., Fontanini, A., Wieskopf, J. S., and Katz, D. B. (2008). Learning-related plasticity of temporal coding in simultaneously recorded amygdala-cortical ensembles. *J. Neurosci.* 28, 2864–2873. doi: 10.1523/JNEUROSCI.4063-07.2008

Gur, M., Beylin, A., and Snodderly, D. M. (1997). Response variability of neurons in primary visual cortex (V1) of alert monkeys. *J. Neurosci.* 17, 2914–2920.

Holland, P. W., and Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Commun. Statist. Theory Methods A* 6, 813–827. doi: 10.1080/03610927708827533

Horst, N. K., and Laubach, M. (2013). Reward-related activity in the medial prefrontal cortex is driven by consumption. *Front. Neurosci.* 7:56. doi: 10.3389/fnins.2013.00056

Inberg, S., Elkobi, A., Edri, E., and Rosenblum, K. (2013). Taste familiarity is inversely correlated with Arc/Arg3.1 hemispheric lateralization. *J. Neurosci.* 33, 11734–11743. doi: 10.1523/JNEUROSCI.0801-13.2013

Jezzini, A., Mazzucato, L., La Camera, G., and Fontanini, A. (2013). Processing of hedonic and chemosensory features of taste in medial prefrontal and insular networks. *J. Neurosci.* 33, 18966–18978. doi: 10.1523/JNEUROSCI.2974-13.2013

Jones, L. M., Fontanini, A., Sadacca, B. F., Miller, P., and Katz, D. B. (2007). Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18772–18777. doi: 10.1073/pnas.0705546104

Katz, D. B., Simon, S. A., and Nicolelis, M. A. (2001). Dynamic and multimodal responses of gustatory cortical neurons in awake rats. *J. Neurosci.* 21, 4478–4489.

Kemere, C., Santhanam, G., Yu, B. M., Afshar, A., Ryu, S. I., Meng, T. H., et al. (2008). Detecting neural-state transitions using hidden Markov models for motor cortical prostheses. *J. Neurophysiol.* 100, 2441–2452. doi: 10.1152/jn.00924.2007

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature* 425, 954–956. doi: 10.1038/nature02078

Kiani, R., Cueva, C. J., Reppas, J. B., Peixoto, D., Ryu, S. I., and Newsome, W. T. (2015). Natural grouping of neural responses reveals spatially segregated clusters in prearcuate cortex. *Neuron* 85, 1359–1373. doi: 10.1016/j.neuron.2015.02.014

La Camera, G., Giugliano, M., Senn, W., and Fusi, S. (2008). The response of cortical neurons to *in vivo*-like input current: theory and experiment : I. Noisy inputs with stationary statistics. *Biol. Cybern.* 99, 279–301. doi: 10.1007/s00422-008-0272-7

La Camera, G., Rauch, A., Lüscher, H. R., Senn, W., and Fusi, S. (2004). Minimal models of adapted neuronal response to *in vivo*-like input currents. *Neural Comput.* 16, 2101–2124. doi: 10.1162/0899766041732468

La Camera, G., Rauch, A., Thurbon, D., Lüscher, H. R., Senn, W., and Fusi, S. (2006). Multiple time scales of temporal response in pyramidal and fast spiking cortical neurons. *J. Neurophysiol.* 96, 3448–3464. doi: 10.1152/jn.00453.2006

Lánský, P., and Sato, S. (1999). The stochastic diffusion models of nerve membrane depolarization and interspike interval generation. *J. Peripher. Nerv. Syst.* 4, 27–42.

Litwin-Kumar, A., and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* 15, 1498–1505. doi: 10.1038/nn.3220

Luczak, A., Bartho, P., and Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* 62, 413–425. doi: 10.1016/j.neuron.2009.03.014

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. doi: 10.1162/neco.2008.02-08-713

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

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. doi: 10.1162/089976600300014872

Mascaro, M., and Amit, D. J. (1999). Effective neural response function for collective population states. *Network* 10, 351–373. doi: 10.1088/0954-898X_10_4_305

Mazzucato, L., Fontanini, A., and La Camera, G. (2015). Dynamics of multistable states during ongoing and evoked cortical activity. *J. Neurosci.* 35, 8214–8231. doi: 10.1523/JNEUROSCI.4819-14.2015

McDonnell, M. D., and Ward, L. M. (2011). The benefits of noise in neural systems: bridging theory and experiment. *Nat. Rev. Neurosci.* 12, 415–426. doi: 10.1038/nrn3061

Miller, P., and Katz, D. B. (2010). Stochastic transitions between neural states in taste processing and decision-making. *J. Neurosci.* 30, 2559–2570. doi: 10.1523/JNEUROSCI.3047-09.2010

Montavon, G., Braun, M. L., and Müller, K.-R. (2011). Kernel analysis of deep networks. *J. Mach. Learn. Res.* 12, 2563–2581.

Moreno-Bote, R. (2014). Poisson-like spiking in circuits with probabilistic synapses. *PLoS Comput. Biol.* 10:e1003522. doi: 10.1371/journal.pcbi.1003522

Nienborg, H., Cohen, M. R., and Cumming, B. G. (2012). Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Ann. Rev. Neurosci.* 35, 463–483. doi: 10.1146/annurev-neuro-062111-150403

Phillips, M. I., and Norgren, R. (1970). A rapid method for permanent implantation of an intraoral fistula in rats. *Behav. Res. Methods Instrum.* 2:124. doi: 10.3758/BF03211020

Ponce-Alvarez, A., Nácher, V., Luna, R., Riehle, A., and Romo, R. (2012). Dynamics of cortical neuronal ensembles transit from decision making to storage for later report. *J. Neurosci.* 32, 11956–11969. doi: 10.1523/JNEUROSCI.6176-11.2012

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes the Art of Scientific Computing, 3rd Edn.* Cambridge;New York, NY: Cambridge University Press.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 257–286. doi: 10.1109/5.18626

Rauch, A., La Camera, G., Luscher, H. R., Senn, W., and Fusi, S. (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to *in vivo*-like input currents. *J. Neurophysiol.* 90, 1598–1612. doi: 10.1152/jn.00293.2003

Renart, A., Brunel, N., and Wang, X.-J. (2004). "Mean-field theory of recurrent cortical networks: from irregularly spiking neurons to working memory," in *Computational Neuroscience: A Comprehensive Approach,* ed J. Feng (Boca Raton, FL: CRC), 431–490.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. doi: 10.1126/science.1179850

Richardson, M. J. (2004). The effects of synaptic conductance on the voltage distribution and firing rate of spiking neurons. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 69:051918. doi: 10.1103/PhysRevE.69.051918

Rickert, J., Riehle, A., Aertsen, A., Rotter, S., and Nawrot, M. P. (2009). Dynamic encoding of movement direction in motor cortical neurons. *J. Neurosci.* 29, 13870–13882. doi: 10.1523/JNEUROSCI.5441-08.2009

Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., et al. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590. doi: 10.1038/nature12160

Sadacca, B. F., Rothwax, J. T., and Katz, D. B. (2012). Sodium concentration coding gives way to evaluative coding in cortex and amygdala. *J. Neurosci.* 32, 9999–10011. doi: 10.1523/JNEUROSCI.6059-11.2012

Samuelsen, C. L., Gardner, M. P., and Fontanini, A. (2012). Effects of cue-triggered expectation on cortical processing of taste. *Neuron* 74, 410–422. doi: 10.1016/j.neuron.2012.02.031

Schneidman, E., Berry, M. J. II., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012. doi: 10.1038/nature04701

Seidemann, E., Meilijson, I., Abeles, M., Bergman, H., and Vaadia, E. (1996). Simultaneously recorded single units in the frontal cortex go through sequences of discrete and stable states in monkeys performing a delayed localization task. *J. Neurosci.* 16, 752–768.

Shadlen, M. N., and Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* 4, 569–579. doi: 10.1016/0959-4388(94)90059-0

Staude, B., Rotter, S., and Grün, S. (2010). CuBIC: cumulant based inference of higher-order correlations in massively parallel spike trains. *J. Comput. Neurosci.* 29, 327–350. doi: 10.1007/s10827-009-0195-x

Tkacik, G., Prentice, J. S., Balasubramanian, V., and Schneidman, E. (2010). Optimal population coding by noisy spiking neurons. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14419–14424. doi: 10.1073/pnas.1004906107

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. doi: 10.1126/science.286.5446.1943

Tuckwell, H. C. (1988). *Introduction to Theoretical Neurobiology*. Cambridge, UK: Cambridge University Press.

van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726. doi: 10.1126/science.274.5293.1724

van Vreeswijk, C., and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.* 10, 1321–1371. doi: 10.1162/089976698300017214

White, B., Abbott, L. F., and Fiser, J. (2012). Suppression of cortical neural variability is stimulus- and state-dependent. *J. Neurophysiol.* 108, 2383–2392. doi: 10.1152/jn.00723.2011

Yamamoto, T., Matsuo, R., Kiyomitsu, Y., and Kitamura, R. (1988). Sensory inputs from the oral region to the cerebral cortex in behaving rats: an analysis of unit responses in cortical somatosensory and taste areas during ingestive behavior. *J. Neurophysiol.* 60, 1303–1321.

Yamamoto, T., Yuyama, N., and Kawamura, Y. (1981). Cortical neurons responding to tactile, thermal and taste stimulations of the rat's tongue. *Brain Res.* 221, 202–206. doi: 10.1016/0006-8993(81)91075-1

Zucchini, W., and MacDonald, I. L. (2009). *Hidden Markov Models for Time Series : An Introduction Using R*. Boca Raton, FL: CRC Press. doi: 10.1201/9781420010893

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.

frontiers in Systems Neuroscience

ORIGINAL RESEARCH
published: 14 January 2016
doi: 10.3389/fnsys.2015.00184

Statistical Frequency-Dependent Analysis of Trial-to-Trial Variability in Single Time Series by Recurrence Plots

Tamara Tošić [1, 2, 3], Kristin K. Sellers [4, 5], Flavio Fröhlich [4, 5, 6, 7, 8], Mariia Fedotenkova [1, 2, 3], Peter beim Graben [9, 10] and Axel Hutt [1, 2, 3]*

[1] Team Neurosys, Inria, Villers-lès-Nancy, France, [2] Loria, Centre National de la Recherche Scientifique, UMR no 7503, Villers-lès-Nancy, France, [3] Université de Lorraine, Loria, UMR no 7503, Villers-lès-Nancy, France, [4] Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [5] Neurobiology Curriculum, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [6] Department of Cell Biology and Physiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [7] Department of Biomedical Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [8] Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [9] Department of German Studies and Linguistics, Berlin, Germany, [10] Bernstein Center for Computational Neuroscience, Berlin, Germany

For decades, research in neuroscience has supported the hypothesis that brain dynamics exhibits recurrent metastable states connected by transients, which together encode fundamental neural information processing. To understand the system's dynamics it is important to detect such recurrence domains, but it is challenging to extract them from experimental neuroscience datasets due to the large trial-to-trial variability. The proposed methodology extracts recurrent metastable states in univariate time series by transforming datasets into their time-frequency representations and computing recurrence plots based on instantaneous spectral power values in various frequency bands. Additionally, a new statistical inference analysis compares different trial recurrence plots with corresponding surrogates to obtain statistically significant recurrent structures. This combination of methods is validated by applying it to two artificial datasets. In a final study of visually-evoked Local Field Potentials in partially anesthetized ferrets, the methodology is able to reveal recurrence structures of neural responses with trial-to-trial variability. Focusing on different frequency bands, the δ-band activity is much less recurrent than α-band activity. Moreover, α-activity is susceptible to pre-stimuli, while δ-activity is much less sensitive to pre-stimuli. This difference in recurrence structures in different frequency bands indicates diverse underlying information processing steps in the brain.

Keywords: trial-to-trial variability, time-frequency analysis, local field potentials, recurrence plot analysis, statistical inference, surrogate data, anesthesia, ferret

Edited by:
Emili Balaguer-Ballester, Bournemouth University, UK and Bernstein Center for Computational Neuroscience Heidelberg-Mannheim, Germany

Reviewed by:
Jorge Otero-Millan, Johns Hopkins University, USA
Hazem Toutounji, Central Institute of Mental Health, Germany

****Correspondence:***
Tamara Tošić
tamtos@gmail.com

Received: *31 March 2015*
Accepted: *18 December 2015*
Published: *14 January 2016*

Citation:
Tošić T, Sellers KK, Fröhlich F, Fedotenkova M, beim Graben P and Hutt A (2016) Statistical Frequency-Dependent Analysis of Trial-to-Trial Variability in Single Time Series by Recurrence Plots. Front. Syst. Neurosci. 9:184. doi: 10.3389/fnsys.2015.00184

1. INTRODUCTION

Investigation of metastable states (MS) and transients of complex dynamical systems has become increasingly important over the last decades. In this context, dynamical systems spend longer time intervals in MSs than in transients between MSs. The large interest in studying such states comes from the belief that a complex temporal behavior of systems may be decomposed into

a simple sequence of alternating MSs and transients between them. This reduced description is a model that captures the essential dynamic elements of rather complex underlying dynamics. Applications range from spin glasses (Larralde and Leyvraz, 2005) to molecular configurations (Deuflhard and Weber, 2005) and geoscientific applications (Froyland et al., 2007). In neuroscience, the related concept of sequential metastable attractors has received increasing attention in the last years (Friston, 1997; Oullier and Kelso, 2006; Rabinovich et al., 2008b; Yildiz and Kiebel, 2011; Hudson et al., 2014; Tognoli and Kelso, 2014). Primarily, works are motivated by the experimental observation of signal features showing alternations of dynamical behavior at fast and slow time scales (Hutt and Riedel, 2003; Hutt, 2004; Mazor and Laurent, 2005; Allefeld et al., 2009).

Originally the concept of metastability refers to slow relaxation dynamics in statistical physics (Larralde and Leyvraz, 2005; Tokman et al., 2011). In a much wider sense, this notion is nowadays used for regions in the phase space of a dynamical system with relatively large dwell that are connected by transients (Friston, 1997; Rabinovich et al., 2008b; Tognoli and Kelso, 2014). Paradigmatic examples for those MSs are *almost invariant sets* (Froyland, 2005) and *recurrence domains* (beim Graben and Hutt, 2013), such as saddles connected by heteroclinic trajectories (Rabinovich et al., 2008a) or the "wings" of the Lorenz attractor (Lorenz, 1963). For this attractor in particular, it is attractive itself and has two recurrence domains centered around two unstable foci. Geometrically, these domains are spatially separated and the system's trajectory alternately approaches to and departs from the foci. The system spends much longer time in the vicinity of a focus compared to transient intervals between the two foci. Therefore, one may refer to a Lorenz wing as to a MS: the system remains for a longer time in one partition cell of the phase space before it performs a rapid transition to another partition cell of the phase space. A MS is thus identified with a recurrence domain, while non-recurrent portions of a trajectory can be compared with transients.

In neuroscience, metastability assumed increasing experimental evidence over recent years. Lehmann et al. (1987), Wackermann et al. (1993) observed sequences of metastable electroencephalogram (EEG) topographies, which they called *brain microstates*. Hutt and Riedel (2003), Hutt (2004), beim Graben and Hutt (2015) argued that components of the event-related brain potentials (ERPs) reflecting perceptional and cognitive processes could be identified with metastable brain states. Mazor and Laurent, for instance, reported sequences of metastable states in a reconstructed activation space of the locust's neural odor circuit (Mazor and Laurent, 2005). Allefeld et al. (2009) were able to detect metastable states in epileptic EEG time series through spectral clustering methods, and most recently, Hudson et al. (2014) revealed metastable transition networks in the recovery from anesthesia. Consequently, to understand underlying neural mechanisms much better, it is necessary to develop advanced techniques to detect these recurrence structures in experimental time series.

For the identification of metastability in time series, their characteristic slow time scales must be separated from the fast dynamics of phase space trajectories. The method known as *Perron clustering* (Deuflhard and Weber, 2005), separates the system's phase states into partitions that can approximate Markov chain states (Deuflhard and Weber, 2005; Froyland, 2005; Larralde and Leyvraz, 2005; Gaveau and Schulman, 2006; Allefeld et al., 2009). Applying spectral clustering methods to the resulting transition matrix yields the time scales of the process, while their corresponding (left-)eigenvectors allow the unification of cells into a partition of metastable states (Gaveau and Schulman, 2006; Allefeld et al., 2009). Another approach by Hutt and Riedel (2003) utilizes the slowing-down of the system's trajectory in the vicinity of saddles by means of phase space clustering. Most recently, beim Graben and Hutt suggested to combine recurrence plot techniques and symbolic dynamics in order to partition a system's phase space into its recurrence domains (beim Graben and Hutt, 2013, 2015). The application of the latter method to experimental event-related potentials has identified metastable attractors to so-called ERP-components, known to reflect cognitive processing stages in neural information processing.

Developing novel analysis tools for representation and tracking of non-linear transient patterns faces numerous challenges, such as reducing the signal dimensionality while preserving the information significant for the detection task or building methods robust to acquisition noise. Recurrence analysis has been used for identifying transient patterns in experimental EEG (Shalbaf et al., 2015), for classifying patients based on EEG time series (McCarthy et al., 2014) and for prediction of responses during anesthesia (Huang et al., 2006). A key feature of recurrence analysis is to identify sequential states in a multi-dimensional signal space, as shown in most previous studies (beim Graben and Hutt, 2013, 2015). If the experimental data under study is multi-dimensional, for instance a multi-channel EEG recording, the data serves directly as the input to the recurrence analysis. However, it is not valid to compute recurrence plots in the case of univariate time series and hence the data can not be analyzed directly. Therefore, it is necessary to transform the univariate signal to a multivariate (multi-dimensional) signal. Typically this is done by delay-embedding techniques (Webber and Zbilut, 1994; Iwanski and Bradley, 1998) inspired by Takens' theorem (Takens, 1981). The corresponding embedding dimension and delay time in these techniques are chosen rather independent from the dynamic features of the data since typically these are not known *a priori*.

In neuroscience, patterns occurring in certain frequency bands play distinct roles in neural information processing (Kandel et al., 2000; Schnitzler and Gross, 2005). We argue that this additional knowledge can be taken into account and the present work proposes a novel technique based on time-frequency representations of univariate signals. Here, the signal is transformed into its time-frequency representation of spectral power which spans a new phase space in which the signal trajectory evolves. Hence, one may call this transformation *spectral power embedding* since the new phase space encodes instantaneous power in certain frequency bands. The additional advantage of this approach is that it permits to analyse the recurrence structure of data in selected frequency bands. For completeness, we mention that a signal is fully defined by its instantaneous amplitude and phase.

In this work we propose a new method for the detection of metastable states in univariate neural signals. To obtain statistically significant evidence of recurrence structures in signals, we conduct a statistical test over the set of novel, frequency-selective recurrence plots (RP). Below we describe methodologies for building such frequency-selective RPs and performing statistical inference tests. These tests indicate how stable the recurrence plots are with respect to trial-to-trial variability. This novel statistical evaluation is necessary in the analysis of neurophysiological data, since trial-to-trial variability is a well-known experimental finding in such signals. In our work we analyse synthetic transient oscillations and one state variable of the Lorenz attractor involving acquisition noise to validate the methodology. Finally, the study of experimental Local Field Potentials obtained in partially anesthetized ferrets (*Mustela putorius furo*) during a visual stimulus experiment allows to extract new insights into neural information processing. For instance, we show that temporal recurrence occurs in the α-frequency band but not in the δ-frequency band. This result suggests that in the α-band the brain processes the information step-wise (*state by state*) while no step-wise process is performed in the δ-band.

2. MATERIALS AND METHODS

In this section we introduce the novel method for studying temporal recurrences common in recurrence plots of different trials. Section 2.1 introduces classical recurrence plots and describes corresponding parameters. Then, we provide a novel method to compute recurrence plots from their time-frequency representations. In Section 2.2 we propose the statistical test method that analyses the similarity of RPs and finds their statistically significant parts. Finally, in Section 2.3 we describe the datasets used in this work.

2.1. Recurrence Plots and Novel Time-Frequency Representations

Recurrence is a fundamental property of dynamical systems which characterizes the behavior of the system in phase space (Poincaré, 1890). A recurrent signal instance is a moment in time when the trajectory returns to a neigborhood of a location in phase space it has already visited previously.

Deterministic dynamical systems are described by their trajectory. A trajectory $\underline{\mathbf{x}}(t) \in \mathbb{R}^n$, $t \in \mathbb{R}$ is sampled at times $t = i\Delta t, i \in \{1, 2, \ldots, N\}$, where Δt is the sampling time interval and N is the total number of samples. For notation simplicity, in this paper we denote the signal sample $\underline{\mathbf{x}}(i\Delta t)$ by $\underline{\mathbf{x}}(i)$. Then, a recurrence plot (RP) is defined as the $N \times N$ matrix $\mathbf{R}$, whose elements $r_{i,j}$ take values $r_{i,j} = 1$ when two trajectory samples lay within the open ball $\mathcal{B}(\epsilon)$ of radius ϵ

$$r_{i,j} = \begin{cases} 1, & \text{if } d(\underline{\mathbf{x}}(i), \underline{\mathbf{x}}(j)) < \epsilon, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $d(\cdot, \cdot)$ is a distance function and $i, j \in \{1, \ldots, N\}$. Hence, recurrence plots are two-dimensional binary matrices obtained by distance based thresholding and its elements take values $r_{i,j} \in \{0, 1\}$. In this work, pixels in RPs will be color-coded white for values $r_{i,j} = 0$ and black otherwise.

For an arbitrary chosen ϵ value we can not guarantee that some of the significant dynamic features are not discarded by thresholding. To minimize such a thresholding error, we compute the optimal threshold value ϵ^* which maximizes the symbolic entropy for a given distance function, as proposed in beim Graben and Hutt (2013). In more detail, under the assumption that recurrence domains are uniformly distributed for a given recurrence plot, the method constructs disjunct and transitive symbolic recurrence plot matrices from multivariate data. This method permits to identify MSs in a recurrence plot and maps each state (and the transients between the states) to a symbol. Consequently, one maps the high-dimensional dynamics of the system to a sequence of symbols. Let p_k be the probability of the occurrence of the state k, i.e., the number of the occurrences of the symbol k divided by the total number of occurrences of all symbols. Then maximizing the entropy

$$H(\epsilon) = -\frac{1}{S_{k,\epsilon}} \sum_{k=1}^{S_{k,\epsilon}} p_k \log(p_k), \tag{2}$$

for a range of ϵ-values yields that value of ϵ for which the distribution of occurrence probabilities $\{p_k\}$ approaches uniformity, i.e., for which all states are equally probable. Here, $S_{k,\epsilon}$ is the number of states for a given ϵ. Then the optimal value

$$\epsilon^* = \arg\max_{\epsilon} H(\epsilon)$$

maximizes the entropy of the extracted symbolic sequence and hence the recurrence structure of the data. This optimal value is computed for each dataset separately.

After defining conventional RPs and computation of the optimal parameter ϵ, the remaining part of this section focusses on how to build frequency-selective recurrence plots. Many biophysiological signals have characteristic frequency signatures. For example, the human heart beats about sixty times per minute in average, i.e., at the frequency of 1 Hz. Another example are eye blinks that induce signal changes in the α-frequency band (frequencies in the interval 8–12 Hz) in EEG recordings. To take into account the distinct signatures of spectral bands present in neural signals, we propose a novel concept for building recurrence plots from time-frequency signal representations, instead of building them directly from univariate data or constructing them by employing delay-embedding techniques. Such representations, in general adapted for non-stationary signal analysis, give insights into frequency bands of importance and provide additional flexibility to recurrence plot analysis that is not present in time-domain, for e.g., the possibility to weight the importance of some frequency bands. In the literature there are several ways to choose values of the frequency bands. We use the following frequency interval definitions: the δ-frequency band denotes the interval [0.5 Hz; 4 Hz], the θ-frequency band the interval [4 Hz; 8 Hz], α-band [8 Hz; 12 Hz], β-band [12 Hz; 20 Hz], and the γ-band denotes the interval [20 Hz; 40 Hz].

We build novel recurrence plots in three steps, as shown in **Figure 1**: (i) we expand the set of T univariate trials

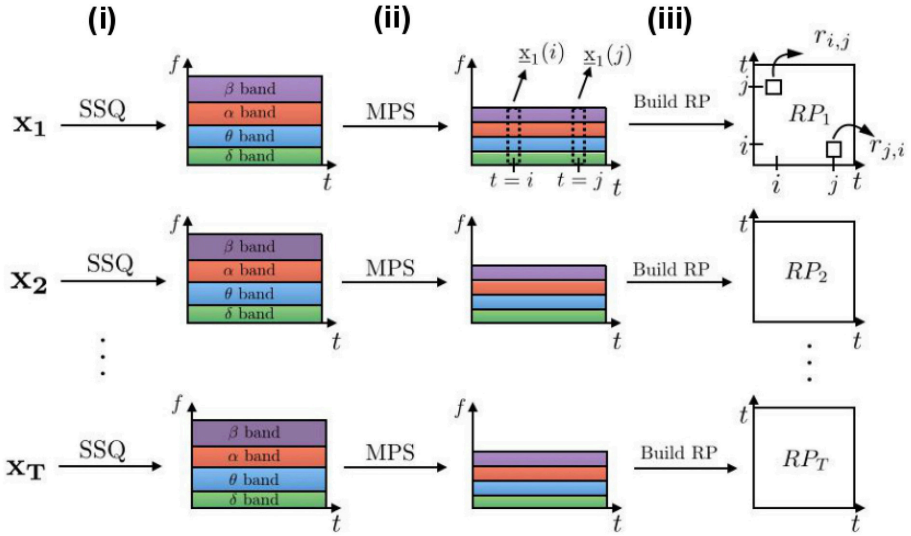


FIGURE 1 | Building frequency-selective recurrence plots from *T* time series. Processing blocks are represented by arrows: **(i)** SSQ is a synchrosqueezing transform block used to obtain time-frequency representations of signals; **(ii)** the MPS block computes mean values of the power spectrum for each of the chosen frequency bands, which reduces the signal dimensionality; **(iii)** This signal is the basis for the recurrence analysis leading to recurrence plots $\{RP_k\}_{k=1}^{T}$.

$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$ to their corresponding time-frequency domains; (ii) we compute the mean power of the spectrum over certain sets of frequencies. These mean power time series may be called $\mathbf{y}_j(t) \in \mathbb{R}^S$, $j = 1, \ldots, T$, where S is the number of frequency bands; (iii) we compute recurrence plots by computing distances between vectors $\underline{y}(i)$ and $\underline{y}(j)$, $i, j \in \{1, \ldots, N\}$ as in Equation (1). In the following paragraphs we describe these blocks in more detail.

The first block in **Figure 1** provides a time-frequency representation of the signal. In classical spectrogram calculations, the stronger (weaker) is the localization of signals in time, the larger (smaller) are their localization windows in frequency. This effect is called the uncertainty principle implied in the Fourier transform. A synchrosqueezing (SSQ) transform overcomes this deficiency by performing wavelet-based filtering and signal power reassignment to the appropriate frequencies. In addition, Meignen et al. (2012) and Auger et al. (2013) show that SSQ is superior for processing neural signals when compared to conventional spectral analysis methods, such as continuous wavelet transform or spectrogram. Hence, we use the SSQ transform defined in Section 2.1.1 as the processing block (i) in **Figure 1**.

The second block in **Figure 1** computes the mean value of the power spectrum (MPS) for sets of frequencies, see Section 2.1.2 for details. This is one of the basic features for studying neural signals. We assume that the dynamics of the neural system encoded in frequencies is proportional to the power spectrum in sets of frequencies. This analysis step provides multi-variate time series whose dimension is equal to the resulting vector of averaged frequency bands.

Finally, in the third processing block in the figure we compute recurrence plots from the obtained time-frequency dataset as in Equation (1). If we do not explicitly mention otherwise, we use features from all the frequency bands to compute recurrence plots. In the experimental ferret dataset, we additionally present cases when recurrence plots are calculated from the single frequency band features such as δ- or α-frequency bands, since these bands play an important role in the loss of consciousness under anesthesia.

To summarize, the proposed method for building recurrence plots from time-frequency representations grasps band-related features and allows flexibility in the analysis of particular frequency bands, which is not possible in the classical RP analysis. Our approach however requires additional computations of the synchrosqueezing transform and mean power of the spectrum.

2.1.1. Synchrosqueezing Transform

For completeness of this work, in this section we provide the mathematical definition of the synchrosqueezing transform (Meignen et al., 2012; Auger et al., 2013), that we use as a processing block in the proposed algorithm, see **Figure 1**. We presume that input signals are composed of several components with time-varying oscillatory characteristics. In other words, we assume that signals $f(t)$ can be well approximated with K signal components, $f(t) = \sum_{k=1}^{K} f_k(t) + e(t)$, $f_k(t) = A_k(t)e^{2\pi i \phi_k(t)}$, where $A_k(t)$ and $\phi_k'(t) = \frac{1}{2\pi}\frac{d\phi_k(t)}{dt}$ denote the amplitude and the instantaneous frequency (IF) of each component and $e(t)$ represents a small error. We assume that the components f_k have slowly time-varying amplitudes $A_k(t)$ and sufficiently smooth IFs. These conditions assure that signal components are well separated in frequencies and the complete definition is available in Thakur et al. (2013), Def. II.1 (codes available online in Thakur, 2013).

Let a wavelet $\psi(t)$ be a square integrable and normalized function. Then, its scaled and time-shifted variants $\psi(\frac{t-b}{a})$ represent a set of scaled bandpass filters. In the following,

we denote the frequency of one signal component by $\omega_k \approx 2\pi \frac{d\phi_k(t)}{dt}$. A Continuous wavelet transform (CWT) of the function f at scale a and time shift b is defined by $W_f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\overline{\psi(\frac{t-b}{a})}dt$, which represents a convolution of scaled and band-passed filters with the signal. The shifts of wavelet function are driven by the scale value a. For example, for the first signal component with frequency ω_1, the value of the wavelet coefficient $W_f(a_1, b)$ spreads around the scale factor $a_1 = \frac{\omega_\psi}{\omega_1}$, where ω_ψ is the central wavelet frequency. Therefore, the estimated IF in the neighborhood of this value of the scale is equal to the frequency ω_1. The synchrosqueezing transform $T(\omega_q, b)$ uses estimates of the instantaneous frequency $\omega_f(a, b)$, computed for each scale-time pair (a, b) by $\omega_f(a,b) = -iW_f(a,b)\frac{\partial W_f(a,b)}{\partial b}$ to reallocate the energy of the wavelet coefficients. Let Δa_p ($\Delta\omega$) denote resolution steps in scale (frequency). Then, this transform, defined by $T(\omega_q, b) = \sum_{a_p:|\omega_f(a_p,b)-\omega_q|\leq\Delta\omega/2} W_f(a_p, b)a^{-3/2}\Delta a_p$ enhances frequency localization of oscillating components of the signal and provides more precise time-frequency representations of the signal. In analogy to the spectrogram used in classical short-time Fourier analysis, we plot values

$$S(\omega_q, b) = |T(\omega_q, b)|^2 \tag{3}$$

for each pair (ω_q, b) in time-frequency plots, see **Figures 5A,E**, **6A,C**.

2.1.2. Mean power spectrum

For each frequency band with Q components, the mean power spectrum value is defined by

$$MPS(t) = \frac{1}{Q}\sum_{q=1}^{Q} S(\omega_q, t), \tag{4}$$

where $S(\omega_q, t)$ is defined in Equation (3), ω_q are frequencies of one frequency band and t is time.

2.2. Statistical analysis

We study statistical properties of frequency-selective RPs obtained from time-frequency trial representations. By virtue of noise effects and an expected trial-to-trial variability, recurrence plot structures are expected to vary from trial to trial. To evaluate the recurrence plots statistically, we perform a statistical inference analysis based on a classical chi-squared test (Yates, 1934). To this end, we construct surrogate recurrence plots and employ an inference test.

Classically, surrogate sets of univariate signals (Schreiber and Schmitz, 2000) preserve some of the important features of the original time series, for example the spectrum magnitude, while they replace the phase values by a random sequence of values. The reasoning behind this randomization is that time domain reshaping destroys non-stationarities, so the local spectral components will vary while the global spectrum remains the same. As a consequence, the mean and variance of the signal do not change (Borgnat et al., 2010; Richard et al., 2010).

In this work, we build the surrogate dataset with the same power spectrum as in the original data, where the information component encoded in time is randomized, cf. **Figure 2A**. For each time index of the signal we randomly select a novel index value, such that all the index values are chosen exactly once (permutations without repetition). Then, we rearrange the time-frequency representation of trials according to the chosen index values and compute recurrence plots of surrogates by repeating steps (ii) and (iii) shown in **Figure 1**. This procedure is repeated S times per trial. **Figure 2A** illustrates how to obtain the surrogate set from T trials. Examples of an original RP and a corresponding surrogate RP are provided in **Figure 2B**.

We compare pixel-related statistical measures between the set of the original recurrence plots from different trials and their surrogates to determine whether original RPs preserve the common underlying signal dynamics in statistically significant way. This comparison is illustrated in **Figure 2C**. In detail, we denote the set of T recurrence plots obtained from the original trial data by $\{RP_k\}_{k=1}^{T}$ and its surrogate set by $\{SRP_k\}_{k=1}^{S\cdot T}$. In our simulations, there are $T = 10$ trials in total, where the number of surrogates generated per trial is $S = 100$. The full set of surrogates counts $S \cdot T = 1000$ surrogate RPs. At first, we perform *pixel-wise* statistical analysis tests between the corresponding pixels of the original and the surrogate recurrence plots. Let $\bar{B} = \{r_{i,j}^{\{k\}}\}_{k=1}^{T}$ be the vector that consists of the set of pixels with same coordinates in the original RPs and $\hat{B} = \{r_{i,j}^{\{k\}}\}_{k=1}^{S\cdot T}$ is the corresponding vector of pixel values for surrogates. Vectors $\bar{B}$ and $\hat{B}$ consist of values from the set {0, 1}, since RP elements $r_{i,j}$ by definition take binary values, cf. Equation (1). To perform a chi-square test for categorical data, we build a two-by-two contingency table. For explanation, this tables first row takes values from the original RPs and the second row contains values from surrogate RP. The first table column marks the number of values $r_{i,j} = 1$ and the second column the number of elements $r_{i,j} = 0$. The elements of this table (two rows and two columns) have the coordinates (l, m), $l, m \in \{1, 2\}$. Then, the chi-square statistics for the pixel (i, j), $i, j \in \{1, \ldots, N\}$ is computed by

$$\chi^2(i,j) = \sum_{l\in\{1,2\}}\sum_{m\in\{1,2\}} \frac{(f_o^{(i,j)}(l,m) - f_e^{(i,j)}(l,m))^2}{f_e^{(i,j)}(l,m)}\,.$$

Here, $f_o^{(i,j)}(l, m)$ is the observed table value at the coordinate (l, m) for the pixel (i, j) and $f_e^{(i,j)}(l, m)$ is its expected frequency. The latter value is computed as $f_e^{(i,j)}(l, m) = n_r(l)n_c(m)/q$, where $n_r(l)$ is the total number of elements in the row l, $n_c(m)$ is the number of elements in the column m and q is the total number of elements in the two-by-two table. The calculated chi-square value is compared with the result in the chi-square table for predefined values of the degree of freedom $df = 1$ and the significance level $\alpha_s = 0.05$. If the calculated chi-square value is larger than the value in the table, the hypothesis that signals share the same distribution is rejected, see Yates (1934) for more details. In this

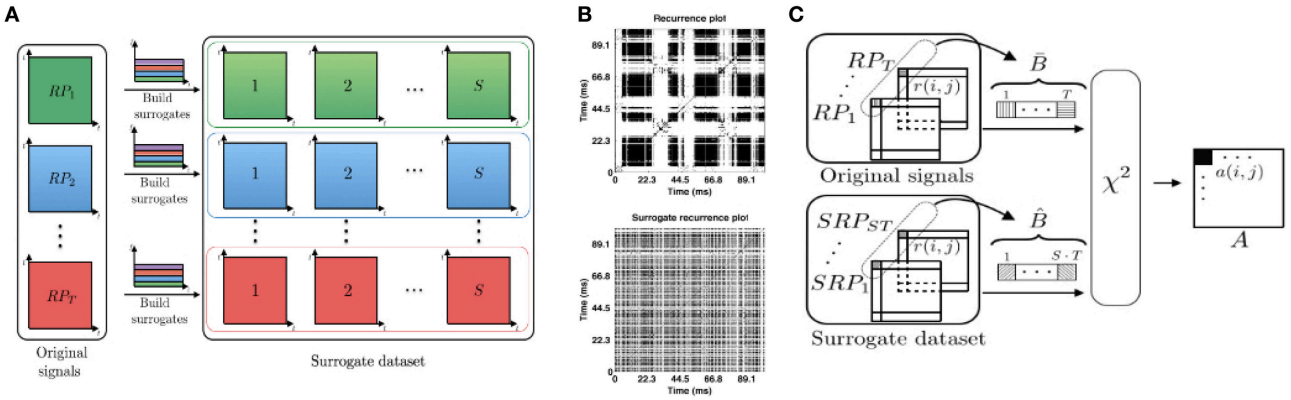


FIGURE 2 | (A) Surrogate set construction: for each of T time-frequency signal representations used for computing the original set of RPs, one computes S surrogate RPs. The resulting S surrogate recurrence plots have the same energy in time-frequency domain. In total, there are $S \cdot T$ surrogates. **(B)** Examples of the RP (top) and the surrogate RP (bottom) for the transient oscillations dataset. **(C)** Illustration of the pixel-wise χ^2 statistical test, more details in the text.

work, the outcomes of chi-square tests are visually represented as matrices $A = [a_{i,j}]$, $i, j \in \{1, \ldots, N\}$ whose elements take values

$$a_{i,j} = \begin{cases} 1, & \text{if distributions of trial and surrogate sets} \\ & \text{are different,} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

In this work all the figures follow the same color code as for illustrating recurrence plots, i.e., white pixels denote values $a_{i,j} = 0$ and black pixels stand for $a_{i,j} = 1$.

Since single elements in RPs are correlated to neighboring elements caused by the underlying dynamics, the underlying assumption of independent recurrence matrix elements does not hold and corrections of the significance test should be applied, such as the Bonferroni correction. To this end, in the examples of artificial datasets, we have performed a t-test which is based on the hypothesis that original and surrogate signals have the same distribution of mean values. The statistics are computed based on the pixels and their mean values in a 5×5 neighborhood around each pixel. In addition, we have applied a Bonferonni correction.

2.3. Datasets

To illustrate different analysis steps and to validate the power of the proposed method, we first apply the proposed algorithm to two artificial datasets. Then, the methodology is applied to experimental datasets. Single trials of these datasets are illustrated in **Figure 3** and their origin is described in detail below. For both artificial datasets, we model the tria-to-trial variability by a temporal shift of the data in time combined with additive measurement noise.

2.3.1. Transient Oscillations

A modified Lotka-Volterra model with $n = 3$ interactive elements (Rabinovich et al., 2008a,b)

$$\frac{dx_i(t)}{dt} = x_i(t)\left(\sigma_i - \sum_{j=1}^{n} \rho_{i,j} x_j(t)\right), \quad (6)$$

serves as an abstract model of event-related brain potentials (beim Graben and Hutt, 2015). Here $x_i(t) \geq 0$, $i \in \{1, 2, 3\}$ is the activity rate of the element i, σ_n is the growth rate of the n-th population and ρ_i define interactions between elements. In our setup, $\sigma_1 = 1$, $\sigma_2 = 1.2$ and $\sigma_3 = 1.6$, $\rho_{ii} = 1$, $\rho_{12} = 1.33$, $\rho_{13} = 1.125$, $\rho_{21} = 0.7$, $\rho_{23} = 1.25$, $\rho_{31} = 2.1$, and $\rho_{32} = 0.83$. The output signal $s(t)$ is a linear superposition of transient oscillations with frequencies $\nu_1 = 170$ Hz, $\nu_2 = 20$ Hz, $\nu_3 = 75$ Hz, where at one time instance, only one of these three components is dominant, see more details below. We point out that these frequencies are chosen rather arbitrarily for an optimal illustration. The activity rate x_i defines the amplitude of the component a_i with frequency ν_i and the output signal obeys

$$s(t) = \sum_{i=1}^{3} a_i(t) \sin(2\pi \nu_i t) + \xi(t) \quad , \quad a_i(t) = e^{-(x_i - \sigma_i)^2 / 2\eta_i^2}, \quad (7)$$

with $\eta_1 = 0.5$, $\eta_2 = 0.33$, $\eta_3 = 0.4$. By this construction, the amplitudes a_i increase and decrease in a certain time window outside of which they almost vanish. These windows of the three oscillation modes $i = 1, 2, 3$ do not overlap and the transitions between them are rather rapid. The variable $\xi(t)$ represents measurement noise and its random values are i.i.d. Gaussian noise with zero mean and variance 0.5. The sampling rate is 450 Hz. We generate 10 trials which are time-jittered by shifting the trials by 1 sample to later instances, while each trial is subject to additive noise different in each trial. A single trial is given in **Figure 3A**.

2.3.2. Lorenz dataset

The Lorenz system (Lorenz, 1963) is a well-studied three-dimensional differential equation system

$$\frac{dx}{dt} = -\sigma x + \sigma y\,, \quad \frac{dy}{dt} = \rho x - y - xz\,, \quad \frac{dz}{dt} = -\beta z + xy$$

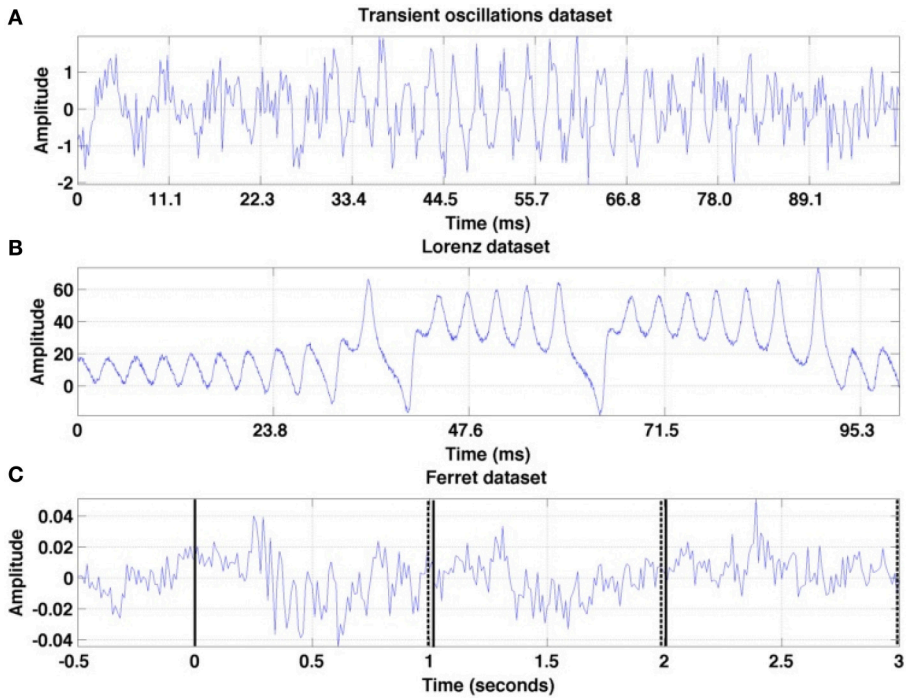


FIGURE 3 | Illustration of the first trial in each of the three time series under study. (A) Transient oscillations **(B)** Lorenz attractor **(C)** Ferret dataset: trial from the set *session one* is recorded at a granular layer electrode. The vertical solid lines denote the stimulus onsets and the set of dashed lines mark the stimulus offsets.

with $\sigma = 10$, $\rho = 28$, $\beta = 8/3$. Its solutions show non-trivial transient dynamics and their wings represent metastable states as explained above (cf. beim Graben and Hutt, 2013). We study the univariate time series $x(t)$, which is the solution of the above given system of equations. This time series may represent a macroscopic measured signal such as EEG recording (Skarda and Freeman, 1987; Basar, 2006), capturing activity from different metastable sources. The sampling rate is equal to 2100 Hz. We generate ten signal trials time-jittered by shifting the signal by 1 sample to later instances and add i.i.d. zero mean Gaussian noise with unity variance to the signals. One trial signal is illustrated in **Figure 3B**.

2.3.3. Ferret Dataset

The experimental dataset under study in the present work are Local Field Potential (LFP) measurements collected as described in Sellers et al. (2013, 2015a,b). Briefly, female ferrets were anesthetized, intubated, and underwent surgery to gain access to primary visual cortex (V1, ∼3 mm anterior to lambda and 9 mm lateral to the midline). Anesthesia induction was achieved with an intramuscular injection of ketamine (30 mg/kg) and xylazine (1–2 mg/kg), and anesthesia maintenance was achieved with 1.0% isoflurane (10–11 cc, 50 bpm, 100% medical grade oxygen), with continuous IV infusion of xylazine (1.5 mg/kg/h xylazine with 4.25 mL/h 5% dextrose lactated ringer's). Animals were head-fixed in front of the presentation screen and a 32-channel depth probe was acutely inserted into cortex (50 microns contact spacing along the z-axis, NeuroNexus, Ann Arbor, MI) and was positioned to cover all cortical layers. The reference

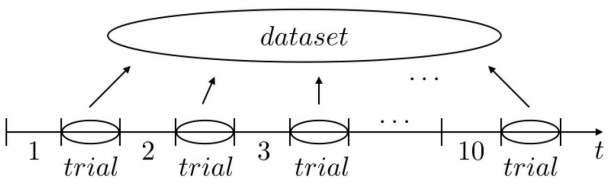


FIGURE 4 | Experimental paradigm of the ferret experiment. In the original recording protocol (Sellers et al., 2013, 2015a,b), 10 visual stimuli are interleaved by several types of different stimuli, represented by numbers in the figure. Our dataset consists of the responses to sine-wave luminance gratings only. Other types of stimuli are the black screen (intervals marked by {4, 6, 9}), checkerboard noise for {2, 7} or fox images stimuli. The latter set consists of the weakly spatially filtered image of foxes in the intervals marked by {3, 5} and strongly spatially filtered image of foxes in the interval set {1, 5, 10}.

electrode was located on the same shank (0.5 mm above the top recording site) and was positioned in 4% agar in saline above the brain. The full-field visual stimulus was presented on a 52 × 29 cm monitor with 120 Hz refresh rate and full high-definition resolution (1920 × 1080 pixels, *GD*235 Hz, Acer Inc, New Taipei City, Taiwan) at 47 cm distance from the animal. Each trial was 30 s long and consisted of three parts: (i) recording interval $[0 - 10)$s is a baseline (screen is black); (ii) $t_s \in [10 - 20)$ s is the presentation of the sine-wave luminance gratings; (iii) $[20 - 30)$ s is "post-baseline" (screen is again black). Visual stimuli were interleaved with other types of stimuli (all in randomized order), for instance with a black screen or with a strongly spatially filtered image of foxes (foxes are natural enemies of ferrets), see **Figure 4**.

The sine-wave luminance grating was presented at a rate of 1 Hz for 10 s (during each 1 s period, progressive frames transitioned from black to white to black and all the screen pixels had the same color for any given frame). In the subsequent analysis, we consider a subset of recordings. The dataset under study starts 0.5 s before stimulus onset and lasts until 3 s at the end of the third stimulus cycle. This stimulus is a black screen at $t = 0$, 2, and 3 s with luminance maxima at $t = 0.5$, 1.5 s, and 2.5 s. A single trial is illustrated in **Figure 3C**.

Electrophysiological recordings were conducted during stimulus presentation. Unfiltered signals were first amplified with MPA8I head-stages with gain 10 (Multichannel Systems, Reutlingen, Germany), then further amplified with gain 500 (Model 3500, A-M Systems, Carlsborg, WA), digitized at 20 kHz (Power 1401, Cambridge Electronic Design, Cambridge, UK), downsampled to 1 kHz afterwards, and digitally stored using Spike2 software (Cambridge Electronic Design). In total, 20 trials across two sessions conducted on different days were analyzed, the session sets of 10 trials are called *session one* and *session two* in the following. Datasets are downsampled to the sampling rate equal to 100 Hz. All procedures were approved by the University of North Carolina-Chapel Hill Institutional Animal Care and Use Committee (UNC-CH IACUC) and exceed guidelines set forth by the National Institutes of Health and U.S. Department of Agriculture.

3. RESULTS

In this section, we first apply the proposed method to two artificial datasets to verify if it reveals dynamics given the noisy set of trials. After validation of the method, we apply it on the experimental dataset (ferret dataset) and study whether it well extracts the dynamics from the recorded trials.

To understand the results, we first shortly describe recurrence plots for several simple test signals. As previously mentioned, black pixels denote recurrence events and white ones its absence. All recurrence plots have a black diagonal line, by definition (see Equation 1). Signals without any recurrence have a white square RP with a black diagonal line. Random noise signals have a random distribution of black pixels in the plot, with the exception of the black diagonal line. A simple periodic signal has a recurrence plot that consists of the black diagonal line and other black lines that are parallel to the diagonal, where the distance between them will reveal the period of the signal. More complex signals that have recurrent states may show different structures in RPs, for example, checkerboard-like patterns. These black colored fields, to which we refer as to recurrence domains, may have different sizes and shapes. For two artificial datasets we expect to observe repetitive black patterns that correspond to repetitive states within signal components. For the experimental dataset, we expect to observe recurrence patterns that are directed by the onset of the visual stimulus.

3.1. Artificial datasets

We demonstrate our methodology in **Figure 5**, which shows the analysis steps for the examples of transient oscillations (**Figures 5A–D**) and the Lorenz attractor (**Figures 5E–H**).

The time-frequency representation of one transient oscillation trial is shown in **Figure 5A**. As previously mentioned in Section 2.3.1, the corresponding signal exhibits three periodic components. We visually inspect the figure and observe high power spectrum values around the following time windows: (i) {(0, 11), (42, 54), (85, 97)} ms around $\nu_1 = 170$ Hz (dark red horizontal line segments); (ii) {(19, 26), (64, 72)} ms around $\nu_2 = 20$ Hz (broad orange areas); (iii) {(24, 38), (74, 87)} ms around $\nu_3 = 75$ Hz (dark red horizontal line segments). Note that for other trials these values may fluctuate because the frequency and time window of the current active component vary due to noise that models trial-to-trial variability. **Figure 5B** shows recurrent blocks (in black) in a single trial which correspond well to the dynamics observed in the data, cf. **Figures 5A**, **3A**. For explanation, these recurrence blocks correspond to MSs and the white parts represent transients between them.

The time-frequency representation of one Lorenz attractor trial is given in **Figure 5E**. The approximate time intervals during which the system stays in each of the two wings are visually inspected from the power spectrum values. For the wing in time intervals {(0, 30), (90, 100)} ms, **Figure 5E** shows a peak at ∼30 Hz corresponding to the oscillation frequency in the Lorenz wing, see **Figure 3B**. The other wing is reached in the time intervals {(40, 60), (65, 80)} ms in accordance to the power peak at about 40 Hz. Note that for other trials time intervals may be different due to varying trials in the set. **Figure 5F** shows recurrent blocks in a single trajectory. The recurrence blocks repeat in the correct time windows and represent the different wings, i.e., the MSs.

Time-frequency representations of single trials are the basis for the recurrence analysis leading to recurrence plots given in **Figures 5B,F** for the respective datasets. These plots show the metastable dynamics of the transient oscillations and the Lorenz trajectories in the corresponding time windows as recurrent structures. The recurrent, i.e., repetitive, structure is visible in the illustrated trial of the corresponding data. Now, considering several trials these recurrent structures may vary due to the trial-to-trial variability. Nevertheless, to study the recurrent structure *common* to all trials, we employ the statistical inference method and extract statistically significant areas of recurrence plots, as shown in **Figures 5C,D,G,H**. The recurrent structure is obvious in these plots, reflecting the underlying recurrence structure in the artificial signals. In addition, these results demonstrate that the methodology extracts recurrence structures common in several trials, although the recurrent structure is less obvious in single trials, **Figures 5B,C**. **Figures 5D,H** show the multiple comparison-test results for both artificial datasets. The white area increases and the black areas are more focussed on the red squares, i.e., spurious recurrences (black dots) are removed and and separated well from transient (white areas). Hence the multiple-comparison test improves the statistical inference.

We point out again, that the extraction of the recurrent structure from the univariate data shown in **Figure 5** is possible only by the spectral power embedding, i.e., the transformation of the univariate data into multivariate data. The subsequent preliminary statistical inference allows to identify the recurrent

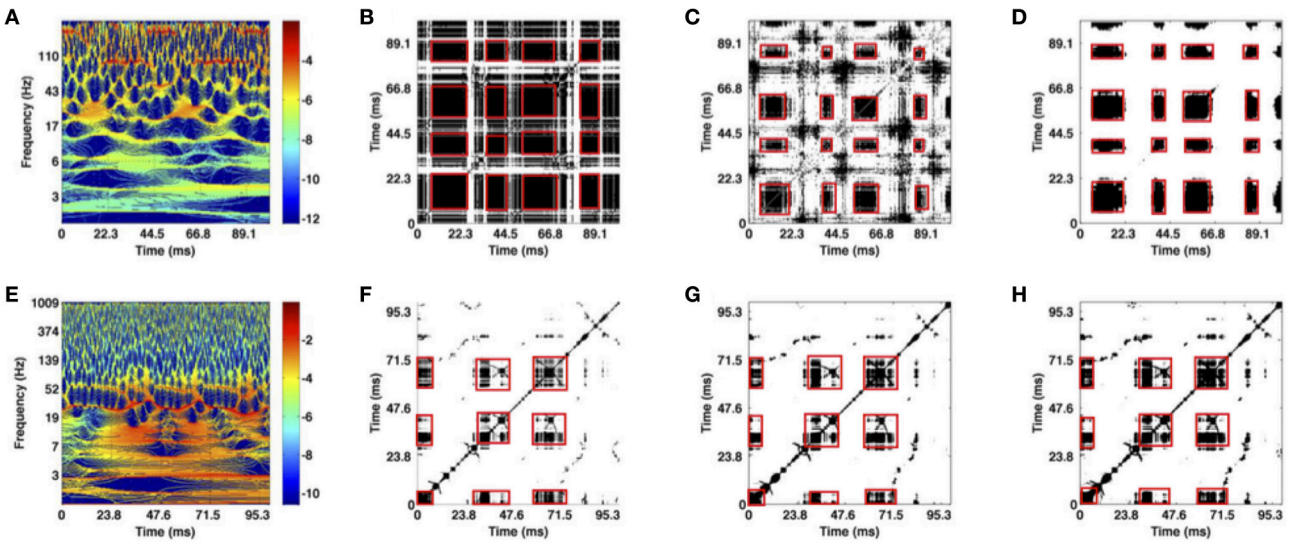


FIGURE 5 | Illustration of time-frequency (A,E), optimal recurrence plot (B,F) for the trials of two artificial datasets shown in Figures 3A,B, statistically significant recurrence plots (C,G) based on 10 trials and classical *t*-test results (D,H). Values plotted in black represent areas where the original and the surrogate set differ significantly (rejected test). **(A–D)** Transient oscillations; **(E–H)** Lorenz attractor. The red boxes mark recurrent metastable states.

MSs which are common in all trials with a confidence of 0.95. By virtue of the spectral power embedding, the method permits to select certain frequency bands to study recurrence structure in specific frequency bands. This new element renders the spectral power embedding more flexible and hence superior to previous embedding techniques, such as the delay embedding based on Takens theorem. To illustrate this, the subsequent section shows results from experimental data in different frequency bands.

3.2. Experimental Data

After studying artificially generated trials and verifying that the proposed method extracts well the dynamics features given by repetitive black structures, we now investigate whether such structures can be found in the experimental data as well.

Figure 6 provides the time-frequency representations of two single trials of the same session **Figures 6A,C** and the corresponding recurrence plots **Figures 6B,D**. We observe a high trial-to-trial variability between both trials. This can be observed both in the time-frequency representations and the resulting recurrence plots. For instance, **Figure 6B** shows a single recurrent state in the data except in the time window during the first stimulus at $t \in [0.25\,\text{s}; 1\,\text{s}]$. Hence the system remains close to the resting state ($t < 0$) during the first stimulus. Conversely, **Figure 6D** reveals that the baseline activity, i.e., activity before stimulation in the time interval $[-0.5\,\text{s}; 0\,\text{s}]$, recurres in the interval $[2.5\,\text{s}; 3\,\text{s}]$. In addition, the activity at about $t = 2$ s resembles the activity just after $t = 2.5$ s. These different findings for two trials are surprising since the experimental presentation of the visual stimulus is well-controlled and the stimulus is simple enough to expect almost identical neural responses.

To reveal the recurrent structure that is common in all trials, we now study the trial-to-trial variability of recurrence plots and aim to reveal whether the signal trials preserve the same dynamical behavior, cf. **Figure 7**. Applying the statistical method, we investigate the similarities of the results obtained from ten trials measured by a single granular sensor and from the set of 10 averaged signal trials, where the average is taken over eight granular layer sensors. This analysis is done for both ferret datasets. Moreover, we detail the analysis considering particular frequency bands which are of interest for anesthesia. To this end, we compute recurrence plots using the values of the power spectrum coefficients in the corresponding frequency bands as illustrated in **Figure 1**.

Figure 7 shows the statistically significant parts of the recurrence plots for the δ- and the α-frequency band and for all frequencies (chi-square and *t*-test results). The figure reveals that there is no statistically significant recurrent structure in the δ frequency band in signals under study. Conversely, the α-frequency band exhibits significant recurrent structures in the single granular electrode in both datasets, cf. **Figures 7A,B**. For instance, in *session one* the first response to the stimulus at $t = 0$ s returns at $t = 1$ s. Results for all frequency bands differ to results obtained in the α frequency band. The differences are dependent on the experimental sessions suggesting the presence of strong recurrences in bands different to α and δ or strong noise artifacts. To gain further insights into the dependence on frequency bands, we consider single trials which represent spatial averages of time series from adjacent granular layers. This average denoises the time series. **Figures 7C,D** shows the corresponding results. **Figures 7C,D** show results from data in both experimental sessions revealing a similar recurrence structure now. Considering all frequency bands yields recurrences similar to the one obtained in the α-band. Specifically, the prominent cross-shaped structure located at $t = 0.5$ s indicates a MS common to all data with $t \geq 0.5$ s. Additional recurrences occur in the time intervals $[-0.5; 0]$,

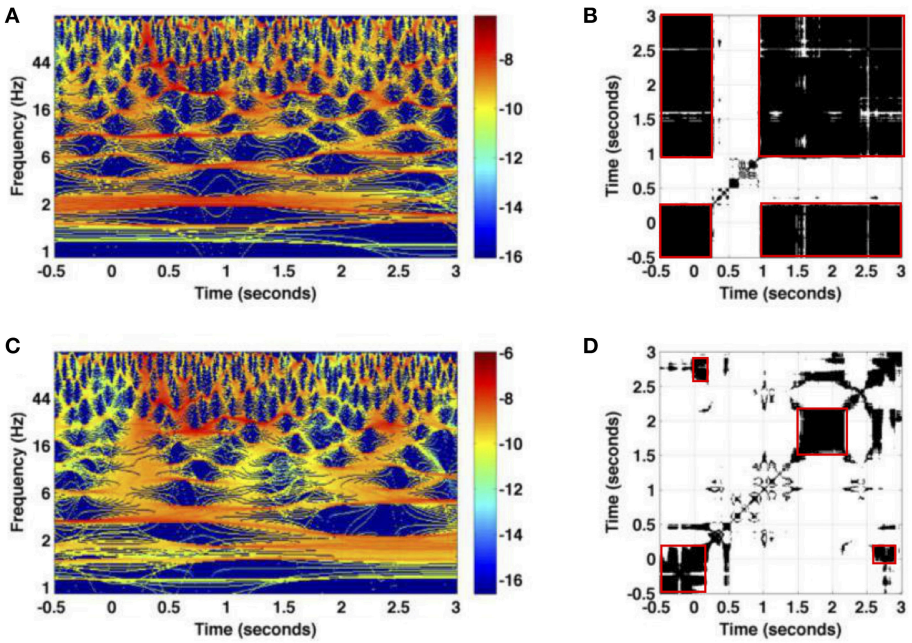


FIGURE 6 | Illustration of trial variability for the experimental dataset *session one*. (A,C) The logarithm values of the power spectrum for two trials measured at an electrode in the granular layer; **(B,D)** the corresponding recurrence plots. Non-zero values of recurrence plots are encoded in black. Red boxes denote recurrent metastable states.

[1.7; 2.2], and [2.7; 3.0] ms. These results are consistent in the chi-square significance tests and the t-test involving corrections for multiple comparison. At last, we mention the prominent lack of recurrence in the baseline time interval observed in a single electrode in *session one*, cf. **Figure 7B**. Since it does neither occur in *session one* nor in the spatially averaged data shown in **Figures 7C,D**, it appears to be spurious and is neglected.

The experimental paradigm includes visual presentations of stimulus types in a randomized order. The previous paragraphs show neurophysiological responses to the sine-wave stimulus only. To gain further insight into the trial-to-trial variability subject to various pre-stimuli, we have selected two subsets of sine-wave trials that have two different preceding stimuli, namely the "black screen" subset denoted by *subset one* and the subset of "strongly spatially filtered version of foxes" denoted by *subset two*. *Subset one* includes the trials $\{4, 6, 9\}$, while *subset two* is composed of the trials $\{1, 5, 10\}$ of datasets *session one* and *session two*. The comparison of various pre-stimuli data is done by the chi-squared difference measure based on recurrence plots of both subsets. **Figure 8** shows the statistically significant recurrences that are common in stimulus responses on both types of pre-stimuli. The diagonal lines are absent from figures, which suggests that at each time instance two comparison signals differ. Poor but visible recurrent structures in δ-band are grouped into two distinct blocks which distinguish the activity before the stimuli (around $t = 0$ s) and during the stimuli, for $t \in (1, 3)$ s. In α-band, the figure shows more prominent recurrences, such as the patterns around $t = 1.5$ and 2.7 s. We point out that recurrences within δ- and α-bands do not overlap, except in the pre-stimuli period, for $t \in \{-0.5, 0\}$ s. Finally, considering all the frequency bands together does not reveal significant similarities of two pre-stimuli. Results from the chi-square test and the t-test involving multiple comparison correction are similar. However, it is interesting to note that the t-test reveals more significant common recurrences than obtained with the naive chi-square test.

4. DISCUSSION

The present work introduces a new recurrence analysis methodology for univariate time series. The first new element is the transformation to a time-frequency representation leading to a multivariate time series of spectral power. This new technique generates a new high-dimensional phase space in which the instantaneous power of the signal evolves. This high-dimensional phase space is mandatory to apply recurrence analysis. In addition, it permits to compute recurrence plots for specific frequency bands. The second new element is the statistical analysis of recurrence plots that takes into account spurious recurrence structures and allows to suppress them. The combination of the two methods permits to extract temporal recurrence structures in data which may reflect underlying transient dynamics in a certain range of frequencies that would have been hidden in conventional methods. To our best knowledge these two techniques have not been considered before.

The first results for two artificial datasets illustrate the methodology and indicate that method detects recurrences in a variable dataset (noise-induced trial variability) by the statistical analysis as seen in **Figure 5**. These results on artificial datasets prove that the method reveals underlying recurrences in a set of trials if they are present in these trials.

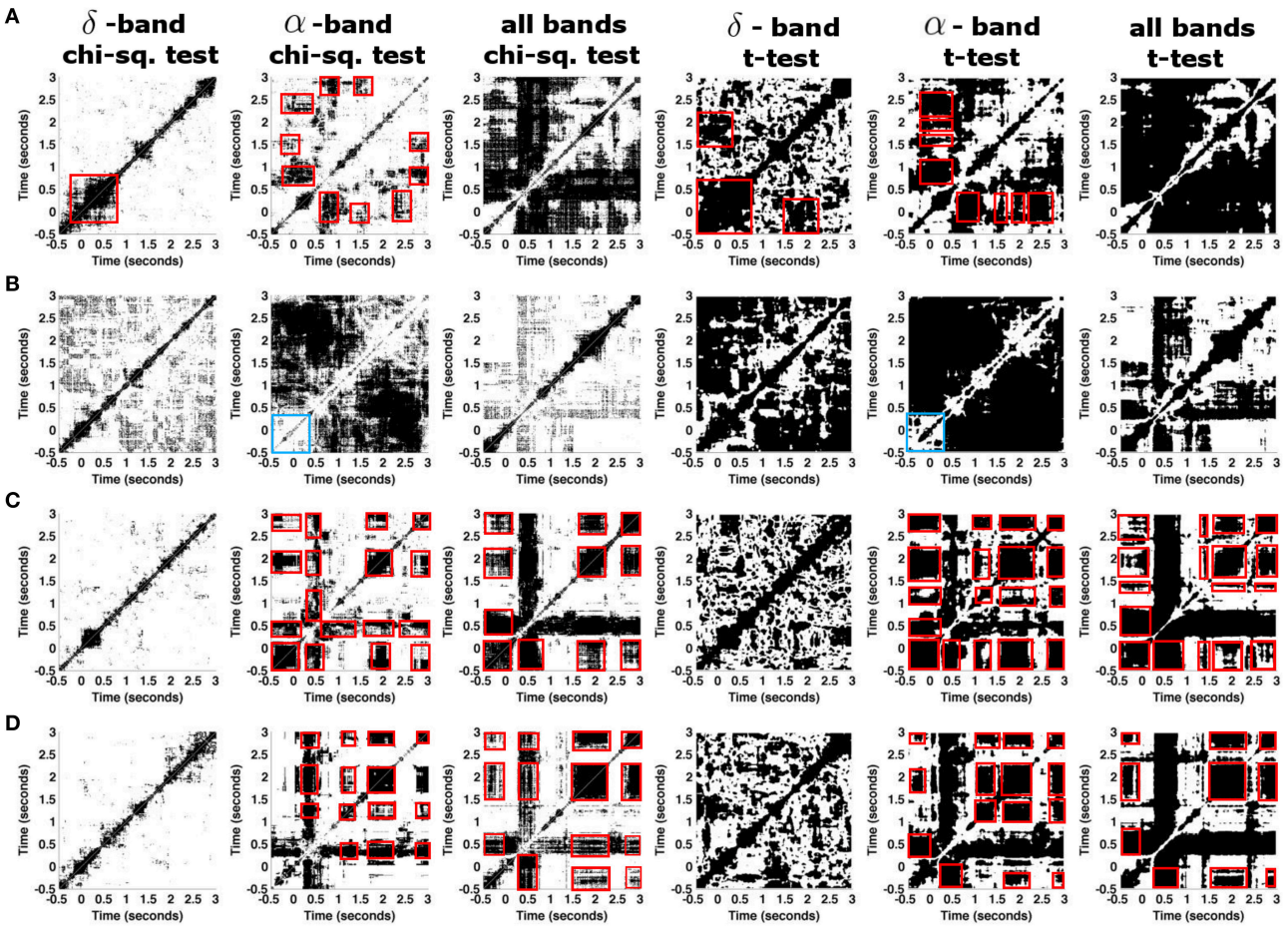


FIGURE 7 | Statistically significant parts of recurrence plots for different ferret datasets. (A) Results from a dataset of *session one* that consists of 10 trial recordings of the single granular sensor s_{10}. **(B)** Same as in **(A)** for the dataset *session two*. **(C)** Results are calculated based on 10 averaged trials in dataset *session one*, where each trial is averaged over 8 granular layer time series. **(D)** Same as in **(C)** for dataset *session two*. Red boxes indicate MSs, blue boxes in **(B)** indicate a prominent lack of recurrence in the α−band.

The subsequent analysis of single Local Field Potentials measured experimentally in ferret visual cortex reveals a high trial-to-trial variability, cf. **Figure 6**. The trial-to-trial variability is surprising due to the well-controlled experiment revealing an intrinsic ongoing activity (Arieli et al., 1995). This result demonstrates that it is mandatory to take into account recurrence variability in several trials. This is done by the methodology proposed. Detailed recurrence analysis of specific frequency bands in **Figure 7** reveals missing recurrences in the δ band whereas α-activity exhibits statistically significant temporal recurrence. This important finding reflects a fundamental difference of the nature of δ- and α-activity which has been shown in previous experimental studies on the neural origin of both signal features (Alkire et al., 2000; Ching et al., 2010; Hashemi et al., 2014). Our results suggest that the brain may decode information processing steps in different frequency bands. This might be of importance in previous studies and may shed some new light on neural processes, such as on metastable states in EEG during the emergence from unconsciousness (Hudson et al., 2014) and metastable states in bird songs (Yildiz and Kiebel, 2011).

The effects of pre-stimuli have been hypothesized (Van Rullen et al., 2011; Lundqvist et al., 2013) and we have investigated the effect of pre-stimuli. The performed analysis is based on a rather small set of trials reflecting the responses to identical stimuli. To have sufficiently large dataset for tests, we merged trials coming from two recording sessions. We note that trials coming from two sessions may not be independent, which may introduce errors. We found negligible effects in the δ frequency bands but differences in the temporal recurrence structure in the α frequency band. This result indicates that α-activity is more sensitive to pre-stimuli than δ-activity in the experimental setup under study. This finding is in full line with previous theoretical (Lundqvist et al., 2013) and experimental (Romei et al., 2008) studies on the importance of phase and power of prestimulus α-activity. In addition, we notice the absence of the diagonal line and other strong recurrence patterns visible in **Figure 7**. This may be the result of merging trials from different sessions, which was necessary to obtain larger test set for the analysis.

The present work shows that trial-to-trial variability in neurophysiological data occurs in spite of well-controlled

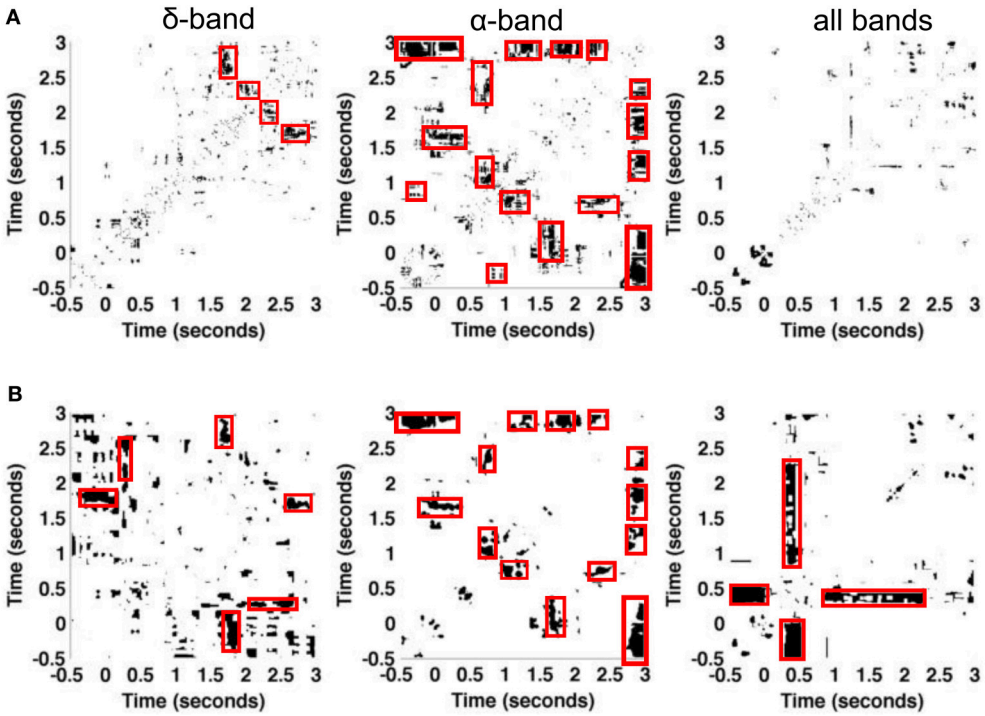


FIGURE 8 | Illustration of the influence of different visual pre-stimuli on resulting plots in experimental dataset. Statistically significant areas of recurrence plots are obtained by **(A)** pixel-wise chi-square tests and **(B)** *t*-test between trials with pre-stimulus *black screen* and those whose pre-stimulus is *fox image*. The data is taken from a single granular layer electrode in datasets *session one* and *session two* together. Both the pre-stimulus *black screen* and *fox image* have occurred three times among the 10 trials in each dataset. Significantly different values are coded as white pixels, statistically similar values are coded as black pixels.

and simple response-driven experimental conditions and demonstrate how to extract recurrent structures nevertheless. The methodology proposed makes it necessary to choose a well-adapted technique to transform the univariate times series to a multivariate time frequency signal. In addition to our current choice of a spectral reassignment technique, we have employed a conventional wavelet technique using complex Morlet mother wavelets and performed the same recurrence analysis (results are not shown). It turns out that this conventional method does not provide high-quality extraction of transient recurrent structures, given by the reassignment method. This may result from the worse time-frequency resolution of conventional Morlet wavelets. Future work will further investigate the best choice of multi-resolution time-frequency methods. Moreover, the methodology considers surrogate data generated by a temporal random shuffling of data and hence destructing all temporal structure. Future work may include the destruction of the recurrence structure by phase randomization in certain frequency bands (Li et al., 2010).

To conclude, in this work we propose a novel analysis method for trial-to-trial variability of recurrence plots in univariate time series applying a novel statistical analysis technique. This extension of recurrence analysis by a statistical technique is motivated by the fact that many physiological datasets have a limited number of trials but posses the intrinsic recurrence property of patterns of interest. Inspired by the fact that particular physiological patterns very often occur in specific frequency bands, we first build novel recurrence plots from a time-frequency signal representation. A low dimensional time-frequency signal that is built by the band median filter is then used to obtain original trial recurrence plots. Next, we use a chi-squared statistics to obtain statistically important areas of recurrence plots. The work reveals a strong trial-to-trial variability of recurrences in experimental data in spite of the well-controlled experimental paradigm. Moreover, it turns out that recurrences occur in the α-frequency band, whereas activity in the δ-frequency band does not exhibit a temporal recurrent structure indicating frequency-dependent metastable states.

5. DATA SHARING

We provide the time-series of the transient oscillation dataset and Lorenz dataset on the webpage of the corresponding author (https://sites.google.com/site/tamtos/datasets).

AUTHOR CONTRIBUTIONS

The majority of the analysis steps and the implementations have been performed by TT. KS and FF have provided the experimental data and neurophysiological insights. MF has

contributed additional time-frequency analysis results and PB and AH have conceived the study. All authors wrote the manuscript together.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute Of Mental Health of the National Institutes of Health under Award Number R01MH101547. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. AH acknowledges funding from the European Research Council for support under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 257253. PbG acknowledges support by a Heisenberg Fellowship of the German Research Foundation DFG (GR 3711/1-2).

REFERENCES

Alkire, M., Haier, R. J., and Fallon, J. H. (2000). Toward a unified theory of narcosis: brain imaging evidence for a thalamocortical switch as the neurophysiologic basis of anesthetic-induced unconsciousness. *Conscious. Cogn.* 9, 370–386. doi: 10.1006/ccog.1999.0423

Allefeld, C., Atmanspacher, H., and Wackermann, J. (2009). Mental states as macrostates emerging from EEG dynamics. *Chaos* 19, 015102. doi: 10.1063/1.3072788

Arieli, A., Shoham, D., Hildesheim, R., and Grinvald, A. (1995). Coherent spatio-temporal pattern of on-going activity revealed by real-time optical imaging coupled with single unit recording in the cat visual cortex. *J. Neurophysiol.* 73, 2072–2093. doi: 10.1109/MSP.2013.2265316

Auger, F., Flandrin, P., Lin, Y.-T., McLauhlin, S., Meignen, S., Oberlin, T., et al. (2013). Time-frequency reassignment and synchrosqueezing. *IEEE Signal Process. Mag.* 80, 32–41. doi: 10.1109/MSP.2013.2265316

Basar, E. (2006). *Brain Function and Oscillations, Chaos in Brain Function*. Heidelberg: Springer.

beim Graben, P., and Hutt, A. (2013). Detecting recurrence domains of dynamical systems by symbolic dynamics. *Phys. Rev. Lett.* 110:154101. doi: 10.1103/PhysRevLett.110.154101

beim Graben, P., and Hutt, A. (2015). Detecting event-related recurrences by symbolic analysis: applications to human language processing. *Proc. R. Soc. Lond. A* 373:20140089. doi: 10.1098/rsta.2014.0089

Borgnat, P., Flandrin, P., Honeine, P., Richard, C., and Xiao, J. (2010). Testing stationarity with surrogates : a time-frequency approach. *IEEE Trans. Signal Process.* 58, 3459–3470.

Ching, S., Cimenser, A., Purdon, P. L., Brown, E. N., and Kopell, N. J. (2010). Thalamocortical model for a propofol-induced-rhythm associated with loss of consciousness. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22665–22670. doi: 10.1073/pnas.1017069108

Deuflhard, P., and Weber, M. (2005). Robust perron cluster analysis in conformation dynamics. *Linear Algebra Appl.* 398, 161–184. doi: 10.1016/j.laa.2004.10.026

Friston, K. J. (1997). Transients, metastability, and neuronal dynamics. *Neuroimage* 5, 164–171. doi: 10.1006/nimg.1997.0259

Froyland, G. (2005). Statistically optimal almost-invariant sets. *Physica D* 200, 205–219. doi: 10.1016/j.physd.2004.11.008

Froyland, G., Padberg, K., England, M. H., and Treguier, A. M. (2007). Detection of coherent oceanic structures via transfer operators. *Phys. Rev. Lett.* 98, 131–134. doi: 10.1103/physrevlett.98.224503

Gaveau, B., and Schulman, L. S. (2006). Multiple phases in stochastic dynamics: geometry and probabilities. *Phys. Rev. E* 73:036124. doi: 10.1103/physreve.73.036124

Hashemi, M., Hutt, A., and Sleigh, J. (2014). Anesthetic action on extra-synaptic receptors: effects in neural population models of EEG activity. *J. Front. Syst. Neurosci.* 8:232. doi: 10.3389/fnsys.2014.00232

Huang, L., Wang, W., and Singare, S. (2006). "Recurrence quantification analysis of eeg predicts responses to incision during anesthesia," in *Neural Information Processing, Lecture Notes in Computer Science*, Vol. 4234 (Berlin; Heidelberg: Springer), 58–65.

Hudson, A. E., Calderon, D. P., Pfaff, D. W., and Proekt, A. (2014). Recovery of consciousness is mediated by a network of discrete metastable activity states. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9283–9288. doi: 10.1073/pnas.1408296111

Hutt, A. (2004). An analytical framework for modeling evoked and event-related potentials. *Int. J. Bifurcat. Chaos* 14, 653–666. doi: 10.1142/S0218127404009351

Hutt, A., and Riedel, H. (2003). Analysis and modeling of quasi-stationary multivariate time series and their application to middle latency auditory evoked potentials. *Physica D* 177, 203–232. doi: 10.1016/S0167-2789(02)00747-9

Iwanski, J., and Bradley, E. (1998). Recurrence plot analysis: to embed or not to embed ? *Chaos* 8, 861–871. doi: 10.1063/1.166372

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of Neural Science, 4th Edn.* New York, NY: McGraw-Hill.

Larralde, H., and Leyvraz, F. (2005). Metastability for Markov processes with detailed balance. *Phys. Rev. Lett.* 94:160201. doi: 10.1103/physrevlett.94.160201

Lehmann, D., Ozaki, H., and Pal, I. (1987). EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. *Electroencephalogr. Clin. Neurophysiol.* 67, 271–288. doi: 10.1016/0013-4694(87)90025-3

Li, C., Ding, G., Wu, G., and Poon, C. (2010). Band-phase-randomized surrogate data reveal high-frequency chaos in heart rate variability. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2010, 2806–2809. doi: 10.1109/IEMBS.2010.5626061

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.* 20, 130–141. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Lundqvist, M., Herman, P., and Lansner, A. (2013). Effect of prestimulus alpha power, phase, and synchronization on stimulus detection rates in a biophysical attractor network model. *J. Neurosci.* 33, 11917–11824. doi: 10.1523/JNEUROSCI.5155-12.2013

Mazor, O., and Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* 48, 661–673. doi: 10.1016/j.neuron.2005.09.032

McCarthy, D. E., Punjabi, N. M., Kim, P. Y., Frilot, C. II, and Marino, A. A. (2014). Recurrence analysis of the EEG during sleep accurately identifies subjects with mental health symptoms. *Psychiatry Res.* 224, 335–340. doi: 10.1016/j.pscychresns.2014.10.004

Meignen, S., Oberlin, T., and McLaughlin, S. (2012). A new algorithm for multicomponent signals analysis based on synchrosqueezing: with an application to signal sampling and denoising. *IEEE Trans. Signal Process.* 60, 5787–5798. doi: 10.1109/TSP.2012.2212891

Oullier, O., and Kelso, J. A. S. (2006). Neuroeconomics and the metastable brain. *Trends Cogn. Sci.* 10, 353–354. doi: 10.1016/j.tics.2006.06.009

Poincaré, H. (1890). Sur la probleme des trois corps et les équations de la dynamique. *Acta Math.* 13, 1–271.

Rabinovich, M. I., Huerta, R., Varona, P., and Afraimovich, V. (2008a). Transient cognitive dynamics, metastability and decision making. *PLOS Comput. Biol.* 4:e1000072. doi: 10.1371/journal.pcbi.1000072

Rabinovich, M. I., Huerta, R., and Laurent, G. (2008b). Transient dynamics for neural processing. *Science* 321, 48–50. doi: 10.1126/science.1155564

Richard, C., Ferrari, A., Amoud, H., Honeine, P., Flandrin, P., and Borgnat, P. (2010). "Statistical hypothesis testing with time-frequency surrogates to check signal stationarity," in *IEEE International Conference on Acoustics, Speech and Signal Processing* (Dallas, TX), 3666–3669.

Romei, V., Brodbeck, V., Michel, C., Amedi, A., Pascual-Leone, A., and Thut, G. (2008). Spontaneous fluctuations in posterior alpha-band eeg activity reflect variability in excitability of human visual areas. *Cereb. Cortex* 18, 2010–2018. doi: 10.1093/cercor/bhm229

Schnitzler, A., and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nat. Rev. Neurosci.* 4, 285–296. doi: 10.1038/nrn1650

Schreiber, T., and Schmitz, A. (2000). Surrogate time series. *Physica D* 142, 346–382. doi: 10.1016/S0167-2789(00)00043-9

Sellers, K., Bennett, D., Hutt, A., Williams, J., and Frohlich, F. (2015a). Awake versus anesthetized: layer-specific sensory processing in visual cortex and functional connectivity between cortical areas. *J. Neurophysiol.* 113, 3798–3815. doi: 10.1152/jn.00923.2014

Sellers, K., Bennett, D. V., and Frohlich, F. (2015b). Frequency-band signatures of visual responses to naturalistic input in ferret primary visual cortex during free viewing. *Brain Res.* 1598, 31–45. doi: 10.1016/j.brainres.2014.12.016

Sellers, K., Bennett, D. V., Hutt, A., and Frohlich, F. (2013). Anesthesia differentially modulates spontaneous network dynamics by cortical area and layer. *J. Neurophysiol.* 110, 2739–2751. doi: 10.1152/jn.00404.2013

Shalbaf, R., Behnam, H., Sleigh, J., Steyn-Ross, D., and Steyn-Ross, M. (2015). Frontal-temporal synchronization of eeg signals quantified by order patterns cross recurrence analysis during propofol anesthesia. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 468–474. doi: 10.1109/TNSRE.2014.2350537

Skarda, C., and Freeman, W. (1987). How brains make chaos in order to make sense of the world. *Behav. Brain Sci.* 10, 161. doi: 10.1017/S0140525X00047336

Takens, F. (1981). "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, *Lecture Notes in Mathematics*, Vol. 898 (Berlin; Heidelberg: Springer), 366–381.

Thakur, G. (2013). *Synchrosqueezing Transform Codes.* Available online at: https://github.com/ebrevdo/synchrosqueezing

Thakur, G., Brevdo, E., Fučkar, N.-S., and Wu, H.-T. (2013). The synchrosqueezing algorithm for time-varying spectral analysis: robustness properties and new paleoclimate applications. *Signal Process.* 93, 1079–1094. doi: 10.1016/j.sigpro.2012.11.029

Tognoli, E., and Kelso, J. (2014). The metastable brain. *Neuron* 81, 35–48. doi: 10.1016/j.neuron.2013.12.022

Tokman, C. G., Hunt, B. R., and Wright, P. (2011). Approximating invariant densities of metastablesystems. *Ergodic Theory Dyn. Syst.* 31, 1345–1361. doi: 10.1017/S0143385710000337

Van Rullen, R., Busch, N., Drewes, J., and Dubois, J. (2011). Ongoing eeg phase as a trial-by-trial predictor of perceptual and attentional variability. *Front. Psychol.* 2:60. doi: 10.3389/fpsyg.2011.00060

Wackermann, J., Lehmann, D., Michel, C. M., and Strik, W. K. (1993). Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. *Int. J. Psychophysiol.* 14, 269–283. doi: 10.1016/0167-8760(93)90041-M

Webber, C. L. Jr., and Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* 76, 965–973.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *Suppl. J. R. Stat. Soci.* 1, 217–235. doi: 10.2307/2983604

Yildiz, I. B., and Kiebel, S. J. (2011). A hierarchical neuronal model for generation and online recognition of birdsongs. *PLoS Comput. Biol.* 7:e1002303. doi: 10.1371/journal.pcbi.1002303

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.

frontiers in Systems Neuroscience

ORIGINAL RESEARCH
published: 06 March 2015
doi: 10.3389/fnsys.2015.00028

Improved discriminability of spatiotemporal neural patterns in rat motor cortical areas as directional choice learning progresses

Hongwei Mao[1], *Yuan Yuan*[1] *and Jennie Si*[1, 2, 3]*

[1] *Electrical Engineering, School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA,*
[2] *Graduate Faculty of the School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ, USA,*
[3] *Affiliate Faculty of the Interdisciplinary Graduate Program in Neuroscience, Arizona State University, Tempe, AZ, USA*

Animals learn to choose a proper action among alternatives to improve their odds of success in food foraging and other activities critical for survival. Through trial-and-error, they learn correct associations between their choices and external stimuli. While a neural network that underlies such learning process has been identified at a high level, it is still unclear how individual neurons and a neural ensemble adapt as learning progresses. In this study, we monitored the activity of single units in the rat medial and lateral agranular (AGm and AGl, respectively) areas as rats learned to make a left or right side lever press in response to a left or right side light cue. We noticed that rat movement parameters during the performance of the directional choice task quickly became stereotyped during the first 2–3 days or sessions. But learning the directional choice problem took weeks to occur. Accompanying rats' behavioral performance adaptation, we observed neural modulation by directional choice in recorded single units. Our analysis shows that ensemble mean firing rates in the cue-on period did not change significantly as learning progressed, and the ensemble mean rate difference between left and right side choices did not show a clear trend of change either. However, the spatiotemporal firing patterns of the neural ensemble exhibited improved discriminability between the two directional choices through learning. These results suggest a spatiotemporal neural coding scheme in a motor cortical neural ensemble that may be responsible for and contributing to learning the directional choice task.

Keywords: associative learning, action selection, agranular medial and lateral areas, plasticity, support vector machines

OPEN ACCESS

Edited by:
Daniel Durstewitz, Heidelberg University, Germany

Reviewed by:
Bruno B. Averbeck, National Institute of Mental Health, USA
James M. Hyman, University of British Columbia, Canada
Brice Bathellier, Centre National de la Recherche Scientifique, France

***Correspondence:**
Jennie Si, Electrical Engineering, School of Electrical, Computer and Energy Engineering, Arizona State University, 650 E Tyler Mall, Tempe, AZ 85287, USA
si@asu.edu

Received: *16 October 2014*
Accepted: *16 February 2015*
Published: *06 March 2015*

Citation:
Mao H, Yuan Y and Si J (2015) Improved discriminability of spatiotemporal neural patterns in rat motor cortical areas as directional choice learning progresses. Front. Syst. Neurosci. 9:28. doi: 10.3389/fnsys.2015.00028

Introduction

When selecting an action among alternatives in response to an external stimulus, an animal usually makes its choice according to consequences of the actions taken. Animals choose those actions that have resulted in rewards in the past and thus, learning takes place by correctly associating a stimulus with an appropriate response. A neural network that underlies the acquisition of this stimulus-response association has largely been identified (Murray et al., 2000), and it points to the prefrontal cortex (PFC) and the basal ganglia as two key nodes for solving an associative learning task (Pasupathy and Miller, 2005). Within the frontal lobe, a rostro-caudal hierarchical

organization supporting cognitive control functions such as action selection has been hypothesized (for a review see Badre, 2008). Primate studies have shown that premotor regions are also involved in learning and holding stimulus-response representations under the influence of prefrontal regions through top-down control (Koechlin et al., 2003; Boettiger and D'Esposito, 2005; Fluet et al., 2010). Additionally, the primary motor cortex has been suggested for encoding information beyond movement kinematics (Carpenter et al., 1999; Matsuzaka et al., 2007) such as features of visual stimuli that are behaviorally relevant (Zach et al., 2008; Eisenberg et al., 2011). Furthermore, the motor cortex is highly plastic for learning sensory-motor associations (Sanes and Donoghue, 2000). Putting it all together, the motor cortical regions in the frontal cortex are implicated for learning which action to select according to stimulus-response association.

Two types of adaptation could co-exist during sensorimotor association learning: motor skill learning that improves the execution of motor responses and associative learning that links sensory cues with specific response actions (Cohen and Nicolelis, 2004). Motor skill learning alone could induce neural plasticity ranging from synaptic connections (Xu et al., 2009), changing neural firing rates (Li et al., 2001; Kargo and Nitz, 2004; Rokni et al., 2007), to the motor map (Kleim et al., 1998, 2004) in both young and adult motor cortices. Therefore, motor skill learning could become a confound factor when studying sensorimotor association learning and should be treated with care. Aside from well-studied motor skill learning, whether and how motor cortical activity would adapt during associative learning is still unclear and requires further investigation.

In a previous study (Cohen and Nicolelis, 2004), rats learned to associate directional movements in response to either a high or a low tone. Significant neuronal firing rate changes in the primary motor cortex were observed on the first day when an animal's movement skill improved, but not in the following 2 days when movement parameters were stable and associative learning dominated. In some other studies where animals learned to respond to external sensory cues with appropriate actions, learning-related neural dynamics were evident in motor cortical neural ensemble activity patterns (Laubach et al., 2000; Huber et al., 2012). Based on these results we hypothesize that neural adaptation induced by learning sensorimotor associations would be reflected in changes in spatiotemporal neural firing patterns in motor cortical areas. To test this hypothesis, we had rats learn to perform a directional choice task. The goal of the task was to make a left or right side lever press in response to a left or right side light cue, respectively. Single units were recorded from rat's medial agranular (AGm) and lateral agranular (AGl) areas. Spatiotemporal neural firing patterns were investigated using support vector machines. Improved discriminability in neural patterns was observed as learning progressed.

Materials and Methods

Animal Handling and Surgery

All procedures were in accordance with guidelines of the National Institutes of Health and approved by the institutional Animal Care and Use Committee at Arizona State University. Rats (Long-Evans, male) arrived at the age of about 2 weeks weighing around 50 grams and were handled daily by experimenters to get accustomed to the environment. They started pre-training after reaching 200 grams to master the motor skill of lever pressing, which only involved pressing a single lever (no choice) in response to a light cue above the lever. The pre-training apparatus was similar to that used for recording to help familiarize rats with the recording environment. After achieving a behavioral accuracy of 90% or above for at least 3 consecutive days on the pre-training task, and once their weight reached 400 grams, rats were implanted with a chronic electrode array.

For electrode implant surgery, rats were anesthetized by an intramuscular injection of KXA (10 mg/ml ketamine, 2 mg/ml xylazine, and 0.1 mg/ml acepromazine; 0.1 ml/100g), shaved in the incision area, and placed in a stereotaxic frame. A heated water blanket was used to maintain rat's body temperature at around 35°C. Rat's heart rate and oxygen level were monitored throughout surgery with a pulse oximeter. KXA updates (0.05 ml/100g) were administrated approximately every hour during surgery after the initial shot. Craniotomy was performed over the AGm and AGl areas of the left hemisphere of the rat brain. A microwire array was centered at 2 mm lateral and 3 mm rostral from the bregma (**Figure 1C**), and lowered about 1.8–2.3 mm underneath dura, aiming for layer V pyramidal neurons. An acrylic head cap was formed to support the electrode array. The head cap was fixed to the skull with three screws. A subcutaneous injection of 0.1 ml meloxicam was given for pain relief after surgery, and three more shots were given for the following 3 consecutive days. The rats had 7–10 days or as needed to recover before they were food restricted for recording sessions.

Behavioral Task

Rats were freely moving in the recording chamber, and self-paced to start a trial by pressing the retractable center ready lever. One of the five cue lights (from left to right: LL, L, C, R, and RR) would appear (**Figure 1A**) immediately upon ready LP. The left and right response levers would extend 2 s after cue light onset (**Figure 1B**). Pressing the left lever once would "move" the light one position to the right and pressing the right lever would "move" the light to the left. Once the light cue reached and then remained at the center position for at least 1 s, the trial ended as a success. Otherwise a trial was considered a failure if the light cue ended up at any position other than the center. A feedback tone was played immediately upon the end of a trial: a low frequency tone of 1 kHz in case of a success and a high frequency tone of 12 kHz in case of a failure. A sugar pellet reward was delivered 0.5 s after the feedback tone for a successful trial. The inter-trial interval was 8 s for successful trials and 15 s for failed trials. The five cues were presented in a pseudo-random fashion with equal probability of presence.

Recording Sessions

After rats recovered from surgery, daily recording sessions began, each of which lasted about 60 min. Rats were food restricted during the recording period while the body mass was closely monitored.

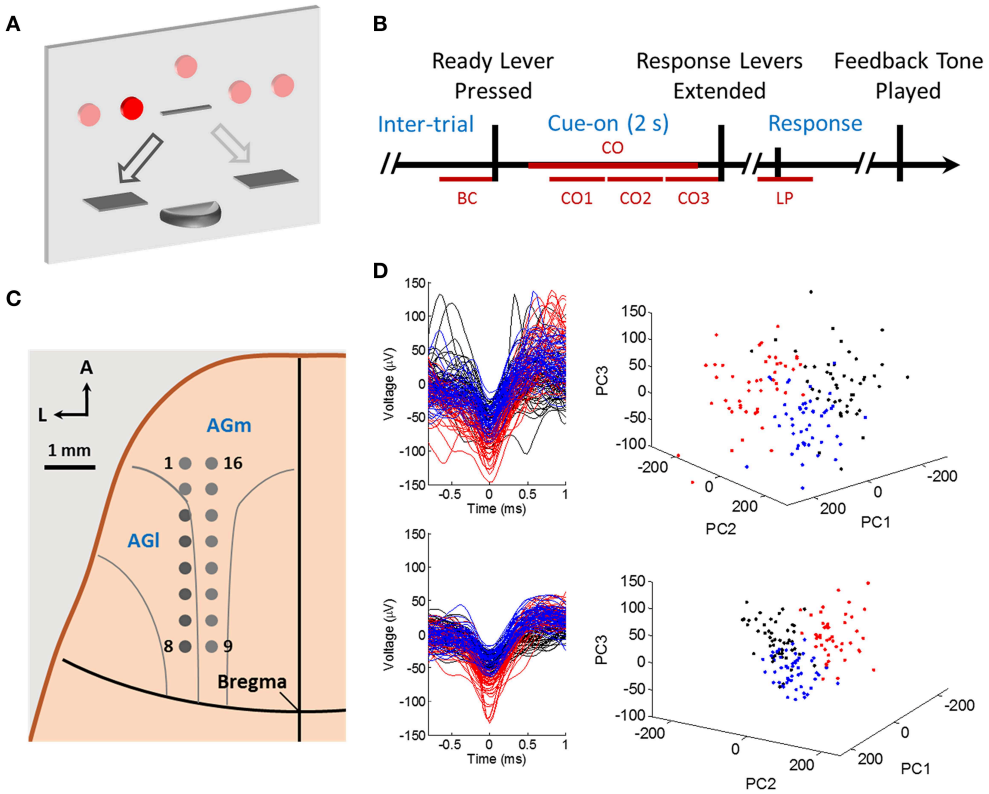


FIGURE 1 | The behavioral task, trial timeline, and recording sites. (A) The control panel setup. The rat was cued by an LED light in one of five positions in any given trial, and could use the left and right response lever to "move" light to the right and left by one position, respectively. The goal was to reach the center light by appropriate lever presses in order to receive a sugar pellet reward. **(B)** The task timeline. The rat would self-start a trial by pressing the center ready lever at his own will, and simultaneously a light cue would turn on. Then there would be a 2 s cue-on period, after which the two response levers would extend simultaneously. The rat would choose and press one of the response levers. A feedback tone would be played at trial end indicating the outcome. **(C)** A 2 × 8 microwire array was chronically implanted in the left hemisphere of the rat brain, aiming for layer V AGm and AGl neurons. **(D)** Example single unit recordings from rat A09 (top row) and rat K11 (bottom row). Left: unsorted waveforms (black), spike waveforms used in neural analysis (blue) and waveforms of another unit not used due to very low firing rate (red). The 50 waveforms from each class were randomly selected and plotted. Right: 3-D PCA projection of the waveforms.

The implanted electrodes were arranged in a 2 × 8 matrix, with 500 or 375 μm row separation, and 500 μm electrode spacing. The polyimide-isolated tungsten microwires were 50 μm in diameter and 5 mm in length. The electrode tips were cut at a sharp 60° angle (TDT Inc., FL). A total of 16 channels of raw waveforms were recorded simultaneously using a RX5 Pentusa Base Station or a RX7 Microstimulator Base Station (TDT Inc., FL). Neural signals picked up by electrodes were passed to a unity gain preamplifier (bandpass 2.2 Hz ~ 7.5 kHz) through an Omnetics or a ZIF-Clip headstage, and then sampled and stored at 24.414 kHz by the base station.

Rat behavior while performing the task was monitored and recorded using cameras (25 fps). Rat head position was determined offline by the implanted head cap as well as left and right ear positions. The head position was tracked and extracted to indicate rat movement trajectory, which was calculated for left side and right side movements separately over recorded trials. Variation in movement trajectory was obtained as the distance between a given movement trajectory during a trial and the mean trajectory in a single recording session.

Spike Sorting

Action potentials were detected and classified off-line using our own M-Sorter software (Yuan et al., 2012), which is based on the multiscale correlation of wavelet coefficients (MCWC) detection algorithm (Yang et al., 2011). The M-Sorter has been tested and compared with two popular sorters: the Wave Clus and the automatic mode of Offline Sorter by Plexon (T-Distribution EM method). The M-Sorter consistently outperformed or was at least comparable to the compared sorters (Yuan et al., 2012). One isolated unit with highest firing rate was extracted from each of the electrodes. Experimenters also inspected spike waveforms, inter spike intervals, and other measures to ensure the quality of single unit clusters (**Figure 1D**). According to the sites of implanted electrodes, recorded neurons were in the AGm and AGl areas of the rat frontal cortex (Paxinos and Watson, 2005) involving

forelimb, neck, and vibrissae areas (Neafsey et al., 1986; Remple et al., 2001). Intracortical microstimulation was also performed to confirm implant electrode location.

For each rat, only those electrodes that consistently picked up unit action potentials in all sessions were included in neural activity analysis. By doing so, we were able to analyze neural ensembles of the same size over sessions to make the results comparable. The analyses in this study as described below were based on neural ensembles. Therefore, we did not require tracking same neurons over learning sessions (Laubach et al., 2000; Cohen and Nicolelis, 2004).

Firing Rate Modulation

In this study, L-L trials are used to denote those trials in which rats reported left side choices by pressing the lever on the left side in response to left side cues, and similarly we define R-R trials. Single unit firing rates in single trials were calculated using a 100 ms data window sliding at 20 ms steps (50 bin/s) through the cue-on task period (**Figure 1B**). The mean firing rate for a data window (CO1, CO2, CO3, or CO) was the average of all binned firing rates in the respective data window.

Let the ith neuron's mean firing rate in session k for all L-L and R-R trials be denoted as $M_L^i(k)$ and $M_R^i(k)$, respectively. The ith neuron's mean firing rate was then defined as $M^i(k) = \frac{1}{2}(M_L^i(k) + M_R^i(k))$. The ith neuron's firing rate difference between L-L and R-R trials in session k was calculated as $D^i(k) = M_L^i(k) - M_R^i(k)$.

The ensemble mean firing rate $M(k)$ of N isolated units in session k was the average over all recorded trials of isolated units, i.e., $M(k) = \frac{1}{N}\sum_{i=1}^{N} M^i(k)$.

The ensemble mean firing rate difference between L-L and R-R trials was then the average of the absolute value of single unit rate differences, i.e., $D(k) = \frac{1}{N}\sum_{i=1}^{N} |D^i(k)|$. As such, each recording session corresponded with one measurement for the ensemble mean rate and another measurement for the ensemble mean rate difference.

In addition to firing rates, firing variability was also monitored. First, we calculated the standard deviation of firing rate of unit i in session k, $S_L^i(k)$ and $S_R^i(k)$, for L-L and R-R trials, respectively. Then, the mean standard deviations, $S_L(k)$ and $S_R(k)$, were calculated as the average across units, respectively.

To study how the ensemble mean firing rate and ensemble mean rate difference would change during learning from session to session, linear regressions were performed against normalized session numbers (between 0 and 1). The sign of the regression line slope was determined according to its confidence interval. A positive slope corresponded with increased rate measures while a negative slope with decreased rate measures. No change in rate measures was associated with a regression line slope that was not significantly different from zero. Similar linear regression analysis was used to examine changes of other measurements as described below.

In order to summarize results of multiple rats, the ensemble mean firing rates of single sessions were Z-scored (zero mean and standard deviation equal to one) over sessions for each rat. Then Z-scored ensemble mean rates from all rats were pooled together for linear regression analysis. Other measurements, including ensemble mean rate difference, mean standard deviation of firing rate, and SVM classification results as described below, were processed in a similar manner when their trends over sessions were explored by summarizing multiple rats' data.

SVM Classification of Neural Representations

We modeled neural firing patterns of L-L and R-R trials by training linear kernel support vector machines (SVMs). The input to the SVMs was spatiotemporal neural firing activity in the cue-on task period of a single trial while the output of the SVMs was the directional choice of left or right. All analyses were performed using customized Matlab programs (Mathworks Inc., MA).

SVMs solve a binary classification problem by determining a separating hyperplane with a maximized margin between two classes (Burges, 1998). Once the separating hyperplane is found, an SVM makes a classification decision for a given data sample x according to the value of the decision function: $df(\boldsymbol{x}) = \sum_i \alpha_i K(\boldsymbol{s_i}, \boldsymbol{x}) + b$, where support vector $\boldsymbol{s_i}$, weight α_i and bias b are determined in the training process automatically once input and output data are presented for training, and the kernel function K is a dot product in case of a linear kernel. If $df(\boldsymbol{x}) \geq 0$, $\boldsymbol{x}$ is classified as an L-L trial, otherwise it is classified as an R-R trial. The decision function value could be interpreted as the distance from the sample point to the separating hyperplane. The greater this distance the less ambiguous the final classification.

In our analysis, a 1500 ms data window in the cue-on period (CO: 300 to 1800 ms after cue onset, **Figure 1B**) was used. This window was divided into three non-overlapping 500 ms time bins. Spike counts in these bins formed one vector representation for each spike train of each unit. Spike count vectors of simultaneously recorded units were then concatenated to form a spike count vector representation of the recorded neural ensemble (**Figure 2**). Thus, there was one ensemble vector or one data sample for each trial, and SVMs were trained based on data samples from both classes (L-L and R-R trials) in each recording session for each rat.

To obtain statistically representative results, a total of 100 SVM classifiers were trained and tested for each session. In each of the 100 classifiers, a constant number of trials were randomly chosen from both L-L and R-R classes. Specifically, 20% of the randomly chosen trials from each class formed the test set, and the remaining 80% formed the raw training set, which was further

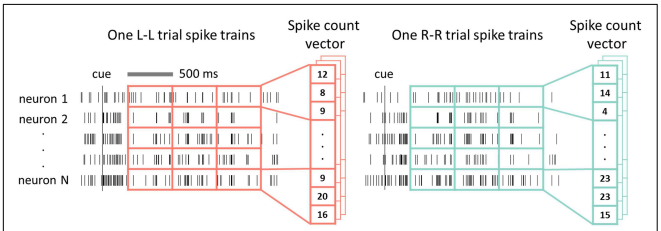


FIGURE 2 | Data preparation for SVM decoding. Spike trains of all simultaneously recorded single neurons from one task trial formed one data sample where the spike counts in non-overlap bins (e.g., 500 ms bins) were concatenated to form a spike count vector. Two classes of data samples from L-L and R-R trials composed the data set.

processed as follows before training an SVM. First, five trials of the same class were randomly selected from the raw training set, each with a respective spike count vector. Then an average spike count vector was obtained based on the five spike count vectors, and was used as an input training data sample for the SVM model. A test data sample simply was a single spike count vector of a test trial. The SVM performance measure was based on averaged test set classification accuracy from the 100 SVM classifiers. This procedure was repeated for each recording session over the entire directional choice learning process. SVM based classification performance of L-L vs. R-R trials over multiple sessions were then inspected using linear regression.

As a control, SVM classification analysis was also performed using a 500 ms time window around response lever press (LP, −100 to 400 ms around the press, **Figure 1B**). Data was prepared in a similar way as described above and 100 ms time bins were used to compute spike counts. To make classification performance comparable, three 500 ms windows within the cue-on period were selected (CO1, 500 to 1000 ms after cue onset; CO2, 1000 to 1500 ms; CO3, 1500 to 2000 ms; **Figure 1B**). SVM classification analysis was repeated in these data windows and compared with that using data in LP.

Note that in our analysis, 100 ms bin size was used for 500 ms data window (CO1, CO2, CO3, and LP) based direction predictions by SVM. But for SVM classification analyses where the CO window (**Figure 2**) was involved, 500 ms bins were used to form spike count vectors.

Results

Behavioral Results

Male Long-Evans rats ($n = 9$) started learning the directional choice task by trial and error from a naïve state. Behavioral accuracy in each recording session was monitored and calculated as the number of correct trials over the total number of trials in that session. Rats gradually improved the accuracy over sessions, from 30.8% (average, range from 14.1 to 47.3%) in session 1 to 76.0% (average, range from 55.3 to 93.4%) in session 18 (rank-sum test, $p < 0.001$; **Figure 3A**). Linear regressions of behavioral accuracies vs. session numbers revealed that seven rats significantly improved their performance except rat A09 and I10. Actually rat A09 didn't learn the right side choices, and rat I10 struggled with both left and right side choices. Among the seven rats, one of them (J11) reached 75% accuracy, all the other six rats went above 80%, and two rats (W09 and O10) even achieved over 90% accuracy. We therefore used data from the seven rats when reporting results against normalized session numbers as learning progressed. When results are based on data from all nine rats, it will be specified accordingly.

Once the rat self-started a new trial (**Figure 1B**), he had 2 s to choose from the two response levers prior to their extension. Upon response lever extension, he could make a press of his choice within 1 s. In this analysis, the response latency was calculated as the time from response lever extension to the first press on the chosen lever. This latency decreased from 0.44 ± 0.16 s in

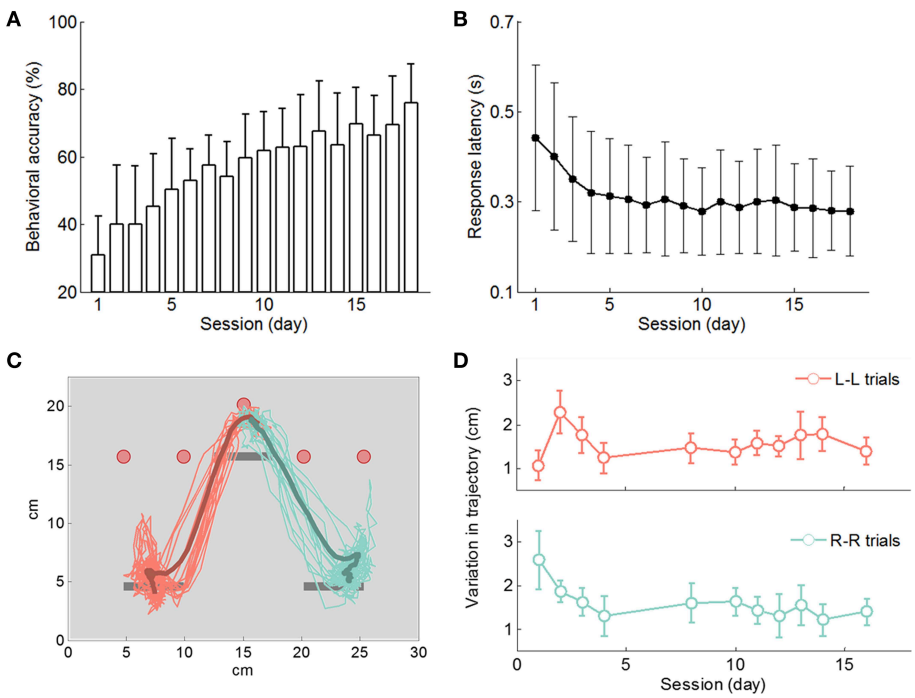


FIGURE 3 | Behavioral results. (A) Rats ($n = 9$) improved behavioral accuracy to a level above 70% in 18 days on average. **(B)** The response latency decreased mainly during the initial days and stayed relatively stable afterwards. **(C)** Movement trajectory (head position) during the cue onset period projected onto the front view of the interface panel. Thin traces: trajectory records from single trials. Bold curves: mean trajectory traces of left side and right side movements. **(D)** Variation in movement trajectory (distance between the trajectory trace of a single trial and the mean trajectory) decreased during initial learning sessions.

session 1 to 0.32 ± 0.14 s in session 4 (mean ± STD; ANOVA, $p < 10^{-5}$; **Figure 3B**; nine rats). After the first three learning sessions the response latency became stabilized, and from the fourth session onwards, the latency measurements over sessions were not significantly different (linear regression slope at −0.0021 s/session, 99% confidence interval of [−0.0058 0.0016]). The reduction in response latency is an indicator of improved act of lever pressing.

During task performance, rats usually made quick movements toward their chosen lever right after light cue onset and their movements became stereotypical in a few sessions. Video analysis of rat movement trajectory during the cue-on period (**Figure 3C**) confirmed this observation. Typically, rats started moving in their chosen direction shortly after cue onset. Their directional movements were completed by about 1 s after cue onset and then rats stayed in front of the response lever waiting for lever extension. To measure differences in rat directional movement from trial to trial, variation in movement trajectory between a single-trial trajectory and the mean trajectory of a session was calculated for each session. As shown in **Figure 3D**, variations in movement trajectory decreased during the first three sessions and remained stable afterwards. Additional video analysis results of rats' directional movement trajectory are available in Yuan et al. (2014). These observations together with response latency results show that motor skill learning occurred during the first few days and therefore, it could be dissociated from the rest of the associative learning process.

Firing Rate Modulation by Directional Choice

In this study, trials with correct first response lever press were used in analysis and within those trials, we mainly focused on the cue-on task period (**Figure 1B**). A total of 220 sessions were recorded from nine rats. We included 190 sessions for analysis and excluded the remaining 30 sessions because of inadequate numbers of trials (less than 20 L-L or 20 R-R trials). Those sessions mainly included the first few sessions of each rat when behavioral accuracy was low and motor skill learning was possibly present. All together for this study, we had 11,060 L-L trials and 10,717 R-R trials from the 190 sessions with 58 L-L trials and 56 R-R trials per session on average. Of the 190 recorded sessions from nine rats, we collected 839 unit records (337 from AGl and 502 from AGm), 4.4 unit records per session on average, ranging from 3 to 6 unit records in the ensemble. Here we consider an isolated unit each day a unit record.

Single unit firing activities of L-L and R-R trials in each session were first inspected by spike rasters and peri-event time histograms (PETHs). Examples of single unit firing rate modulations by L-L and R-R trials are shown in **Figure 4**.

In the 500 ms time window before cue onset, the averaged (over all single units, nine rats, and sessions) single unit firing rate difference between L-L and R-R trials was −0.02 Hz, which was not significantly different from zero (one-sample *t*-test, $p > 0.89$). If the same single unit firing rate differences between L-L and R-R trials were evaluated in three cue-on sub-windows (CO1, CO2, and CO3 in **Figure 1B**), they were 2.76, 3.12, and 2.17 Hz, respectively, all of which were significantly greater than zero (one-sample *t*-test, $p < 10^{-5}$; **Figure 5A**).

We then evaluated time-resolved (100 ms bins for every 20 ms) single unit firing rates of L-L and R-R trials, and the firing rate difference between the two during the cue-on period (**Figure 5B**). For the pool of single units, the averaged (over all single units, nine rats, and sessions) time-resolved firing rate difference did not emerge from 0 until 400 ms after cue onset (one-sample *t*-test, $p < 0.001$), and it sustained through the rest of the cue-on window. These results show that firing rate modulation of single neurons was prominent in motor cortical areas during the cue-on period.

To study how firing rate modulation at a population level varied as learning progressed, we calculated the ensemble mean rate and ensemble mean rate difference between L-L and R-R trials (see Materials and Methods) session by session. The results from using rat B11's data are given in **Figure 5C** as an example where the firing rates of a 1500 ms cue-on window (CO, **Figure 1B**) were used. To summarize results from all seven rats, we normalized session numbers. According to **Figure 5D**, the Z-scored ensemble mean firing rate did not change significantly through the learning process (99% confidence interval of linear regression slope: [–0.48 1.05]). When the learning process was divided into three stages of equal numbers of sessions, the Z-scored ensemble mean rates were −0.21 ± 1.08, 0.16 ± 0.85, and 0.00 ± 1.00 Hz (mean ± STD), which were not significantly different (ANOVA, $p > 0.1$). The Z-scored ensemble mean rate difference between L-L and R-R trials tended to increase slightly with a slope of 0.59 (**Figure 5D**), but its 99% confidence interval was [−0.16 1.34] indicating it was not significantly different from zero. When calculated in the three learning stages, the Z-scored ensemble mean rate differences were −0.16 ± 1.09, −0.19 ± 0.83, and 0.32 ± 0.96 Hz, respectively, showing higher rate differences between L-L and R-R trials in the last stage compared with the previous two stages (ANOVA, $p < 0.05$). But **Figure 5D** also shows that some early session had large rate differences. When linear regression was performed for individual rats, the slope was again not significantly different from zero (*t*-test, $p > 0.15$). To summarize, although ensemble mean rate difference tended to become larger near the end of the recorded learning process, the trend was not strongly observed.

The BC data window (**Figure 1B**) was analyzed in a similar way to provide a control. The ensemble mean firing rate leveled over sessions (slope of Z-scored rates: 0.04, 99% confidence interval of slope: [−0.73 0.81]; Z-scored rates in the three stages were: −0.16 ± 1.06, 0.21 ± 0.77, −0.09 ± 1.08, ANOVA $p > 0.1$). The ensemble mean rate difference was relatively stable over sessions as well (slope of Z-scored rate difference: −0.40, 99% interval: [−1.17 0.36]; Z-scored rate difference in the three stages were: 0.17 ± 1.01, −0.04 ± 0.92, and −0.09 ± 1.01, ANOVA $p > 0.4$).

Additionally, the standard deviation of firing rate during the CO data window remained relatively stable across learning sessions. The linear regression slope of mean standard deviation for L-L trials against normalized session number was 0.06, which was not significantly different from zero given that the 99% confidence interval of the slope was [−0.70 0.82] (**Figure 5D**, red dotted line). Similarly for R-R trials, the linear regression slope was 0.04 and its 99% confidence interval was [−0.72 0.80] (**Figure 5D**, green dotted line).

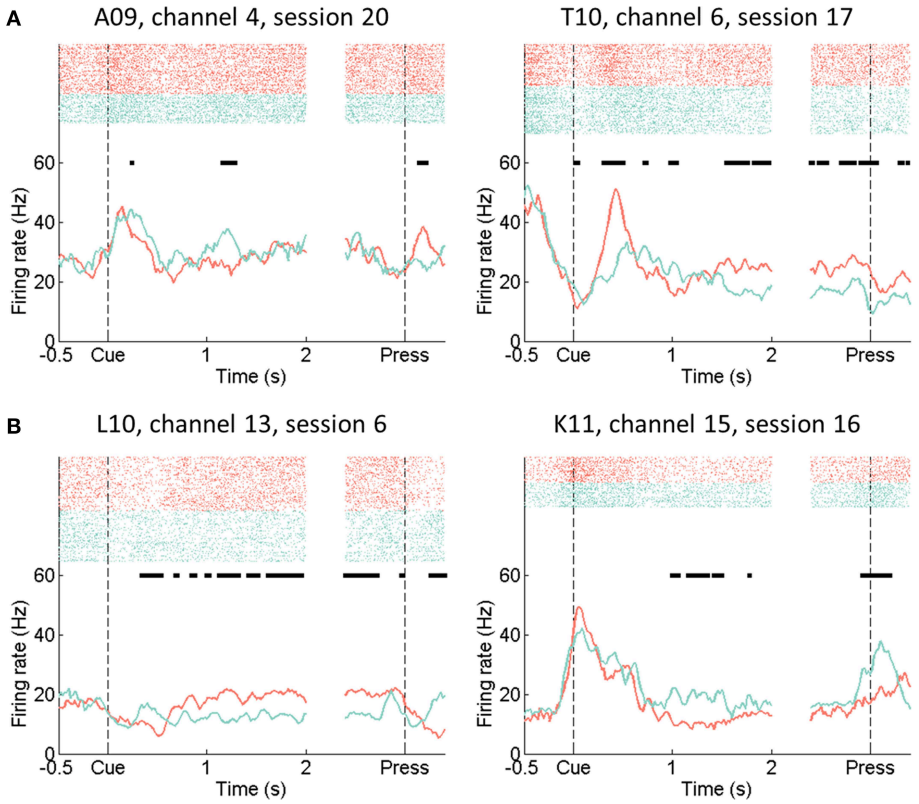


FIGURE 4 | Raster plots and peri-event time histograms of example single units. Firing rate estimated in 100 ms time bins sliding down the time axis with a step size of 20 ms. Red: L-L trials in which the rat moved to the left response lever in response to left side cues. Green: R-R trials with right side movements in response to right side cues. **(A)** Example AGl units. **(B)** Example AGm units. Horizontal bars (black) indicate time bins in which directional firing rate modulation was significant (rank-sum test, $p < 0.01$).

To summarize, the ensemble mean firing rate of all simultaneously recorded single units over all trials did not change significantly over sessions, and the ensemble mean firing rate difference between L-L and R-R trials did not show a clear trend of change either. These findings suggest that ensemble mean rate based measures are not adequate to explain improved behavioral learning of the seven rats under study.

Spatiotemporal Firing Pattern Analysis by SVMs

As learning related behavioral adaptation could not be well-explained by ensemble mean firing rate or ensemble rate difference between L-L and R-R trials in cue-on period, we then inspected the neural data with increased spatiotemporal resolution using SVMs. In the following, we first examined how data preparation and SVM parameters may affect SVM model performance.

Ensemble vs. Single Units

Ensemble spike count vectors were formed by concatenating spike counts (CO window, 500 ms bins) of simultaneously recorded single units. For each session, SVMs were trained and tested using ensemble vectors [vector dimension was 3× (number of single neurons)], and classification performance was characterized by classification accuracy on the test data. On the other hand, SVMs were built and tested using spike count vectors of single units (three dimensional vectors). When comparing the best classification performance using single unit data with that using ensemble data (**Figure 6A**), the ensemble approach outperformed the best single unit approach in 61.6% (117/190) of tested sessions. The mean single trial decoding accuracy among all sessions when using ensemble approach was 76.2%, which was higher than the 74.2% accuracy of the best single units (paired-sample t-test, $p < 10^{-5}$).

Additionally, we examined the impact of the ensemble size on decoding accuracy. Each of the nine rats had at least 3 units per session. Specifically, one rat had 3 units per session, four rats had 4 units per session, three rats had 5 units per session, and one rat had 6 units per session. As shown in **Figure 6A** insert, decoding accuracy increased when larger ensembles were used, but the speed of increase in decoding accuracy by using larger ensembles slowed down as ensemble size increased.

Multiple vs. Single Time Bins

To explore SVM classification performance over time, the cue-on period was divided into non-overlapping 100 ms time bins and SVMs were trained using data of spike counts in a single time bin from all simultaneously recorded units. As shown in **Figure 6B**, the classification accuracy (averaged over all sessions)

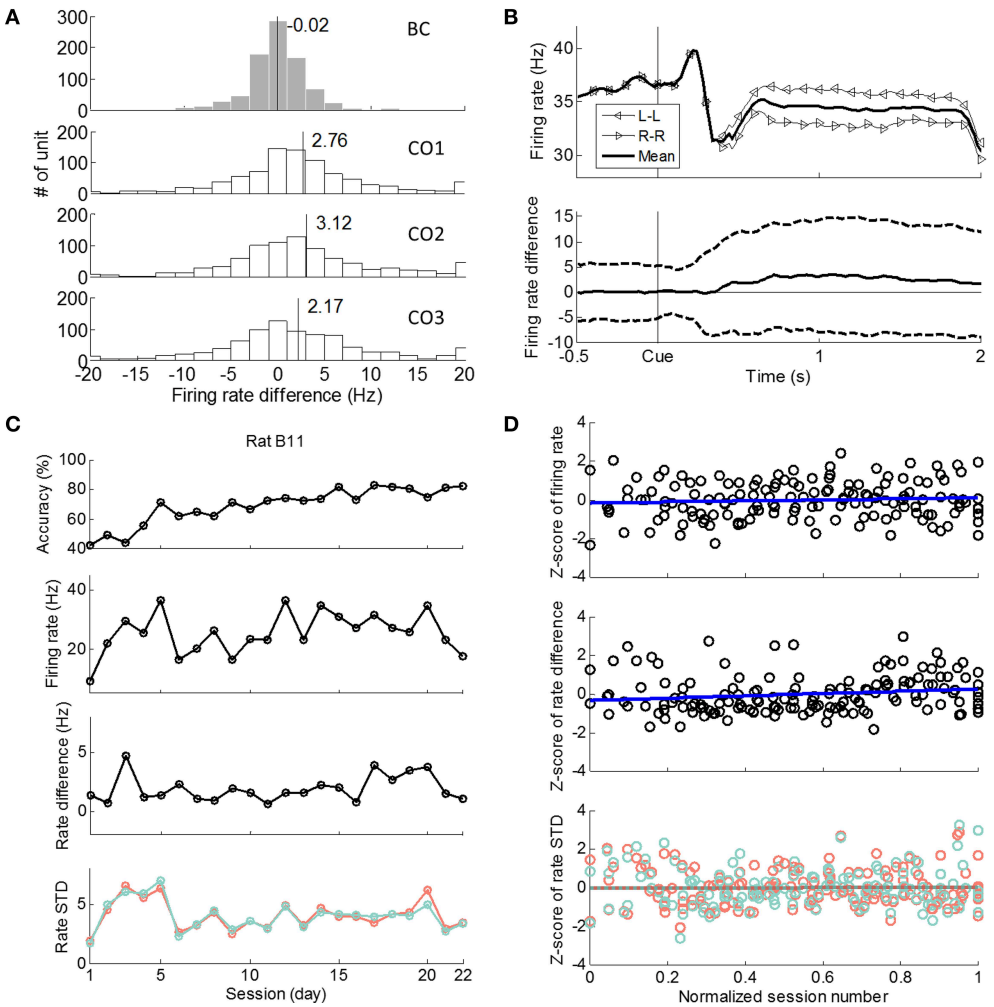


FIGURE 5 | Firing rates and firing rate differences between L-L and R-R trials. (A) Histograms of single unit ($n = 839$) firing rate differences between L-L and R-R trials for nine rats in four 500 ms time windows. BC: 500 ms window before cue onset. CO1: 500 to 1000 ms after cue. CO2: 1000 to 1500 ms after cue. CO3: 1500 to 2000 ms after cue. **(B)** Top: averaged (over all units, nine rats, and sessions) single unit firing rates, $M^i_L(k)$ for L-L trials, $M^i_R(k)$ for R-R trials, and $M^i(k)$ for mean firing rate over L-L and R-R trials. Bottom: mean ± STD of single unit firing rate difference, $D^i(k)$. **(C)** Rat B11 as an example. Ensemble mean firing rate and rate difference between L-L and R-R trials in CO data window (300 to 1800 ms after cue onset). Four panels from top to bottom: (1) behavioral accuracy; (2) ensemble mean firing rate over sessions; (3) ensemble mean rate difference between L-L and R-R trials over sessions; (4) averaged standard deviation of firing rate over sessions for L-L (red) and R-R (green) trials. **(D)** Linear regression of Z-scored ensemble mean firing rate (top) and ensemble mean firing rate difference between L-L and R-R trials (middle) of seven rats in the CO data window. Linear regression of averaged standard deviation (in Z-score) for L-L (red) and R-R (green) trials of seven rats in the CO data window (bottom).

gradually increased after cue onset and then leveled off at around 60% (still above chance level of 50.02% accuracy when training samples from both classes were randomly shuffled, one-sample t-test, $p < 10^{-5}$). This is consistent with our previous observation of sustained firing rate modulation between L-L and R-R trials during cue-on period (**Figure 5B**). As a comparison, when spike counts in 15 bins (100 ms bin width) together (CO window) were used in SVM model for classification, the average decoding accuracy was 73.01% over all sessions, which was significantly higher than the decoding accuracy when single time bins were used (paired-sample t-test, $p < 10^{-5}$). Therefore, temporal firing patterns or spike counts in multiple consecutive time bins were expected to benefit SVM neural decoding.

Size of Time Bin

Then we tested how the size of a time bin may affect SVM classification. The same 1500 ms (CO window) neural ensemble data was used but spikes were counted in non-overlap time bins of different sizes, ranging from 100 to 750ms. Best classification performance was obtained using 500 ms bins with a 76.18% decoding accuracy (**Figure 6C**). Larger time bins (e.g., 750 ms) resulted in slightly lower classification accuracy (76.01% accuracy; paired-sample t-test, $p < 0.01$) probably due to loss of temporal resolution. However, higher temporal resolutions did not help improve classification accuracy either (paired-sample t-test, $p < 0.01$). Given the above discussion, we used spike counts in 500 ms bins for analyses hereafter unless otherwise specified.

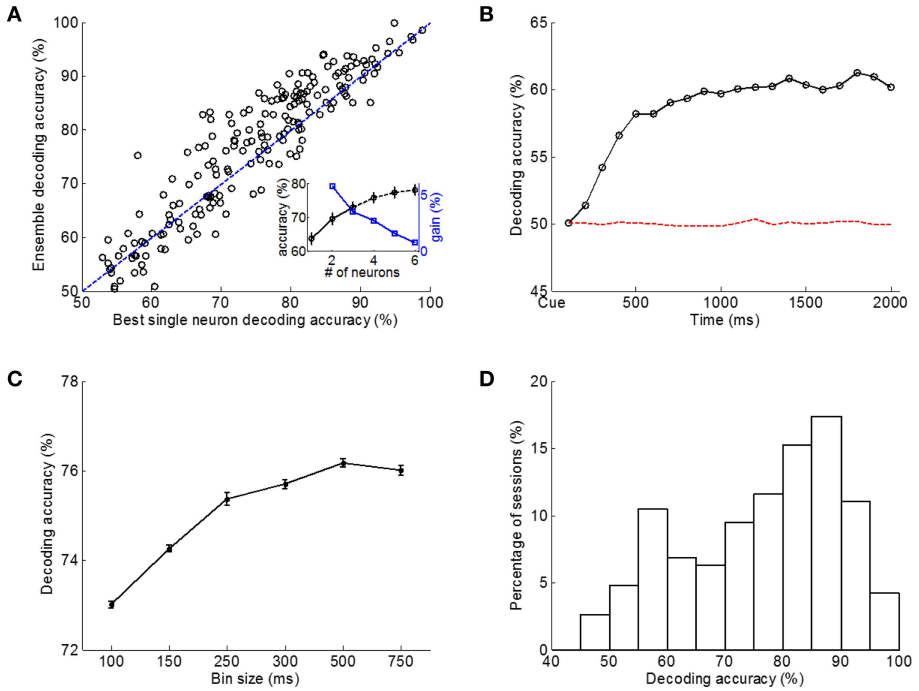


FIGURE 6 | Decoding accuracy between L-L and R-R trials using linear SVM classifiers with different data and classifier configurations. (A) Classification performance using neural ensembles vs. best single neurons. CO data window, 500 ms time bins. Each point represents one of the 190 sessions, and those above the diagonal line indicate better classification performance using the ensemble. Insert: decoding accuracy increased when larger ensembles were used, but the speed of increase in decoding accuracy by using larger ensembles slowed down as ensemble size increased. **(B)** SVM classification accuracy using data of single 100 ms time bins, averaged over all sessions with chance level classification accuracy (dashed line) obtained using shuffled training data. **(C)** Bin size affected classification performance. **(D)** Histogram of single session SVM classification performance (190 sessions). Averaged classification accuracy (test data set) was 76.18% based on CO data window with 500 ms bins.

The histogram of classification performance in single sessions is shown in **Figure 6D**, where spike counts in 500 ms non-overlapping bins during the 1500 ms CO window of neural ensembles were used for decoding. The mean classification accuracy tested with novel single trial data was 76.18% for all sessions from all rats.

To gain additional insight into the firing patterns during the CO window, SVM decoding analysis was also performed between correct and error trials. Two classes of error trials were considered here: R-L trials stand for those left side choices in response to right side cues, and similarly we define L-R trials. Only 11 out of all 190 sessions had at least 20 trials per class for L-L vs. R-L analysis, and 18 sessions were included in R-R vs. L-R analysis. The average decoding accuracy for L-L vs. R-L trials across the 11 sessions was 53.74% which was slightly but significantly higher than chance (t-test, $p < 0.03$), and the accuracy for R-R vs. L-R decoding was 50.79% which was not significantly different from chance (t-test, $p > 0.60$). Specifically in L-L vs. R-L analysis, decoding accuracy was higher than chance in 7 out of the 11 sessions (mean accuracy was 57.05%), lower than chance in one session (44.86% accuracy), and not different from chance in the remaining three sessions (t-test, $\alpha = 0.001$). For the 18 sessions included for R-R vs. L-R analysis, these numbers were six sessions (58.58% accuracy), six sessions (44.38% accuracy), and six sessions. Alongside these decoding results, it's also worth noticing rat behavioral accuracy data. The average behavioral accuracy was 69.1% for the 11 sessions used for L-L vs. R-L decoding, and the behavioral accuracy for the 18 sessions used for R-R vs. L-R decoding was 59.9%. To summarize, when the same directional choice was made in both correct and error trials (e.g., left side choices in both L-L and R-L trials), the neural patterns associated with the two types of trials were largely similar, but there still seemed to be some difference between the two. When the rat was less clear about correct vs. wrong choices (low behavioral accuracy), the neural activities were more similar for correct (R-R) and wrong (L-R) trials. However, this analysis is not conclusive due to limited data available (only 28 out of all 190 sessions were eligible for this analysis).

Adaptation of Spatiotemporal Firing Patterns with Learning

Before presenting evidence on neural adaptation as learning took place, we first illustrate how SVMs can be used for this purpose. **Figure 7A** is an example of how SVM classification took place to separate L-L and R-R trials where in the figure, we showed the first two principal components of the original spatiotemporal neural ensemble data. As shown, training data samples of the two classes formed distinct clusters and the SVM created an optimal

separating line properly. This classifier was then used to predict rat response of left or right side lever press given a novel single trial neural data sample from the test data set (**Figure 7B**). The classification accuracy on the test set and the averaged decision function values (see Materials and Methods) of test samples from both classes could be calculated.

Figure 7C shows histograms of decision function values of test neural data samples from the 100 runs of randomly selected test samples in two sessions (session 9 and 21) from rat B11 as an example. Decision function values from the two sessions were significantly different between the two classes (ANOVA, $p < 10^{-5}$). But the distance between the mean decision function values of the two classes (0.40 vs. 1.14) was larger in the later session when SVM classification accuracy was also higher (53.08% vs. 64.30%).

As classification accuracy increased over learning sessions, we also see an increase in the distance measurements of decision function values of the two classes provided by SVM (**Figure 7D**, rat B11 as an example). Among the seven rats, linear regression of Z-scored classification accuracy (CO data window, 500 ms bins) against normalized session number had a positive slope of 1.59 (**Figure 7E**), with the 99% confidence interval at [0.91 2.26]. If we divide learning sessions into three stages of equal length, the average of the Z-scored classification accuracy gradually increased over the three stages at −0.52, −0.19, and 0.48 (1st vs. 3rd stage: ANOVA, $p < 0.001$; **Figure 7E**). The Z-scored distance measurement between decision function values of the two classes had a positive regression slope of 1.32 (99% confidence interval at [0.62 2.02]), and increased significantly over the three stages as well to reach their respective Z-scored distance measurement of −0.44, −0.15, and 0.40 (1st vs. 3rd stage: ANOVA, $p < 0.001$; **Figure 7E**). These results suggest enhanced discriminability in spatiotemporal neural activity patterns between L-L and R-R trials as learning progressed.

After examining neural activities in the cue-on period in relation to rat's behavioral learning improvement, we attempted to gain additional insight by investigating neural activity patterns during the response lever press period as a control. A 500 ms time window (LP window, from −100 to 400 ms) around response lever press was used, and spike counts in 100 ms non-overlapping bins were used to build SVMs to decode L-L and R-R trial lever presses. For this time window, Z-scored classification accuracy exhibited a weak rising trend (**Figure 7F**, LP window), with the slope of a linear regression at 0.55, which was not significantly different from 0 since the 99% confidence interval of the slope was [−0.23 1.34].

To compare with those results using the LP data window, we repeated the analysis for the three cue-on period windows (**Figure 7F**). For CO1, from 0.5 to 1.0 s after cue onset while directional movement was being performed, improvement of Z-scored classification accuracy was significant (slope: 0.98; 99% confidence interval of slope: [0.25 1.71]). For CO2, from 1.0 to 1.5 s after cue onset when directional movements were mostly completed, improvement of Z-scored classification accuracy was significant as well (slope: 1.48; 99% confidence interval of slope: [0.79 2.17]). An improvement of Z-scored classification accuracy (slope: 1.12; 99% confidence interval of slope: [0.40 1.84]) was also observed in CO3, from 1.5 to 2.0 s after cue onset which was right before extension of response levers. When the regression analysis was carried out on individual rats, regression slopes for cue-on period data windows were significantly greater than zero (t-test; CO1, $p < 0.01$; CO2, $p < 0.005$; CO3, $p < 0.05$), but not significantly different from zero for LP (t-test, $p > 0.39$). When classification accuracies were averaged for each of the three equal-length learning stages, Z-scored classification accuracy increased gradually over stages for each of the three cue-on period windows (CO1: −0.76, −0.37, and 1.13; CO2: −0.86, −0.24, and 1.10; CO3: −1.05, 0.11, and 0.94; ANOVA, 1st vs. 3rd stage, $p < 0.05$), but remained leveled during the last two stages for LP window (−1.15, 0.67, and 0.48; ANOVA, 2nd vs. 3rd stage, $p > 0.9$). **Figure 7G** illustrates the Z-scored classification accuracies in the three stages for the four data windows, and again consistent increment over the three stages was found in cue-on data windows but not in the LP window. Taken together, enhanced discriminability of neural activity patterns over the entire learning process was mainly found during the cue-on period, but not the LP period when rats actually pressed levers.

To further validate the results from SVM based decoding analyses presented above, we used linear discriminant analysis (LDA) as a second classification method. Similar results from LDA classifiers were obtained as those reported in **Figures 7E–G**.

Regression analysis was also performed between classification accuracy (CO data window, 500 ms bins) and the percentage of trials with correct directional choice in single sessions. Results of the seven rats were plotted in **Figure 8A** individually. The linear regression slope was significantly greater than zero (t-test, $p < 0.05$). Thus, decoding of directional choice using neural activity did improve as rats made progress on the learning task.

To show that spatiotemporal patterns indeed facilitated the observed improvement in directional choice decoding, the mean firing rate of the neural ensemble over the whole CO data window (1500 ms bin, 1-D data samples) was used for classification as a comparison. Regression slope of classification accuracy against the percentage of correct choice trials (i.e., L-L and R-R trials) when using the 1-D data samples is −0.0071 (**Figure 8B**, black; $n = 7$), which is not significantly different from zero (99% confidence interval at [−0.0221 0.0078]). When spatiotemporal patterns (CO data window, 500 ms bins) were used for classification, the regression slope is significantly steeper (ANCOVA, $p < 0.05$), which is 0.0380 with 99% confidence interval at [0.0147 0.0612]. Therefore, neural adaptation associated with directional choice learning is better described by spatiotemporal activity patterns than a low resolution neural activity representation.

Discussion

Seven out of nine rats successfully learned to perform a directional choice task from a naïve state. Using trial-and-error, they were able to associate a light cue with a same side lever pressing. Based on rat behavioral data, we observed that rat movement trajectory and the act of lever press became stereotyped within the first few days and therefore, the motor skill learning factor could be excluded from our analysis of associative learning. In this study, we focused on analyzing neural data from the seven

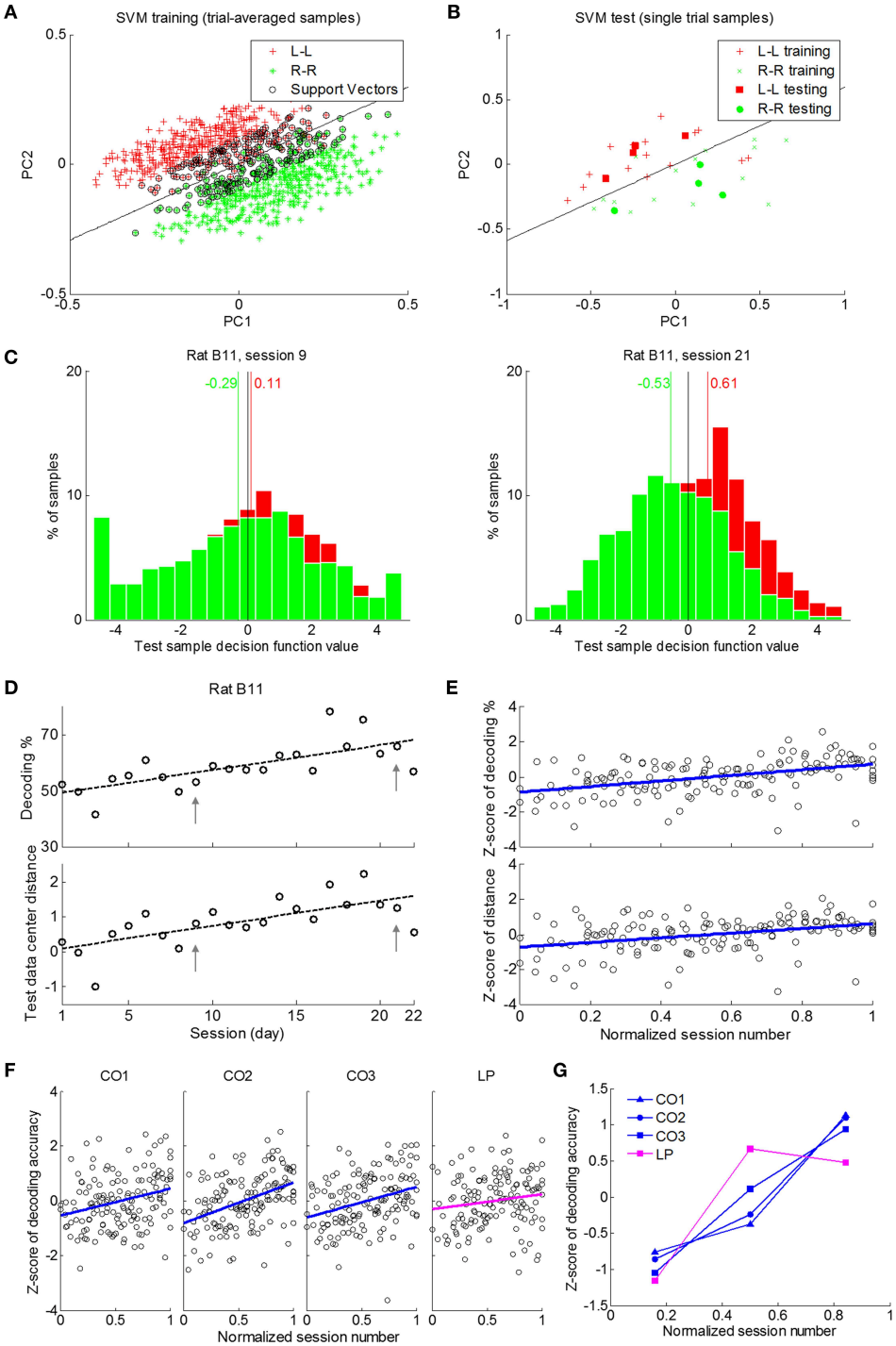


FIGURE 7 | SVM decoding over learning sessions. The CO data window with spike counts in 500 ms time bins was used in all results except those in **(F)** and **(G)** where 100 ms bins were used. **(A)** An illustration of training data samples, support vectors and separating plane (black line) in a 2-D PCA projected space. Data of rat B11 in session 21 were used as an example. **(B)** The original single trial data samples used to generate trial-averaged training data and single-trial test data samples plotted in the same space as in **(A)**. **(C)** Histograms of decision function values of test data set for sessions 9 and 21 of rat B11. **(D)** An example of classification performance over sessions using data of rat B11. Upper panel: Classification accuracy tended to increase with learning. Lower panel: the distance between L-L and R-R data sets in the SVM kernel space increased with learning. **(E)** Classification accuracy and distance between the two classes in SVM kernel space increased through the course of learning (seven rats). **(F)** The increased decoding accuracy was significant in the three cue-on period data windows (CO1, CO2, and CO3), but not in LP window around response lever press. **(G)** Neural activity patterns of response lever press (LP) showed different dynamics in terms of classification accuracies (Z-scored) in three stages of the learning process compared with those in cue-on data windows (CO1, CO2, and CO3).

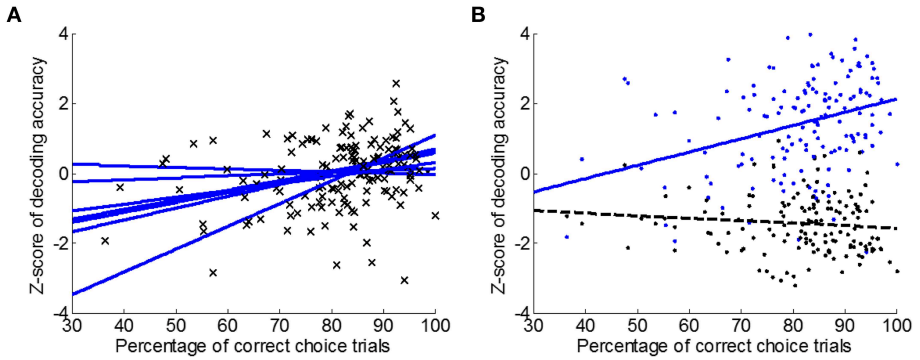


FIGURE 8 | SVM decoding performance vs. behavioral accuracy. (A) Linear regression of Z-scored SVM decoding accuracy (CO data window, 500 ms bins) against the percentage of correct choice trials (i.e., L-L and R-R trials) for individual rats ($n = 7$). Increased decoding accuracy was observed in six out of seven rats. **(B)** Decoding accuracy using spatiotemporal patterns (blue; CO data window, 500 ms bins) improved as behavioral performance improved, but not so when ensemble mean firing rate over the entire CO data window (1500 ms bin) was used for decoding (black).

rats during the cue-on period when they made directional choice decisions. Our results showed that the ensemble mean firing rate over all L-L and R-R trials appeared level over learning sessions. The ensemble mean rate difference between L-L and R-R trials did not show a strong trend of change as task learning progressed either (**Figure 5**). However, when using SVMs to decode directional choice from spatiotemporal neural activity patterns, there was a clear upward trend of SVM decoding accuracy over learning sessions. Correspondingly, there was a clear upward trend in discriminability of neural patterns between left side and right side choices (**Figure 7**). These findings suggest that neural adaptation in rat's motor cortical areas during learning of the directional choice task may lie in the spatiotemporal firing pattern of neural ensembles.

Adaptation of Spatiotemporal Neural Activity Pattern during Task Learning

SVM classifiers were constructed to discriminate neural activity patterns associated with directional choices. During our analysis using SVM, care was given to ensure compatibility when comparing results over learning sessions. First, the same linear kernel SVM model was used for all analyses. Second, each and every classifier was trained with the same numbers of trial samples and tested with the same numbers of samples in the same session (**Figures 6**–**8**). Third, we reported classification accuracy in each session using an averaged result of 100 independent SVM classifiers with randomly selected training and test data samples. As such, SVM classification performance over sessions as reported in **Figures 7**, **8** should be characteristic of neural activity patterns as they adapted with learning.

In this study, we treated single units recorded from the same electrode in different sessions as independent unit records. Actually, our results reported in this study were based on neural ensembles which consisted of all units recorded simultaneously during one session from the same rat. Therefore, we did not intentionally identify and track same neurons over recording sessions. This approach was used before (Cohen and Nicolelis, 2004) and it is adopted in this study since all our results are based on ensemble neural activity.

Motor skill learning is a possible confound when analyzing neural activity in the cue-on period. However, we observed that rats became accustomed to directional movement and response lever press faster than the associative learning aspect of the task. Rat's directional movement became stereotyped quickly after the first few sessions (**Figure 3D** and Yuan et al., 2014) and the response latency of lever press decreased during the first three sessions (**Figure 3B**). Therefore, initial motor skill learning could not explain neural adaptation along an entire learning process lasted for weeks.

There may be other confounding factors in addition to motor skill learning during the first few learning sessions. Notice that right after a press on the center ready lever, at which time the directional cue was presented, there was a 2 s cue-on period. Rats usually moved from the center position to the location of their chosen response lever waiting for the extension of the response levers at the end of the cue-on period. Conceivably during this 2 s window, a rat's anticipation of lever extension and planning of lever press could possibly induce neural modulation. However, behavioral data showed that rats quickly became accustomed to performing the task routine after the first few days as indicated by their stereotyped movement and stable response latency. Therefore, they could quickly become habituated to the extension of response levers as well as the planning of the routine act of response lever press. From that, those potential confounding factors may be excluded from possible reasons for reported neural adaptation.

Would the reported neural pattern adaptation be explained by the repetition of directional movements? Our previous analysis showed that rat directional movement was mostly completed within CO1 window (Yuan et al., 2014), and there was no obvious or systematic movement during CO2 and CO3 windows (**Figure 1B**). Improved discriminability of neural patterns, however, was observed not only in CO1 but also in CO2 and CO3 windows (**Figure 7F**). Response lever press was another action repeatedly performed by the rats during learning of the task.

However, we did not observe significantly improved discrimination of neural patterns in the LP data window (**Figure 7F**). As discussed above, these actions quickly became stereotypical and turned into learned motor skills. Previous primate studies showed that mere repetition of familiar actions did not induce systematic changes in motor cortical neural firing activity (Paz and Vaadia, 2004; Rokni et al., 2007). Taken together, the observed firing pattern adaptation during cue-on period was unlikely to be associated with repeated performance of task related actions.

Reward based stimulus-response association learning is another aspect that may affect neural modulation in the rat frontal areas that we recorded from. Previous studies showed that reward-related action selection is believed to be mediated by the corticostriatal circuitry, linking prefrontal (PFC), premotor, sensorimotor cortices, and the striatum (Balleine et al., 2007). As two important nodes within this circuit of primates, Pasupathy and Miller (2005) found rapid changes in striatum but slow adaptation in PFC during an associative learning task, where the time course of PFC activity had a significantly stronger correlation with the gradual improvement in task performance. In accordance with their findings, the adaptation we observed in rat's frontal areas also correlated with slow improvement of behavioral performance. Given the above considerations, the observed neural adaptation during the cue-on period could be attributed to learning the correct stimulus-response association.

In the rat brain, both AGm and AGl project to basal ganglia (Reep et al., 1987; Cheatwood et al., 2003; Alloway et al., 2009), and both areas receive inputs from basal ganglia through the thalamus (Donoghue and Parham, 1983; Reep et al., 1984). Rat AGm and AGl have connections with a variety of frontal cortical areas as well (Reep et al., 1984, 1990; Hoover and Vertes, 2007). Therefore, these rat motor cortical areas could be in the loop of the reward-related decision making circuit. The neural pattern adaptation in motor cortical ensembles reported here provides neurophysiological evidence for a role of rat motor cortical areas in learning stimulus-response associations, which could be mediated by this neural circuit when rewarded directional choices were learned.

Ensemble Mean Rate vs. Spatiotemporal Activity Pattern

Despite the well-observed phenomenon of dynamic neural modulation in single neurons, relatively uniform firing rates in cortical ensembles have been reported in different brain areas of both primates and rodents while animals performed different tasks (Hoffman and McNaughton, 2002; Carmena et al., 2003; Costa et al., 2006; Pantoja et al., 2007). But that does not include studies that involve learning tasks except a few. In an associative learning task, Cohen and Nicolelis (2004) reported unchanged rate difference over the recorded neural ensemble during the early days, specifically day 2 and day 3. When monkeys were performing a stimulus-response association learning task, firing rate change in single units of motor cortical areas was reported (Mitz et al., 1991; Chen and Wise, 1995; Brasted and Wise, 2004), but it was less certain if and how firing rate of a neural ensemble would change over learning. Here, we report a relatively stable ensemble mean firing rate and ensemble mean rate difference between left and right side choices over a period of about 20–30 sessions covering the entire time course of associative learning. Our results may suggest a balanced increase and decrease in single unit firing rates, which may have contributed to a stable motor cortical ensemble mean firing rate during associative learning. This result is supportive of "the conservation of firing" principle proposed by Nicolelis and Lebedev (2009).

To gain understanding of neural coding beyond ensemble averaged firing rates, investigations of spatiotemporal activity patterns at a fine resolution have brought up new insight on fundamental neural mechanisms in visual attention (Heinze et al., 1994), odor representation (Laurent et al., 1996; Spors and Grinvald, 2002; Rennaker et al., 2007), auditory processing (Kayser et al., 2009), vibrissa deflection coding (Petersen and Diamond, 2000), contextual encoding (Hyman et al., 2012), sequence learning (Ma et al., 2014), and rule learning (Durstewitz et al., 2010), to name a few. In a reaction time study (Laubach et al., 2000) using a rat model, the overall firing rates of the AGm and AGl ensembles did not change significantly but prediction of trial outcome of either correct or error based on spatiotemporal activity patterns improved over learning sessions. Our results appear along similar lines. However, the two experimental protocols are different in a few aspects. The Laubach et al. (2000) experiment used a single stimulus (vibrotactile or auditory) and a lever release for rats to report their detection of presence of sensory cue. In our experiment, alternative choices were associated with distinct stimuli. Besides, Laubach et al. (2000) compared lever release either instructed by a stimulus (correct) or executed spontaneously without stimulus presence (error), while we compared instructed left side and right side choices (both were correct) under distinct cues. This may help rule out confounding factors such as the occurrence of sensory stimuli and prediction of rewards (Carandini and Churchland, 2013).

In the Cohen and Nicolelis (2004) study, prediction of left and right side movements by M1 neural ensembles improved from the first day to the next 2 days. Unfortunately, results in the remaining 8 days as rats' performance continued to improve until reaching a plateau were not available. Here we monitored neural activity patterns through the entire process of associative learning, and demonstrated improved discriminability of spatiotemporal firing patterns in motor cortical ensembles.

Rat Motor Cortical Areas and Associative Learning

The rat AGl area has been considered to correspond with primate primary motor (M1) cortex (Donoghue and Wise, 1982; Donoghue and Parham, 1983). On the other hand, the rat AGm area refers to the medial subdivision of the agranular field of the frontal cortex which differs from the lateral subdivision (AGl) on cytoarchitectonic grounds (Donoghue and Wise, 1982). Other terms referring to this area used in literature include medial precentral area (PrCm, Krettek and Price, 1977), frontal cortical area 2 (Fr2, Zilles, 1985), and secondary motor area (M2, Paxinos and Watson, 2005; MOs, Swanson, 1998). Leaving the inconsistent nomenclature aside, rat AGm was proposed to be

homologous to premotor cortex, supplementary motor area, and frontal and supplementary eye fields in primates (Donoghue and Wise, 1982; Reep et al., 1987; Van Eden et al., 1992; Condé et al., 1995; Sul et al., 2011). However, a clear rat AGm homology in primates has yet to be proved convincingly. Nonetheless, previous neuropsychological studies showed that lesions of rat AGm impaired both the retrieval (Passingham et al., 1988) and the acquisition (Winocur and Eskes, 1998) of visuomotor conditioning, which suggest a role for AGm in stimulus-response associative learning. In line with these reports, we observed firing pattern adaptation during associative learning within the motor cortical neural ensembles, which consisted of AGm neurons and AGl neurons.

It is worth mentioning that in most of the primate studies on associative learning, the animals usually had acquired certain stimulus-response association through long and extensive training. And then, animals would learn a novel pairing (Chen and Wise, 1995; Brasted and Wise, 2004) and/or the reversed pairing (Pasupathy and Miller, 2005; Histed et al., 2009) during one complete recording session. Sometimes the animals would learn a variant of the trained sensorimotor tasks (Mitz et al., 1991; Li et al., 2001; Genovesio et al., 2014), which also could be completed within one session. In these cases, the complete time course of learning could take place in tens of minutes or a single recording session. This may be too soon to result in long-lasting synaptic changes, as suggested by Histed et al. (2009). In our experiment, it took rats several weeks to master the directional choice task. This may allow substantial neural adaptation to take place, possibly through changes at the synaptic level.

Synaptic plasticity has long been hypothesized for being an important neurochemical foundation of learning and memory (Malenka and Bear, 2004; Gilson et al., 2010), and its necessity has been well-supported (Martin et al., 2000). Rat motor cortex is highly capable of functional and structural changes even in adulthood. Reorganization of motor maps has been observed in various experiments (Sanes et al., 1990, 1992; Lee et al., 2003), including animals learning a motor skill (Nudo et al., 1996; Kleim et al., 1998, 2004). Cortical synaptogenesis has been reported during motor training (Jones et al., 1999; Kleim et al., 2004). And recent studies demonstrated learning-induced dendritic spine changes in rodents performing motor tasks (Xu et al., 2009; Yang et al., 2009; Wang et al., 2011). While these changes were related to learning of certain motor skill, whether learning stimulus-response association would induce such changes in rat motor areas is unclear. Given the lengthiness of the task, rats in our experiment may have a chance to experience synaptic modification during the learning process that lasted several weeks. Consequently, an enhanced spatiotemporal neural representation may become increasingly predictable of directional choice.

Acknowledgments

This work was supported in part by the National Science Foundation under Grants ECS-0702057 and ECS-1002391.

References

Alloway, K. D., Smith, J. B., Beauchemin, K. J., and Olson, M. L. (2009). Bilateral projections from rat MI whisker cortex to the neostriatum, thalamus, and claustrum: forebrain circuits for modulating whisking behavior. *J. Comp. Neurol.* 515, 548–564. doi: 10.1002/cne.22073

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn. Sci.* 12, 193–200. doi: 10.1016/j.tics.2008.02.004

Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* 27, 8161–8165. doi: 10.1523/JNEUROSCI.1554-07.2007

Boettiger, C. A., and D'Esposito, M. (2005). Frontal networks for learning and executing arbitrary stimulus-response associations. *J. Neurosci.* 25, 2723–2732. doi: 10.1523/JNEUROSCI.3697-04.2005

Brasted, P. J., and Wise, S. P. (2004). Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. *Eur. J. Neurosci.* 19, 721–740. doi: 10.1111/j.0953-816X.2003.03181.x

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2, 121–167. doi: 10.1023/A:1009715923555

Carandini, M., and Churchland, A. K. (2013). Probing perceptual decisions in rodents. *Nat. Neurosci.* 16, 824–831. doi: 10.1038/nn.3410

Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., et al. (2003). Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS Biol.* 1:e42. doi: 10.1371/journal.pbio.0000042

Carpenter, A. F., Georgopoulos, A. P., and Pellizzer, G. (1999). Motor cortical encoding of serial order in a context-recall task. *Science* 283, 1752–1757. doi: 10.1126/science.283.5408.1752

Cheatwood, J. L., Reep, R. L., and Corwin, J. V. (2003). The associative striatum: cortical and thalamic projections to the dorsocentral striatum in rats. *Brain Res.* 968, 1–14. doi: 10.1016/S0006-8993(02)04212-9

Chen, L., and Wise, S. (1995). Neuronal activity in the supplementary eye field during acquisition of conditional oculomotor associations. *J. Neurophysiol.* 73, 1101–1121.

Cohen, D., and Nicolelis, M. A. (2004). Reduction of single-neuron firing uncertainty by cortical ensembles during motor skill learning. *J. Neurosci.* 24, 3574–3582. doi: 10.1523/JNEUROSCI.5361-03.2004

Condé, F., Maire-Lepoivre, E., Audinat, E., and Crépel, F. (1995). Afferent connections of the medial frontal cortex of the rat. II. Cortical and subcortical afferents. *J. Comp. Neurol.* 352, 567–593. doi: 10.1002/cne.903520407

Costa, R. M., Lin, S.-C., Sotnikova, T. D., Cyr, M., Gainetdinov, R. R., Caron, M. G., et al. (2006). Rapid alterations in corticostriatal ensemble coordination during acute dopamine-dependent motor dysfunction. *Neuron* 52, 359–369. doi: 10.1016/j.neuron.2006.07.030

Donoghue, J. P., and Parham, C. (1983). Afferent connections of the lateral agranular field of the rat motor cortex. *J. Comp. Neurol.* 217, 390–404. doi: 10.1002/cne.902170404

Donoghue, J. P., and Wise, S. P. (1982). The motor cortex of the rat: cytoarchitecture and microstimulation mapping. *J. Comp. Neurol.* 212, 76–88. doi: 10.1002/cne.902120106

Durstewitz, D., Vittoz, N. M., Floresco, S. B., and Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* 66, 438–448. doi: 10.1016/j.neuron.2010.03.029

Eisenberg, M., Shmuelof, L., Vaadia, E., and Zohary, E. (2011). The representation of visual and motor aspects of reaching movements in the human motor cortex. *J. Neurosci.* 31, 12377–12384. doi: 10.1523/JNEUROSCI.0824-11.2011

Fluet, M.-C., Baumann, M. A., and Scherberger, H. (2010). Context-specific grasp movement representation in macaque ventral premotor cortex. *J. Neurosci.* 30, 15175–15184. doi: 10.1523/JNEUROSCI.3343-10.2010

Genovesio, A., Tsujimoto, S., Navarra, G., Falcone, R., and Wise, S. P. (2014). Autonomous encoding of irrelevant goals and outcomes by prefrontal cortex neurons. *J. Neurosci.* 34, 1970–1978. doi: 10.1523/JNEUROSCI.3228-13.2014

Gilson, M., Burkitt, A., and Van Hemmen, J. L. (2010). STDP in recurrent neuronal networks. *Front. Comput. Neurosci.* 4:23. doi: 10.3389/fncom.2010.00023

Heinze, H. J., Mangun, G. R., Burchert, W., Hinrichs, H., Scholz, M., Münte, T. F., et al. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature* 372, 543–546. doi: 10.1038/372543a0

Histed, M. H., Pasupathy, A., and Miller, E. K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* 63, 244–253. doi: 10.1016/j.neuron.2009.06.019

Hoffman, K., and McNaughton, B. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science* 297, 2070–2073. doi: 10.1126/science.1073538

Hoover, W. B., and Vertes, R. P. (2007). Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Struct. Funct.* 212, 149–179. doi: 10.1007/s00429-007-0150-4

Huber, D., Gutnisky, D. A., Peron, S., O'Connor, D. H., Wiegert, J. S., Tian, L., et al. (2012). Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* 484, 473–478. doi: 10.1038/nature11039

Hyman, J. M., Ma, L., Balaguer-Ballester, E., Durstewitz, D., and Seamans, J. K. (2012). Contextual encoding by ensembles of medial prefrontal cortex neurons. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5086–5091. doi: 10.1073/pnas.1114415109

Jones, T. A., Chu, C. J., Grande, L. A., and Gregory, A. D. (1999). Motor skills training enhances lesion-induced structural plasticity in the motor cortex of adult rats. *J. Neurosci.* 19, 10153–10163.

Kargo, W. J., and Nitz, D. A. (2004). Improvements in the signal-to-noise ratio of motor cortex cells distinguish early versus late phases of motor skill learning. *J. Neurosci.* 24, 5560–5569. doi: 10.1523/JNEUROSCI.0562-04.2004

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. doi: 10.1016/j.neuron.2009.01.008

Kleim, J. A., Barbay, S., and Nudo, R. J. (1998). Functional reorganization of the rat motor cortex following motor skill learning. *J. Neurophysiol.* 80, 3321–3325.

Kleim, J. A., Hogg, T. M., Vandenberg, P. M., Cooper, N. R., Bruneau, R., and Remple, M. (2004). Cortical synaptogenesis and motor map reorganization occur during late, but not early, phase of motor skill learning. *J. Neurosci.* 24, 628–633. doi: 10.1523/JNEUROSCI.3440-03.2004

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science* 302, 1181–1185. doi: 10.1126/science.1088545

Krettek, J., and Price, J. (1977). The cortical projections of the mediodorsal nucleus and adjacent thalamic nuclei in the rat. *J. Comp. Neurol.* 171, 157–191. doi: 10.1002/cne.901710204

Laubach, M., Wessberg, J., and Nicolelis, M. A. (2000). Cortical ensemble activity increasingly predicts behaviour outcomes during learning of a motor task. *Nature* 405, 567–571. doi: 10.1038/35014604

Laurent, G., Wehr, M., and Davidowitz, H. (1996). Temporal representations of odors in an olfactory network. *J. Neurosci.* 16, 3837–3847.

Lee, L., Siebner, H. R., Rowe, J. B., Rizzo, V., Rothwell, J. C., Frackowiak, R. S., et al. (2003). Acute remapping within the motor system induced by low-frequency repetitive transcranial magnetic stimulation. *J. Neurosci.* 23, 5308–5318. Available online at: http://www.jneurosci.org/content/23/12/5308.long

Li, C.-S. R., Padoa-Schioppa, C., and Bizzi, E. (2001). Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. *Neuron* 30, 593–607. doi: 10.1016/S0896-6273(01)00301-4

Ma, L., Hyman, J. M., Phillips, A. G., and Seamans, J. K. (2014). Tracking progress toward a goal in corticostriatal ensembles. *J. Neurosci.* 34, 2244–2253. doi: 10.1523/JNEUROSCI.3834-13.2014

Malenka, R. C., and Bear, M. F. (2004). LTP and LTD: an embarrassment of riches. *Neuron* 44, 5–21. doi: 10.1016/j.neuron.2004.09.012

Martin, S. J., Grimwood, P. D., and Morris, R. G. M. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* 23, 649–711. doi: 10.1146/annurev.neuro.23.1.649

Matsuzaka, Y., Picard, N., and Strick, P. L. (2007). Skill representation in the primary motor cortex after long-term practice. *J. Neurophysiol.* 97, 1819–1832. doi: 10.1152/jn.00784.2006

Mitz, A. R., Godschalk, M., and Wise, S. P. (1991). Learning-dependent neuronal activity in the premotor cortex: activity during the acquisition of conditional motor associations. *J. Neurosci.* 11, 1855–1872.

Murray, E. A., Bussey, T. J., and Wise, S. P. (2000). Role of prefrontal cortex in a network for arbitrary visuomotor mapping. *Exp. Brain Res.* 133, 114–129. doi: 10.1007/s002210000406

Neafsey, E. J., Bold, E. L., Haas, G., Hurley-Gius, K. M., Quirk, G., Sievert, C. F., et al. (1986). The organization of the rat motor cortex: a microstimulation mapping study. *Brain Res. Rev.* 11, 77–96. doi: 10.1016/0165-0173(86)90011-1

Nicolelis, M. A., and Lebedev, M. A. (2009). Principles of neural ensemble physiology underlying the operation of brain–machine interfaces. *Nat. Rev. Neurosci.* 10, 530–540. doi: 10.1038/nrn2653

Nudo, R. J., Milliken, G. W., Jenkins, W. M., and Merzenich, M. M. (1996). Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. *J. Neurosci.* 16, 785–807.

Pantoja, J., Ribeiro, S., Wiest, M., Soares, E., Gervasoni, D., Lemos, N. A., et al. (2007). Neuronal activity in the primary somatosensory thalamocortical loop is modulated by reward contingency during tactile discrimination. *J. Neurosci.* 27, 10608–10620. doi: 10.1523/JNEUROSCI.5279-06.2007

Passingham, R., Myers, C., Rawlins, N., Lightfoot, V., and Fearn, S. (1988). Premotor cortex in the rat. *Behav. Neurosci.* 102, 101–109. doi: 10.1037/0735-7044.102.1.101

Pasupathy, A., and Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433, 873–876. doi: 10.1038/nature03287

Paxinos, G., and Watson, C. (2005). *The Rat Brain in Stereotaxic Coordinates.* 5th Edn. San Diego, CA: Academic press.

Paz, R., and Vaadia, E. (2004). Learning-induced improvement in encoding and decoding of specific movement directions by neurons in the primary motor cortex. *PLoS Biol.* 2:e45. doi: 10.1371/journal.pbio.0020045

Petersen, R. S., and Diamond, M. E. (2000). Spatial-temporal distribution of whisker-evoked activity in rat somatosensory cortex and the coding of stimulus location. *J. Neurosci.* 20, 6135–6143. Available online at: http://www.jneurosci.org/content/20/16/6135.long

Reep, R., Corwin, J., Hashimoto, A., and Watson, R. (1987). Efferent connections of the rostral portion of medial agranular cortex in rats. *Brain Res. Bull.* 19, 203–221. doi: 10.1016/0361-9230(87)90086-4

Reep, R. L., Corwin, J. V., Hashimoto, A., and Watson, R. T. (1984). Afferent connections of medial precentral cortex in the rat. *Neurosci. Lett.* 44, 247–252. doi: 10.1016/0304-3940(84)90030-2

Reep, R. L., Goodwin, G. S., and Corwin, J. V. (1990). Topographic organization in the corticocortical connections of medial agranular cortex in rats. *J. Comp. Neurol.* 294, 262–280. doi: 10.1002/cne.902940210

Remple, M. S., Bruneau, R. M., Vandenberg, P. M., Goertzen, C., and Kleim, J. A. (2001). Sensitivity of cortical movement representations to motor experience: evidence that skill learning but not strength training induces cortical reorganization. *Behav. Brain Res.* 123, 133–141. doi: 10.1016/S0166-4328(01)00199-1

Rennaker, R. L., Chen, C.-F. F., Ruyle, A. M., Sloan, A. M., and Wilson, D. A. (2007). Spatial and temporal distribution of odorant-evoked activity in the piriform cortex. *J. Neurosci.* 27, 1534–1542. doi: 10.1523/JNEUROSCI.4072-06.2007

Rokni, U., Richardson, A. G., Bizzi, E., and Seung, H. S. (2007). Motor learning with unstable neural representations. *Neuron* 54, 653–666. doi: 10.1016/j.neuron.2007.04.030

Sanes, J. N., and Donoghue, J. P. (2000). Plasticity and primary motor cortex. *Annu. Rev. Neurosci.* 23, 393–415. doi: 10.1146/annurev.neuro.23.1.393

Sanes, J. N., Suner, S., and Donoghue, J. P. (1990). Dynamic organization of primary motor cortex output to target muscles in adult rats I. Long-term patterns of reorganization following motor or mixed peripheral nerve lesions. *Exp. Brain Res.* 79, 479–491. doi: 10.1007/BF00229318

Sanes, J. N., Wang, J., and Donoghue, J. P. (1992). Immediate and delayed changes of rat motor cortical output representation with new forelimb configurations. *Cereb. Cortex* 2, 141–152. doi: 10.1093/cercor/2.2.141

Spors, H., and Grinvald, A. (2002). Spatio-temporal dynamics of odor representations in the mammalian olfactory bulb. *Neuron* 34, 301–315. doi: 10.1016/S0896-6273(02)00644-X

Sul, J. H., Jo, S., Lee, D., and Jung, M. W. (2011). Role of rodent secondary motor cortex in value-based action selection. *Nat. Neurosci.* 14, 1202–1208. doi: 10.1038/nn.2881

Swanson, L. W. (1998). *Brain Maps: Structure of the Rat Brain*. 2nd Edn. Amsterdam: Elsevier.

Van Eden, C., Lamme, V., and Uylings, H. (1992). Heterotopic cortical afferents to the medial prefrontal cortex in the rat. A combined retrograde and anterograde tracer study. *Eur. J. Neurosci.* 4, 77–97. doi: 10.1111/j.1460-9568.1992.tb00111.x

Wang, L., Conner, J. M., Rickert, J., and Tuszynski, M. H. (2011). Structural plasticity within highly specific neuronal populations identifies a unique parcellation of motor learning in the adult brain. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2545–2550. doi: 10.1073/pnas.1014335108

Winocur, G., and Eskes, G. (1998). Prefrontal cortex and caudate nucleus in conditional associative learning: dissociated effects of selective brain lesions in rats. *Behav. Neurosci.* 112, 89–101. doi: 10.1037/0735-7044.112.1.89

Xu, T., Yu, X., Perlik, A. J., Tobin, W. F., Zweig, J. A., Tennant, K., et al. (2009). Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature* 462, 915–919. doi: 10.1038/nature08389

Yang, C., Olson, B., and Si, J. (2011). A multiscale correlation of wavelet coefficients approach to spike detection. *Neural Comput.* 23, 215–250. doi: 10.1162/NECO_a_00063

Yang, G., Pan, F., and Gan, W.-B. (2009). Stably maintained dendritic spines are associated with lifelong memories. *Nature* 462, 920–924. doi: 10.1038/nature08577

Yuan, Y., Mao, H., and Si, J. (2014). Cortical neural responses to previous trial outcome during learning of a directional choice task. *J. Neurophysiol.* doi: 10.1152/jn.00238.2014. [Epub ahead of print].

Yuan, Y., Yang, C., and Si, J. (2012). The M-Sorter: an automatic and robust spike detection and classification system. *J. Neurosci. Methods* 210, 281–290. doi: 10.1016/j.jneumeth.2012.07.012

Zach, N., Inbar, D., Grinvald, Y., Bergman, H., and Vaadia, E. (2008). Emergence of novel representations in primary motor cortex and premotor neurons during associative learning. *J. Neurosci.* 28, 9545–9556. doi: 10.1523/JNEUROSCI.1965-08.2008

Zilles, K. J. (1985). *The Cortex of the Rat: A Stereotaxic Atlas*. Berlin: Springer-Verlag. doi: 10.1007/978-3-642-70573-1

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.

frontiers in
SYSTEMS NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 26 February 2014
doi: 10.3389/fnsys.2014.00029

Capture of fixation by rotational flow; a deterministic hypothesis regarding scaling and stochasticity in fixational eye movements

***Nicholas M. Wilkinson*[1]* and *Giorgio Metta*[1,2]**

[1] *iCub Facility, Fondazione Istituto Italiano di Tecnologia, Genova, Italy*
[2] *Centre for Robotics and Neural Systems, School of Computing and Mathematics, University of Plymouth, Plymouth, UK*

Edited by:
Ruben Moreno-Bote, Foundation Sant Joan de Deu, Spain

Reviewed by:
Susana Martinez-Conde, Barrow Neurological Institute, USA
Zachary P. Kilpatrick, University of Houston, USA

***Correspondence:**
Nicholas M. Wilkinson, iCub Facility, Fondazione Istituto Italiano di Tecnologia, via Morego, 30 Genova, Italy
e-mail: nicholas.wilkinson@iit.it

Visual scan paths exhibit complex, stochastic dynamics. Even during visual fixation, the eye is in constant motion. Fixational drift and tremor are thought to reflect fluctuations in the *persistent neural activity* of neural integrators in the oculomotor brainstem, which integrate sequences of transient saccadic velocity signals into a short term memory of eye position. Despite intensive research and much progress, the precise mechanisms by which oculomotor posture is maintained remain elusive. Drift exhibits a stochastic statistical profile which has been modeled using random walk formalisms. Tremor is widely dismissed as noise. Here we focus on the dynamical profile of fixational tremor, and argue that tremor may be a signal which usefully reflects the workings of oculomotor postural control. We identify signatures reminiscent of a certain flavor of transient neurodynamics; toric traveling waves which rotate around a central phase singularity. Spiral waves play an organizational role in dynamical systems at many scales throughout nature, though their potential functional role in brain activity remains a matter of educated speculation. Spiral waves have a repertoire of functionally interesting dynamical properties, including persistence, which suggest that they could in theory contribute to persistent neural activity in the oculomotor postural control system. Whilst speculative, the *singularity hypothesis* of oculomotor postural control implies testable predictions, and could provide the beginnings of an integrated dynamical framework for eye movements across scales.

Keywords: fixational eye movement, tremor, traveling waves, spiral wave, phase singularity, Lévy walk, persistent neural activity, neural integrator

1. INTRODUCTION

During fixation the eye is not still. Three main classes of *fixational eye movement* (henceforth FEM) have been identified (Martinez-Conde et al., 2004). Microsaccades are very fast movements which occur relatively infrequently. Drift is a slow, meandering component which occupies most of fixation time. Tremor is a fast, low amplitude aperiodic oscillation imposed on drift. Microsaccades are in many ways much like saccades on a tiny scale (Ko et al., 2010; Kagan and Hafed, 2013; Martinez-Conde et al., 2013; Otero-Millan et al., 2013; Poletti et al., 2013), though they may also be linked to the drift component (Engbert and Mergenthaler, 2006; Engbert et al., 2011). FEM have classically been thought to counteract sensory adaptation. Recent evidence suggests that FEM play a more sophisticated role, optimizing visual flow for the response properties of retinal ganglion cells (Rucci et al., 2007; Kuang et al., 2012) and relocating the highest resolution parts of the retina with great precision (Ko et al., 2010; Poletti et al., 2013). Some theories suggest that FEM perform an active perceptual palpitation of the visual scene which is fundamental to vision (Ahissar and Arieli, 2001, 2012; O'Regan and Noë, 2001). Recently, very high resolution eye movement data based on tracking tiny movements of ocular vein structure in three dimensions has revealed more structure to FEM than had previously been suspected (Li and Zhang, 2012; Zhang and Li, 2012). These studies reported microsaccades which were not straight and ballistic (as previously thought), but curving, and even bent and jerky. Relatively little detailed information was given, but it was reported that the drift-tremor combination took a complex, curling trajectory. These high resolution data may enable new insight into the underlying generative mechanisms of fixational eye movements. Oculomotor postural control is mediated by brainstem circuits (Aksay et al., 2000, 2007; Sparks, 2002) and is strongly associated with *persistent neural activity* (Major and Tank, 2004), which plays the role of *integrating* transient stimulation from superior colliculus reflecting saccadic velocity commands into persistent activity encoding the new eye position. The neuroanatomy and functional circuitry of oculomotor postural control has been intensively studied (e.g., Aksay et al., 2000, 2001, 2003; Miri et al., 2011a,b; Fisher et al., 2013), but the precise mechanisms underlying drift and tremor remain elusive.

Rotational waveforms (aka spiral waves, vortices, tori) are a commonplace, universal dynamical form which play an organizing role in dynamical systems at all scales, from galaxies to weather to evolution to organisms to organs to cells to photons (Toomre, 1969; Da-sheng, 1980; Boerlijst and Hogeweg, 1991; Gray and Jalife, 1996; Winfree, 2001; Molina-Terriza et al., 2007; Schecter et al., 2008; Taniguchi et al., 2013). This generality led Winfree (2001) to suggest toroidal temporal structure as

a fundamental aspect of biological time, a notion for which the reference provides many empirical examples. Spiral waves are a canonical mode of pattern formation in dissipative systems operating far from equilibrium (Kuramoto and Koga, 1981; Cross and Hohenberg, 1993). The brain, by necessity, is one such system (e.g., Kelso, 1995; Ermentrout, 1998). Thus if toroidal waveforms (termed *spiral waves* in two dimensions, and *scroll waves* in three dimensions) do not play an organizing role in normal neurodynamics, then the brain must be considered something of an exception to the rule, which would require explanation. The suspected role of spiral waves in some pathological scenarios such as epilepsy (Milton, 2012) and cardiac fibrillation (Gray et al., 1998) suggests that the nervous system possesses mechanisms for actively suppressing turbulence and spatiotemporal chaos (e.g., Schiff et al., 1994), but observations of spiral waves in non-pathological settings (Jung et al., 1998; Huang et al., 2004, 2010) clarify that such suppression is not complete or universal. Indeed, an active field of study in cardiac defibrillation is the suppression of spatiotemporal chaos and turbulent neural activity by the seeding of spiral waves (e.g., Zhang et al., 2002; Xiao-Ping et al., 2011).

The current contribution hypothesizes a connection between quasi-persistent spiral neurodynamics and *persistent neural activity* in the context of oculomotor postural control (Major and Tank, 2004). Sections 2 and 3 respectively introduce the literature on spiral waves and fixational eye movements. Section 4 details our motivations in proposing the *singularity hypothesis* of postural memory. Our purpose is not an exhaustive review, nor to convince the reader that our hypothesis is necessarily correct, but a targeted presentation of empirical evidence and functional arguments which render the singularity hypothesis interesting, plausible and worth testing. Section 5 offers some concluding remarks. Predictions are presented in boxes in the main text.

2. TRANSIENT NEURODYNAMICS AND SPIRAL WAVES

The classical focus on attractor networks in systems neuroscience (see for review Amit, 1992) is increasingly being enriched by a modern synthesis which also stresses the importance of self-organization and transient neural dynamics (Rabinovich et al., 2001; Maass et al., 2002; Seliger et al., 2003; Durstewitz and Deco, 2008; Friston et al., 2012; Milton, 2012), fractality in physiology (Goldberger and West, 1987; West et al., 1994; Werner, 2010; West, 2010), self-organizing criticality (Bak et al., 1987; Bak, 1996; Jung et al., 1998), chaotic itinerancy (Tsuda, 1991, 2001; Kaneko and Tsuda, 2003) and dynamic pattern formation in non-equilibrium dissipative systems (Cross and Hohenberg, 1993).

Neural spiral waves are an intriguing class of quasi-persistent transient neurodynamics, whose functional potential in brain activity remains an open question. They have received extensive theoretical attention in terms of their abstract properties in networks (e.g., Coombes, 2005; Kilpatrick and Bressloff, 2010b; Ma et al., 2012b), but surprisingly little attention in terms of concrete cases linking their dynamics to perception and behavior. We have conducted preliminary modeling studies employing spiral waves for visual salience mapping (Wilkinson and Metta, 2011; Wilkinson et al., 2011), and spiral neurodynamics have linked to visual geometric hallucination (Bressloff et al., 2001; Kilpatrick and Ermentrout, 2012a,b; Froese et al., 2013). At the motor end, Heitmann, Breakspear and colleagues have developed physiologically explanatory models showing how traveling waves (including spirals) can encode motor trajectories (Heitmann, 2013).

2.0.1. Spiral waves in nature, biology, and the brain

The multiscale ubiquity of spiral waves in nature and biology (Toomre, 1969; Lechleiter et al., 1991; Winfree, 2001), and their interesting dynamical properties (Boerlijst and Hogeweg, 1991; Biktashev and Holden, 1993, 1995; Langham and Barkley, 2013), have motivated many physical, chemical, and mathematical studies. Arthur Winfree pioneered computational and empirical investigations of toroidal dynamics in chemical and biological systems (Winfree, 1967, 1972). Many biological dynamics exhibit toroidal form (Winfree, 2001). The modern understanding of pathological heart fibrillation (and de-fibrillation intervention) is perhaps the most prominent medical application of this work (e.g., Gray et al., 1998; Gray and Chattipakorn, 2005), though cellular calcium dynamics is another important example (Lechleiter et al., 1991). Spirals are reentrant waves which circle around a central rotor known as a *phase singularity* (Winfree, 1991); a point of maximally uncertain phase, surrounded by points of all phases. The central rotor of a whirlpool or tornado provides a physical example in three dimensions.

Propagating calcium waves in astrocyte networks are thought to play an important role in regulating brain activity (Cornell-Bell and Finkbeiner, 1991; Finkbeiner, 1992). Jung et al. (1998) observed that Ca^{2+} spiral waves exhibiting scale-free distributions suggestive of self-organizing criticality (Bak et al., 1987) are characteristic of healthy function, whilst epileptic events are characterized by the breakdown of this scaling. In neural tissue, traveling waves have been observed widely in various species in both sensory and motor cortices (see for review Wu et al., 2008; Sato et al., 2012) via voltage sensitive dye imaging ("VSDI"). VSDI is an invasive optical imaging method which enables measurement of subthreshold changes in membrane potential with high spatiotemporal resolution (Grinvald and Hildesheim, 2004). Spiral dynamics are commonplace in the dynamics of simulated excitable media including networks of model neurons (Milton et al., 1993; Winfree, 2001; Chun-Ni et al., 2010; Yu et al., 2010; Ma et al., 2012a), and have been observed in mammalian (Huang et al., 2004, 2010) and reptilian (Prechtl et al., 1997) cortex. Movies of cortical spiral waves in the VSDI signal (from Huang et al., 2010) can be found http://www9.georgetown.edu/faculty/wuj/propagationwave.html.

It has been suggested that spiral waves may play an organizing role in neural field interactions (Wu et al., 2008; Freeman, 2009; Huang et al., 2010). Short-lived spiral waves are frequently observed in the healthy case (Huang et al., 2010), but the growth of spiral wave formations of large duration and extent has been linked to pathological conditions including heart fibrillation (Gray et al., 1998) and epileptic seizure (Milton and Jung, 2003; Viventi et al., 2011; Milton, 2012; Stacey, 2012). This is suggestive that spiral waves are a part of normal function, whether constitutive or epiphenomenal, but that their (potentially useful) tendency to enslave surrounding dynamics (e.g., Savill et al.,

1997; Yang and Yang, 2007; Huang et al., 2010) has to be carefully controlled.

2.0.2. Dynamical behavior of spiral waves

The dynamical behavior of spiral waves can be complex and is the subject of extensive research. A useful introduction with video visualizations is given at Björn Sandstede's website hosted by the Department of Applied Mathematics at Brown University, USA, http://www.dam.brown.edu/people/sandsted/research.php?project=spirals. At the risk of oversimplifying, the behavior of spiral singularities exhibits three basic components. Firstly, the rotational orbit of the whole spiral wave will exhibit a characteristic frequency and phase, and is reflected in a small, "on the spot" circular rotation of the phase singularity at the spiral tip. The current hypothesis proposes that this component may correspond to the low frequency peak observed in tremor statistics (Spauschus et al., 1999; Greschner et al., 2002), and could explain new high resolution observations of curling trajectories of drift and tremor (Li and Zhang, 2012; Zhang and Li, 2012).

This trajectory may be perturbed in various ways to take on a locally more complex, globally drifting form. This is known as spiral *drift*, and occurs in response to various forms of symmetry breaking perturbations/gradients in the external milieu (Biktashev and Holden, 1995; Wulff, 1996; Sandstede et al., 1999; Biktashev, 2007). Figures 1–11 in the Scholarpedia article Biktashev (2007) (http://www.scholarpedia.org/article/Drift_of_spiral_waves), display images and animated movies of the trajectory of spiral singularities under various forms of symmetry breaking. Note the basic curling trajectory, whose period is equal to that of the wave's orbit. This spiral drift in response to symmetry breaking perturbations in the excitability of the medium is the neural correlate we hypothesize for the well known slow component of fixational drift (Martinez-Conde et al., 2004; Rolfs, 2009).

In addition to the above relatively slow components, fast, aperiodic oscillatory modulations of the basic curling trajectory can result from instabilities at the phase singularity (Winfree, 1991). The singularity is the point at which all surrounding signals cancel exactly, and so small fluctuations in the surround cause this point of balance to jitter unpredictably. Gray et al. (1998); Bray et al. (2001) tracked the spacetime trajectory of phase singularities in cardiac fibrillation data. **Figure 1**, from Bray et al. (2001) depicts the evolution of a real cardiac phase singularity (white tube inside black mesh) in detail over one cycle of the carrying spiral wave. **Figure 2**, also from Bray et al. (2001) graphs longer trajectories of the singularities of four interacting spirals. Note the fast (80–90 Hz) aperiodic oscillation superimposed on the basic curling trajectory, much faster than the period of the carrier wave. Gray et al. (1998) reported similar spiral meander during cardiac fibrillation. Though here in cardiac tissue, this instability at the singularity is a universal feature of spiral waves (Winfree, 1991). The current proposal suggests this instability as the source of the well known high frequency component of fixational tremor (Martinez-Conde et al., 2004; Rolfs, 2009).

2.0.3. The functional role of neural traveling waves

Traveling waves are routinely observed throughout the brain (Wu et al., 2008), and evidence is increasingly suggesting that they play a functional role (Modolo et al., 2011; Sato et al., 2012; Bahramisharif et al., 2013). Heitmann, Breakspear and colleagues have produced a series of physiologically explanatory and plausible models showing how traveling waves can encode motor trajectories read out by dendritic spatial filters (Breakspear et al., 2010; Heitmann et al., 2012, 2013). These are particularly interesting in the current context. In these models, traveling waves encode motor patterns defining *movement*, whilst synchrony constitutes the *resting state*. The current model, in which spiral waves encode for the active holding of posture, sits well in this framework, because spirals, unlike other traveling waves, *have and (almost) hold a location* in a specific sense (Biktasheva and Biktashev, 2003; Langham and Barkley, 2013). This makes them interesting for the kind of active almost-stillness characterizing postural control.

Spiral wave activity has been observed in the VSDI signal, which primarily reflects the field dynamics of sub-threshold membrane potentials (Grinvald and Hildesheim, 2004). These waves can keep cells in a depolarized "ready" state for input, or indeed polarize cells to effectively ignore input (Bahramisharif et al., 2013). This implements a form of spatiotemporally structured *gain control*, widely agreed to be a fundamental aspect of nervous function (Hillyard et al., 1998; Salinas and Thier, 2000; Salinas and Sejnowski, 2001; Rothman et al., 2009; Olsen et al., 2012). Gain fields have been associated with attentional selectivity at both the sensory and motor end (Aston-Jones and Cohen, 2005; Saalmann and Kastner, 2009; Sara and Bouret, 2012).

On this view, fast, aperiodic spiral meander depolarizes a point locus of local cells in the *gamma* band (peaking around 80–90 Hz)

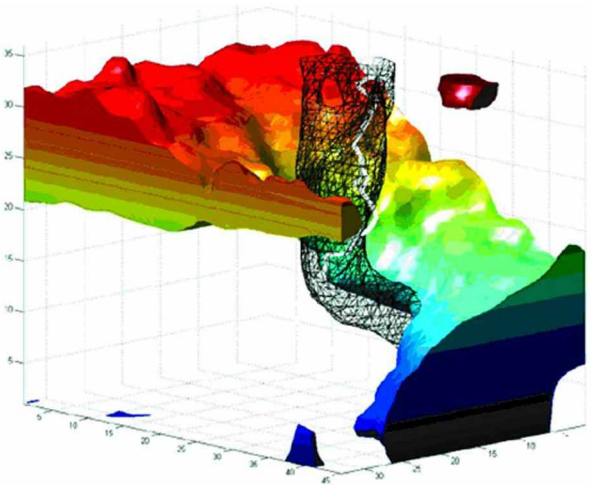


FIGURE 1 | The evolution of a spiral wave over one rotational orbit. Space is represented in the horizontal axes, and time (in milliseconds) on the vertical axis. The black mesh encloses a thresholded area of reduced variance (i.e., low amplitude) at the spiral center, as observed in cortex by Huang et al. (2004, 2010). The white tube within the black mesh tracks the evolution of the phase singularity at the spiral core. Note the fast (80–90 Hz) oscillation of the singularity, which we hypothesize underlies the fast component of fixational tremor. Reproduced from Figure 8B in Bray et al. (2001), copyright John Wiley and Sons Publishing 2001.

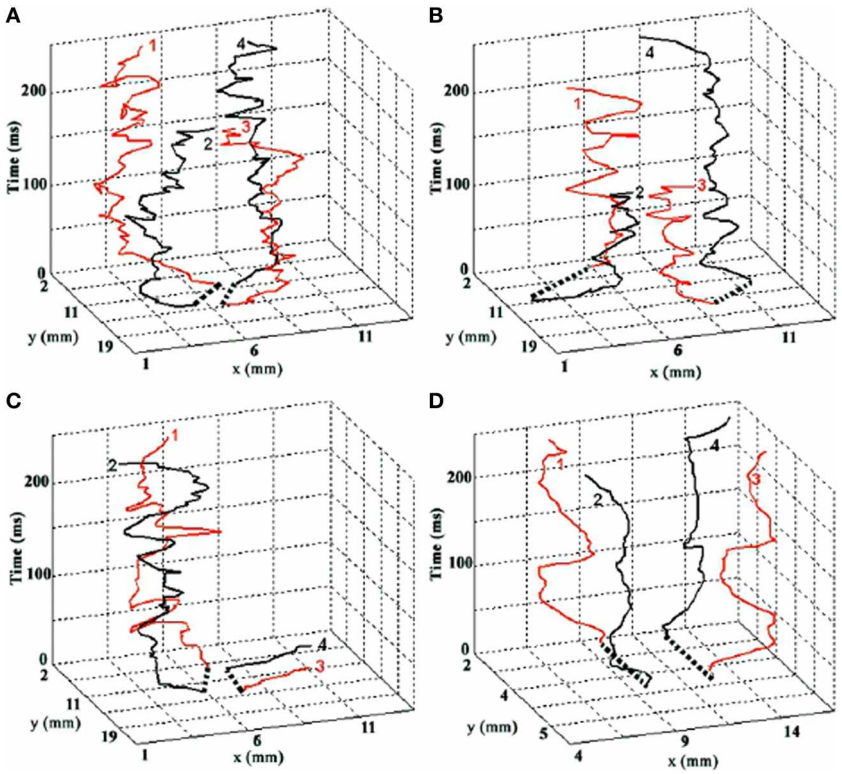


FIGURE 2 | The evolution of four interacting spiral singularities in space (horizontal axes) over time (vertical axis). Time is denoted in milliseconds. Note the fast (80–90 Hz) oscillation of the singularities, which we hypothesize underlies the fast component of fixational tremor. Graphs **A–C** are real trajectories from cardiac data. Graph **D** is from a computational model. Reprinted from Figure 4 in Bray et al. (2001), copyright John Wiley and Sons Publishing 2001.

in a quasi-phaseless manner. The region outside the spiral center is polarized and depolarized periodically by the spin of the spiral arms, on a slower (5–25 Hz) scale dependent on the period of the spiral orbit and the number of spiral arms. Examining the relationship between local field potential and spike rates in the temporal cortex, Zanos et al. (2012) found two populations of cells with just these response characteristics. One population responded at high frequencies in a phase invariant manner, the other at lower frequencies in a phase dependent manner. The (quasi)persistent, self-generating character of spiral waves is particularly interesting in the context of persistent neural responses to transient stimuli. Huang et al. (2010) suggests that spiral waves in visual cortex may be involved in maintaining persistent activity from transient stimuli in the sensory context. A video of their minimal computational model, in which a persistent spiral wave is seeded by transient input, is http://www.jneurosci.org/content/suppl/2004/11/03/24.44.9897.DC1/model-_spiraldrift.mpg. The current hypothesis extends this idea to the context of persistent activity in oculomotor postural control (Aksay et al., 2001; Major and Tank, 2004).

3. FIXATIONAL EYE MOVEMENTS

Fixational eye movements can be quite different between species. Martinez-Conde and Macknik (2008) review comparative studies of FEM in different species, concluding that tremor appears to be the most phylogenetically conserved and fundamental component, consistent with a basic role for spiral wave dynamics in the generative process of FEM. Drift is also widespread, whilst microsaccades appear linked to the existence of foveated ocular architecture. Microsaccades are the most intensively researched component of FEM in humans. These fast relocations of the fixation point appear to play a similar role and manifest similar neural correlates as saccades more generally (Ko et al., 2010; Hafed and Krauzlis, 2012; Kagan and Hafed, 2013; Martinez-Conde et al., 2013; Otero-Millan et al., 2013; Poletti et al., 2013), but also show relations to drift (Engbert and Mergenthaler, 2006; Chen and Hafed, 2013). Microsaccades are relatively infrequent, occuring up to three times per second at most and usually less frequently, in an irregular but individually characteristic fashion (Engbert and Mergenthaler, 2006).

Most of fixation time (>90%) is occupied by a slow *drift* of fixation (Martinez-Conde et al., 2004; Martinez-Conde, 2006; Rolfs, 2009), as depicted in **Figures 3**, **4**. Upon this is superimposed a fast (peaked around 80–90 Hz), low amplitude (approx. one photoreceptor), aperiodic oscillation termed *tremor*. Tremor is usually within the noise range of the recording equipment (Martinez-Conde et al., 2004). As a result, less is known about tremor than other components, and tremor is not resolved in many FEM studies of drift and microsaccades.

3.1. DYNAMICAL CHARACTERISTICS OF FIXATIONAL EYE MOVEMENTS

3.1.1. Random walk modeling of FEM

On the basis of early studies (Cornsweet, 1956; Matin et al., 1970; Findlay, 1971), FEM have been widely held to exhibit the $1/f$ spectrum of classical Brownian noise. Random walk analysis examines the *mean squared displacement* of a diffusing "particle" (in this case the point of fixation) relative to time. In ideal Brownian motion, this relation is linear in time. This characterization has been adopted in some recent models of how vision may cope with (Pitkow et al., 2007; Burak et al., 2010), and indeed exploit (Kuang et al., 2012), jitter of the retinal image due to FEM.

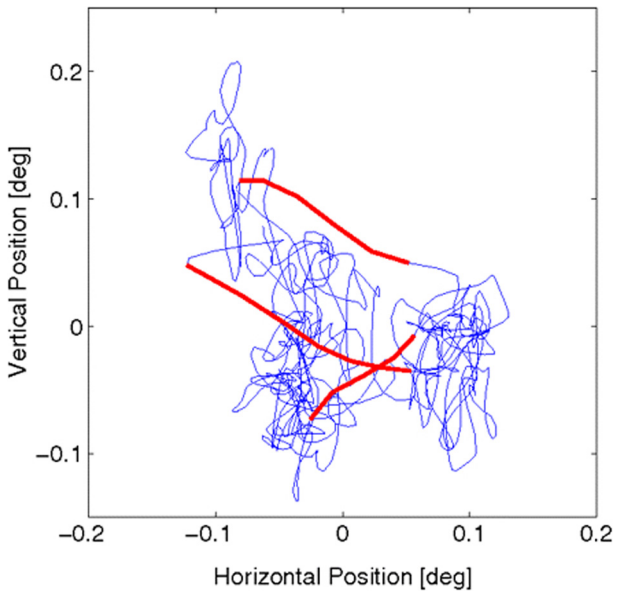


FIGURE 3 | Fixational eye movements and microsaccades, from Figure 1 in Engbert et al. (2011). Data were recorded from fixational eye movements during a fixation of 2 s. Slow movements (blue) are highly erratic, whereas microsaccades (red) are ballistic, small-amplitude epochs with a more linear trajectory (compared with the slow background motions). The sample trajectory was recorded with a sampling frequency of 500 Hz (for details see ref. 29 in Engbert et al., 2011). Reprinted from Engbert et al. (2011), copyright PNAS 2011.

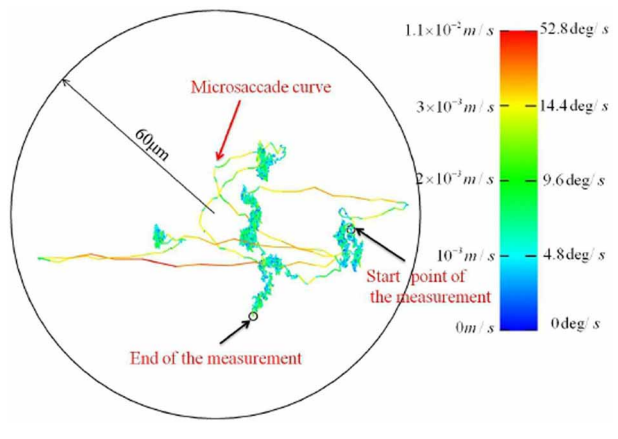


FIGURE 4 | FEM trajectories recorded with a new high resolution video tracking technique. (Adapted from Zhang and Li, 2012).

At the motor end, an important neural correlate of oculomotor postural control is *persistent neural activity* in brainstem regions including the prepositus hypoglossi ("PH") (Delgado-Garcia et al., 1989) and medial vestibular nucleus ("MVN") (Serafin et al., 1991; du Lac and Lisberger, 1995). Persistent here refers to sustained activity on timescales much longer than individual neural spiking timescales, in response to a relatively brief stimulation. Seung (1996) described a model of how persistent neural activity could be maintained through positive feedback, and showed how FEM drift-tremor could reflect a random walk along the line attractor created by the positive feedback dynamics in the motor memory of eye position. Seung suggested that various sources of noise, such as the random fluctuations in the tonic input from vestibular afferents, could be causing the random walk behavior.

Persistent neural activity is associated with short term memory more generally (Major and Tank, 2004), and various potential mechanisms for maintaining persistent activity have been investigated (see for review Brody et al., 2003). Stability issues arising from the positive feedback model were addressed in Koulakov et al. (2002); Goldman et al. (2003). More recently, empirical evidence contrary to the predictions of line attractor models (Aksay et al., 2007; Miri et al., 2011a) has motivated the proposal of modifications of recurrant network models, and the development of new models based on functionally feedforward networks and balanced regimes of excitation and inhibition (Goldman, 2009; Lim and Goldman, 2013). We address this topic in more detail in the following section.

3.1.2. Self-avoiding random walk models

Recent work has shown that FEM exhibit non-trivial temporal correlations whose description require fractional scaling exponents, rather than the unitary scaling exponent of pure Brownian motion. Engbert and Kliegl (2004); Mergenthaler and Engbert (2007) applied random walk analysis to the statistics of fixational eye movements at short (<40 ms) and long time (100–400 ms) scales. The tremor component was not resolved in these studies. At short timescales, the fixation point drifts faster than for normal diffusion/Brownian motion, with scaling exponents in the range 1.3–1.5. This is termed a "persistent" Codling et al. (2008) or "superdiffusive" Metzler and Klafter (2000) process. On longer timescales, however, the distance of the fixation point drifts slower than normal diffusion (termed subdiffusive/antipersistent). We adopt "sub/superdiffusive" here to avoid crossing terminology with that of persistent neural activity.

(Engbert et al., 2011) modeled this behavior in terms of a self-avoiding random walk ("SARW") in a potential. The potential accounts for the long term subdiffusivity, implementing a tendency for "wandering back" to the center of the potential well in the long term. The short term superdiffusivity is modeled by giving the random walk a memory, and adding a term such that the walk next visits the neighbor whose "previously visited" activation is lowest, with ties resolved randomly. As a

result, the walk tends to be structured by the initial direction which is chosen (randomly), as the path backwards is in general more visited. In this model, microsaccades are triggered by visiting a site with a "visited" activation above a threshold. Roberts et al. (2013) describe a similar self-avoiding random walk model, though without the confining potential and the microsaccade threshold, and with an imprecise continuous memory and rather than a lattice representation. Importantly, this study showed that both the superdiffusive drift in the FEM signal, and the self-avoiding random walk approach, generalize to more ecologically valid, dynamic viewing conditions (watching a film).

3.1.3. Recent findings may challenge some existing ideas about FEM

Recently, Zhang and Li (2012) reported technical innovations based on binocular video tracking of ocular vein structure, which enable non-invasive, high resolution imaging of FEM in three rotational degrees of freedom. Importantly, the optical imaging approach avoids the interaction of mechanical measurement devices with ocular tremor. The authors reported a previously unsuspected level of structure at high resolution. See **Figure 3** for a FEM trajectory where tremor is not resolved. **Figure 4** depicts a FEM trajectory as measured by this new technique. These findings are quite new and have not received much attention to date, at least in terms of citations. They likely merit more attention, because if they are reproducible, they offer a challenge to existing conceptions of FEM in a number of ways.

Firstly, the microsaccadic trajectories recorded in these high resolution data are not straight and ballistic, as is widely supposed Martinez-Conde et al. (2004); Martinez-Conde (2006); Rolfs (2009); Martinez-Conde et al. (2013). The microsaccadic trajectories observed by Zhang and Li (2012) were often curved, and could exhibit fine scale changes in both speed and direction. Secondly, and crucially to our argument here, Zhang and Li (2012) reported that drift can take the form of a curled line at a fine spatial scale wherein tremor is resolved. This observation of a trajectory which is habitually self-crossing at a small spatiotemporal scale is not predicted by self-avoiding random walk models of drift generation, as the process should be self-avoiding at small spatiotemporal scales. These findings are a significant part of the motivation for the current hypothesis. Unfortunately, however, we have as yet been unable to obtain the associated time series data. Speculativeness notwithstanding, we believe that the proposal of testable hypotheses is a positive way to motivate the publicization of data and structure further empirical investigations in this area. Another high resolution non-contact FEM measurement method which can resolve tremor has recently been reported by Kenny et al. (2013a,b), so data of sufficient resolution may soon become available from this group.

3.1.4. Tremor; a clue to the mechanisms of FEM, or "just noise"?

Tremor has often been dismissed as "noise," but then so have other aspects of FEM over the years. Whether tremor reflects unrelated background noise or the workings of the neural mechanisms which maintain the dynamical posture of the eye remains very much an open question. Spauschus et al. (1999) found strong binocular coherence of tremor, and concluded that tremor reflects the patterning of low-level drives to oculomotor neurons, rather than motor noise. More recently, a sophisticated method was employed by Thiel et al. (2008), who reported positive evidence for binocular phase synchronization. They concluded that there might be only one center in the brain that produces the fixational movements in both eyes, or a close link between the two centers. The loss or reduction of tremor in certain cases of brain pathology (Michalik, 1987) and in coma (Shakhnovich and Thomas, 1977) also gives reason to suspect a more important, and delicate, source of tremor.

4. MOTIVATIONS FOR THE SPIRAL WAVE HYPOTHESIS OF FEM DRIFT-TREMOR

4.1. STATISTICAL SIMILARITIES BETWEEN DRIFT-TREMOR TRAJECTORIES AND SPIRAL DYNAMICS

Tremor contains a strong spectral peak around 80–90 Hz and a less prominent, variable lower frequency component up to around 25 Hz (Spauschus et al., 1999). A low amplitude (approx. 1 photoreceptor), slow (around 5 Hz), tremor-like ocular oscillation has been recorded at very high resolution in the turtle (Greschner et al., 2002), possibly corresponding to the slow component of tremor in primates, though cross species comparisons must be made with caution (Martinez-Conde and Macknik, 2008). Greschner et al. (2002) reported this oscillation as periodic, though the flat peak in the frequency spectrum around 5 Hz, and the high variability visible in the inset example trajectories, may be suggestive of quasi-periodicity (see their Figure 1A). Either way, the regularity of this low frequency component is interesting because it suggests a certain systematicity to the generative mechanisms. Closer examination of human tremor is required to establish whether an identifiable carrier wave exists at lower frequencies, which is then heavily masked by the fast aperiodic component of tremor. An underlying spiral wave neurodynamics is consistent with both a periodic and a quasi-periodic form for this carrier wave (Barkley et al., 1990; Broer et al., 1996), and predicts the accompanying high frequency aperiodic oscillation (see **Figures 1**, **2**).

See **Figure 5**, reprinted from Figure 7 in Huang et al. (2010), for visual representations of spiral drift recorded in visual cortex. Like FEM drift (Cornsweet, 1956; Matin et al., 1970; Findlay, 1971), spiral drift can exhibit Brownian statistical structure, due to both external forcing/environmental gradients (Sendiña-Nadal et al., 2000; Yuan et al., 2011) and intrinsic dynamics (Biktashev and Holden, 1998). Such gradients could (but need not) reflect the path memory and/or confining potentials in SARW models (Engbert et al., 2011), and perhaps even visual context (Mensh et al., 2004; Chan and Galiana, 2005). Like FEM drift, the velocity of spiral drift in neocortex is variable (Huang et al., 2010), and was found to be higher in induced sleep-like states (see **Figure 5**), consistent with recent observations that time-on-task increases the speed of FEM drift, whilst reducing the peak velocity of microsaccades (Di Stasi et al., 2013).

Prediction The singularity hypothesis predicts that, when the small spatial scale of tremor is resolved, drift-tremor trajectories will take a curling, self-crossing form reflecting the

rotational orbit of a spiral wave. This might also be complex due to interactions between similar systems controlling different degrees of freedom. The fast component of tremor will manifest as an aperiodic modulation of this carrier wave. The presence of a slow (10–20 Hz) rotational component, giving the trajectory a habitually *self-crossing* form at small scales, would distinguish the spiral wave model from self-avoiding random walk models of FEM drift.

4.2. SPIRAL WAVES AND THE NEURODYNAMICS OF OCULOMOTOR POSTURAL CONTROL

4.2.1. Persistent neural activity

Persistent neural activity ("PNA") refers to localized "bumps" of fast firing cells which persist over timescales much longer then the timescales of the individual neurons comprising the bump. PNA is an important neural correlate of working memory (Major and Tank, 2004), and has been studied intensively in the context of brainstem *neural integrators* which encode eye position during oculomotor control (e.g., Aksay et al., 2001, 2007; de Dios Navarro-López et al., 2004; Miri et al., 2011a). Fluctuations in oculomotor PNA are thought to underly FEM drift and tremor, but the precise mechanisms responsible for their generation remain a matter of considerable debate. Current models regarding these mechanisms have been categorized on the basis of whether they posit intrinsic unicellular mechanisms (e.g., Shen, 1989; Loewenstein and Sompolinsky, 2003; Teramae and Fukai, 2005) or network mechanisms (Cannon and Robinson, 1985; Seung, 1996; Seung et al., 2000; Goldman et al., 2003; Goldman, 2009). Some role for network mechanisms is suggested by evidence of correlated activity between cells in the oculomotor integrator network (Aksay et al., 2003), and by the observed covariance of eye position and the frequency and magnitude of the synaptic barrage converging on integrator cells (Huang, 2009). These alternatives are not necessarily mutually exclusive, and multiple mechanisms may operate in the maintainance of PNA in the oculomotor system and elsewhere in the brain (Major and Tank, 2004).

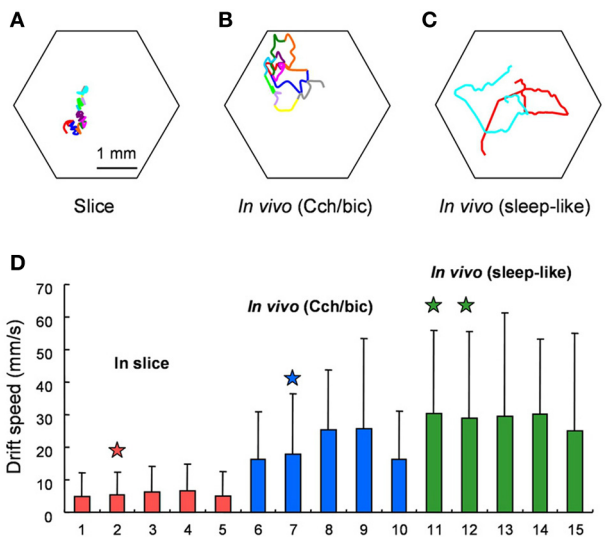


FIGURE 5 | Behavior of spiral waves in mammalian visual cortex under various conditions. From the original; Drifting of Spiral Phase Singularities. **(A)** Trajectory of spiral phase singularity during a 12-cycle spiral waves in cortical slices. **(B)** Trajectory of spiral phase singularity during an 11-cycle spiral waves *in vivo* under Cch/bic application. Hexagon shows the field of view and each color represents one cycle of spiral wave. **(C)** Trajectory of spiral phase singularities during 2 spiral waves (red and cyan, each with 1.5 turn) during sleep-like states. **(D)** Comparison of drifting speed of spiral phase singularity for slices and *in vivo*. Five examples from *in vivo* under Cch/bic, *in vivo* during sleep-like states and slices, respectively, are shown (mean + *SD*). Columns with stars on top are from the examples in **(A,C)**. The standard deviation is large because the drifting of spiral phase singularity is not consistent and there are large variations from time to time. The difference between *in vivo* and slices is statistically significant (Welch's test, $p < 0.001$, 25 *t* tests). The difference between *in vivo* (Cch/bic) and *in vivo* (sleep-like) is also significant. Reprinted from Figure 7 in Huang et al. (2010), copyright Elsevier 2010.

The vestibulo-oculomotor system exhibits fractional dynamics (Anastasio, 1994), and complex time variation in PNA has motivated arguments that a model with multiple timescales of persistent firing may be required (Anastasio, 1998). Indeed, recent evidence for multiple timescales of persistence in oculomotor PNA (Miri et al., 2011a) suggests a higher dimensional attractor dynamics than proposed by earlier line attractor models (Seung, 1996; Seung et al., 2000; Goldman et al., 2003), leading to the development of new models with more complex dynamics (Miri et al., 2011a; Fisher et al., 2013). (Goldman, 2009) describes a functionally feed-forward architecture which reproduces some of the time variation in PNA, showing that positive feedback is not essential in principle, while Lim and Goldman (2013) presents a model based on homeostatic mechanisms which maintain a careful balance of excitation and inhibition. In human psychophysical studies, Khojasteh et al. (2012) found that cross subject averaging hides idiosyncratic nonlinear patterns. All this suggests that considerable complexity inhabits the dynamics of PNA in the oculomotor system (Durstewitz and Seamans, 2006).

The current hypothesis suggests an addition to the repertoire of hypothesized mechanisms for PNA, which falls into the category of network mechanisms, though is distinct from existing network models in a number of ways. Unlike existing network models, which concentrate on modeling neural behavior at the level of firing rates and drift, we focus on the finer spatiotemporal scale of FEM tremor and subthreshold fluctuations in the membrane potential of cells mediating oculomotor integration. Rather than specific circuit design, persistence is based on the transient self-organization of population activity into a reentrant, (quasi)periodic spatiotemporal pattern. Spiral waves require a predominance of excitatory, spatially distributed connections but precise connectivity structure is not required; spirals can easily emerge in randomly connected networks (Milton et al., 1993; Chu et al., 1994; Yuan et al., 2011). This is not to say that specific circuitry is not important or present in the hVPNI; just that it is not a requirement of the current model. The existence of spatially organized, excitatory lateral connectivity is suggested by various studies (Aksay et al., 2001, 2007; Miri et al., 2011a). Disinhibition is crucial to the formation of spirals (and other traveling waves) in cortical tissue (Huang et al., 2004, 2010). (Aksay et al., 2007) identified that mutually inhibitory collateral interactions were not necessary to local integrators within a certain range, suggesting

that these mutually inhibitory interactions regulated local mechanisms rather than driving PNA directly. Such a situation could relate to mutual modulation of the excitability required for the emergence and propagation of traveling waves.

4.2.2. Functional properties of spiral waves

Certain features of spiral waves make them potentially interesting as neurodynamical mediators of PNA. Indeed, spiral waves are a form of neural activity which is persistent (e.g., Milton et al., 1993; Chu et al., 1994), though they are not usually associated with the term as used in the context of short term memory and neural integrators. The combination of A, B, C and D below suggests a mechanism capable of contributing to the maintenance of localized persistent neural activity.

A. Spatiotemporally organized depolarization: Sub-threshold traveling waves in the membrane potential field coordinate network activity in space and time, by defining spatiotemporal regimes of polarization-depolarization (Wu et al., 2008; Huang et al., 2010; Bahramisharif et al., 2013). Cells in the vicinity of the singularity can take arbitrarily differing phase, leading to an almost phaseless depolarizing synaptic barrage in that vicinity.

B. Pseudo-locality: Spiral waves exhibit a duality which gives them both a local, particle-like description (the singularity) and a global wave-like description (the propagating spiral arms) (Biktasheva and Biktashev, 2003). Though the wave is extensive, its behavior is almost entirely based on what happens in the neighborhood of the singularity.

C. Quasi-persistence: Spirals are reentrant waves, whose activity generates the conditions for their own persistence in time (Winfree, 1991), subject to certain conditions (e.g., Ito and Glass, 1991; Fenton et al., 2002; Zhang et al., 2003; Chun-Ni et al., 2010; Ma et al., 2010, 2012a). C combined with A and B, enables a spiral wave to *persistently* depolarize a *spatially localized* region in the neighborhood of the singularity.

D. Seedability: Spiral waves can be induced in an appropriate medium by various methods (Aranson et al., 1994; Williams and Holland, 1999; Leanhardt et al., 2002; Zhang et al., 2002; Xiao-Ping et al., 2011; Yuan et al., 2011; Ma et al., 2012b).

The most effective and tunable method is probably to directly impose an external forcing spiral, as in (Xiao-Ping et al., 2011), or a spiral seed plus a periodic forcing current near the singularity (Zhang et al., 2002). However, persistent spirals can also be induced in networks of integrate-and fire neurons by a brief, non-spiral periodic forcing (Milton et al., 1993; Chu et al., 1994; Huang et al., 2004; Kilpatrick and Bressloff, 2010b; Yuan et al., 2011) given some reasonable connectivity conditions (chiefly spatiality and some kind of inhomogeneity/noise/perturbation which breaks rotational symmetry).

D combined with A, B and C, provides a mechanism whereby an afferent may seed a spatial pattern in an efferent, and then leave that pattern to sustain itself with a certain amount of autonomy. Durstewitz and Deco (2008); Friston et al. (2012) suggest that brain activity is characterized by a high dimension chaotic background state, from which lower dimensional metastable states transiently emerge. **Figure 6**, from (Zhang et al., 2002) nicely visualizes the notion of how a spiral wave seeding might realize such a transient dimensionality reduction in the context of the observed (Aksay et al., 2001) difference between a background "off" state of the integrator, characterized by irregular firing at low rates (the turbulent background), and an "on" state characterized by driving input from the seeding of a spiral wave. de Dios Navarro-López et al. (2004) induced PNA in an oculomotor integrator circuit with brief, cholinergic periodic forcings. Oscillatory neurons observed in the guinea pig nucleus prepositus hypoglossi (Idoux et al., 2006), a region deeply associated with oculomotor integration (Delgado-Garcia et al., 1989; McCrea and Horn, 2006), might exemplify a neural substrate for periodic forcing inputs and the maintainance of traveling wave activity in the population. The persistent spiking of neurons in the vicinity of the induced singularity will be facilitated due to constant depolarization of the cellular membrane by high frequency microstimulation without a strong phasic component, as cells near the singularity may take arbitrarily different phase.

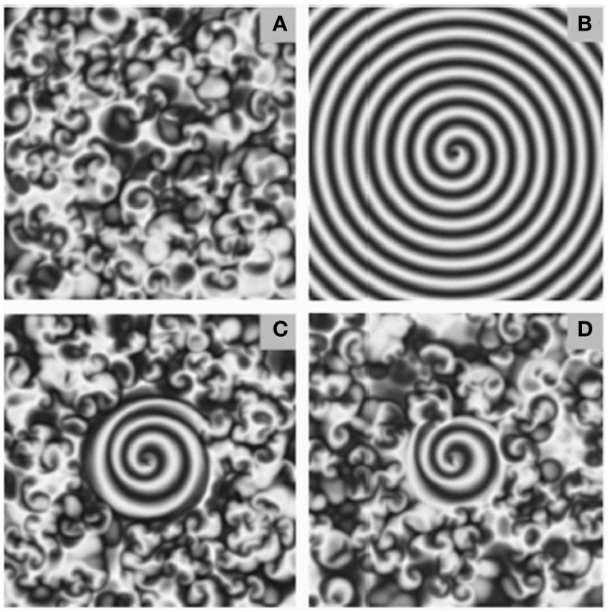


FIGURE 6 | Examples of spiral seeding in a background of chaotic turbulence. Seeding and growing spiral waves in a background of chaotic turbulence. The extent of the spiral wave seeded was dependent on the frequency of periodic forcing at the singularity. Too low, or too high, frequency was less effective. The ratio of the frequency of the forcing signal to that of the angular frequency of the spiral equalled **(A)** 0.6 **(B)** 0.8 **(C)** 1.0 **(D)** 1.2. Figure reprinted from Zhang et al. (2002). Copyright American Physical Society 2002.

The bump of persistent firing activity in PNA has naturally been associated with "bumps" in neural field models (e.g., Tegnér et al., 2002; Owen et al., 2007; Kilpatrick and Ermentrout, 2013). However, it has also been suggested that hVPNI neurons may operate in a fluctuation dominated regime, in part because the firing threshold of the cell *increases with the membrane potential*, and in part because firing always occurs at the apex of membrane potential fluctuations (Huang, 2009). In a fluctuation dominated regime, firing rate is dependent less on mean membrane potential than on fast fluctuations in the level of the depolarizing synaptic barrage. On this view, at the level of subthreshold field dynamics

fast microfluctuation (from a spiral wave) could be more effective than a constant raise in stimulation (from a bump) in inducing persistent firing. Thus spiral waves possess a unique repertoire of functional properties which render them interesting in the context of PNA. However, the presence or absence of a spiral wave only yields a binary distinction between an "on" state and an "off" state. Still missing is a mechanism for a graded temporal memory capable of remembering multiple, arbitrary step changes.

4.2.3. Possible mechanisms for continuous temporal integration

Persistent neural activity is associated with *neural integration* (Major and Tank, 2004), and has been intensively studied in the anatomical context of oculomotor postural control, in particular the *horizontal velocity to position integrator* ("hVPNI") (Aksay et al., 2000). Here, a cell integrates (in the mathematical sense) its inputs over time, providing the ability to hold, and externally nudge, set points. An external nudge is reflected in a persistent shift in average membrane potential, co-occuring with step changes in firing rate and eye position (Aksay et al., 2001). How might this graded integration functionality be implemented by a spiral wave depolarization regime?

One possibility here is that spiral waves of different spatial extent generate different levels of depolarizing input. **Figure 6**, reprinted from Zhang et al. (2002) depicts how different intensities of forcing current can generate different sizes of spiral wave, from a pre-existing background of low level, chaotic turbulence. Increasing the size of the spiral by adding energy could perhaps encode up steps, but it is less obvious how a down step would be implemented. Another possibility is inducing multiple spirals.

Functionally, however, modulation of the frequency of the spiral's rotational orbit could provide the most appropriate variable for graded temporal integration. The excitability of the medium has a strong determining effect on the frequency taken by spirals and other traveling waves (Winfree, 1991). Modulation of the strength of lateral connections might therefore provide a mechanism to induce persistent changes in spiral frequency. Calcium mediated presynaptic facilitation (Mongillo et al., 2008) could provide a mechanism for strengthening lateral connectivity, and other modes of disinhibition could also be relevant (e.g., de Dios Navarro-López et al., 2004). Kilpatrick and Bressloff (2010a,b) describe spirals in neural field models, in which spike frequency adaptation modulates the frequency of network oscillations. Whatever the mechanism of frequency modulation, a faster spiral would generate more action potentials per unit time in the synaptic barrage converging on the cell from weak but numerous lateral connections, maintaining membrane depolarization, and the magnitude of these would be amplified by lateral synaptic facilitation. Both the magnitude and the arrival frequency of depolarizing excitatory postsynaptic potentials converging on active eye position coding cells varies systematically with eye position Aksay et al. (2001); Huang (2009).

Secondary Prediction A spiral frequency based temporal integrator is consistent with the close covariance of the arrival frequency of action potentials with eye position (Huang, 2009), and would predict in addition that a spectral peak in the slower range (around 10–20 Hz) of membrane potential oscillations during PNA will vary systematically with eye position in *individual trial* data (averaging might hide this effect). Note that the frequency modulation approach to graded integration is a secondary hypothesis.

4.2.4. Subthreshold dynamics of the membrane potential during PNA in an oculomotor integrator

Aksay et al. (2001) carried out *in vivo* intracellular recording and perturbation of persistent activity in an oculomotor neural integrator. They tracked the evolution of the cellular membrane potential during step changes in persistent activity associated with position control during fixation events. Their Figure 2 is reprinted here as **Figure 7**. See also their Figure 1 for longer recording period. Note the step like change in membrane potential that accompanies the onset of persistent firing. Further step changes are marked by brief (50–100 ms) overshoot/undershoot depending on direction of change, followed by a persistent change in the mean membrane potential and firing rate. The membrane depolarization was found to be sufficient to explain the associated PNA in a control experiment, suggesting an important role for network mechanisms.

Up (down) steps in membrane potential are correlated with increases (decreases) in both arrival frequency and magnitude of *excitatory postsynaptic potentials* ("EPSPs") (Aksay et al., 2001; Huang, 2009). The membrane potential (V_{is})and the firing rate (F_{intra}) shows signs of an oscillation at around 15 Hz, perhaps corresponding to the rotational orbit of a spiral wave and the slow component of tremor. If the barrage of depolarizing EPSPs are indeed originating from the slow rotational orbit and fast jitter of a spiral wave, then under close examination one would expect to see the leading edge of the spiral waveform reflected in the EPSPs. Huang (2009) examined membrane potential fluctuations during PNA in great detail. Whole-cell patch recordings revealed the existence of many small (0.2–3 mV) excitatory postsynaptic potentials lasting 5–10 ms, and manifesting a "peculiar" sharp-attack, slow-decay form obscured in accompanying sharp electrode recordings. See their section 5 and Figure 5.3, reproduced here as **Figure 8**. This waveform is typical of that generated by the passing of the leading edge of a spiral wave, where the phase gradient is very high (providing the sharp attack). See for example the depictions from Qu et al. (1999) of action potentials caused by cardiac spiral waves http://ajpheart.physiology.org/content/ajpheart/276/1/H269/F8.large.jpg. Note though that other possibilities exist. Based on similar waveforms observed in Mauthner cells (Golding and Spruston, 1998; Korn and Faber, 2005), Huang (2009) suggests that mixed NMDA/AMPA conductances could underly the shape of these potentials.

Prediction At the network level, the singularity hypothesis predicts that spiral waves should be directly identifiable in the VSDI signal at the site of PNA (at least in the anatomical context of brainstem oculomotor integration), and that the success (failure) to induce a spiral wave will distinguish success (failure) to induce PNA. While it is possible in principle

that the membrane potential dynamics just reviewed could reflect a spiral wave which exists elsewhere, neurophysiological evidence suggests that the mechanisms sustaining PNA in the hVPNI are local (Aksay et al., 2007).

A standard test for whether an observation corresponds to a true spiral wave is the existence of a phase singularity, with a local amplitude reduction of field oscillations in the vicinity of the singularity (Winfree, 1991, 2001; Huang et al., 2010), caused by the cancellation of signals from closely located nodes with opposing phase near the singularity. See Figure 1A in Huang et al. (2010), reprinted here as **Figure 9** for neurophysiological recordings of this phenomena. Small fluctations in the balance of this cancellation result in an increase in high frequency fluctuations visible in the VSDI signal in **Figure 9** (and also visible as jitter of the singularity in our **Figure 1**), which could drive a cell effectively in a fluctuation dominated regime (Huang, 2009).

We are not aware of VSDI studies of brainstem neural integrators. Recent methodological advances may combine to provide opportunities for imaging the spatiotemporal dynamics of subthreshold activity in the deep brainstem. Fiber optics offer a means to image non-superficial regions (Flusberg et al., 2005). Combining VSDI and laser scanning microstimulation offers a fast method for anatomical and functional mapping (Xu et al., 2010). Zebra fish larvae have recently been shown to provide an *in vivo* preparation with high optical transparency (Miri et al., 2011a; Fisher et al., 2013). Miri et al. (2011b) used two-photon laser scanning microscopy (Stosiek et al., 2003) to simultaneously image many cells in neural integrator circuits in the larval zebra fish, and introduced a semi-automated approach for identifying behavior measure (in this case eye movement) related cells in the ensuing space-time series.

Combining this methodology with VSDI, which would provide access to subthreshold spatiotemporal dynamics associated with PNA, could test directly whether spiral waves exist and if so, whether they are spatially associated with active eye position integrators and whether their rotational frequency (and/or spatial extent) covaries systematically with eye position. Regardless of whether these predictions are confirmed or denied, VSDI data would likely to be of great utility in the general research effort on the hVPNI.

4.3. SACCADES AND MICROSACCADES

The current contribution focuses on the dynamic maintainance of oculomotor posture between microsaccades. Nonetheless microsaccades, and indeed saccades in general, are naturally relevant to the discussion as a whole. In this section, we briefly address how a spiral wave model of drift-tremor might fit into its saccadic context.

4.3.1. Scale free saccadic behaviors

Recent evidence points to a remarkable continuity in the statistics of saccadic oculomotor control across scales including microsaccades (Otero-Millan et al., 2013). How might a spiral wave model of fixational postural control fit into its containing context of fast (micro and macro) saccadic gaze shifts? Is there a continuously

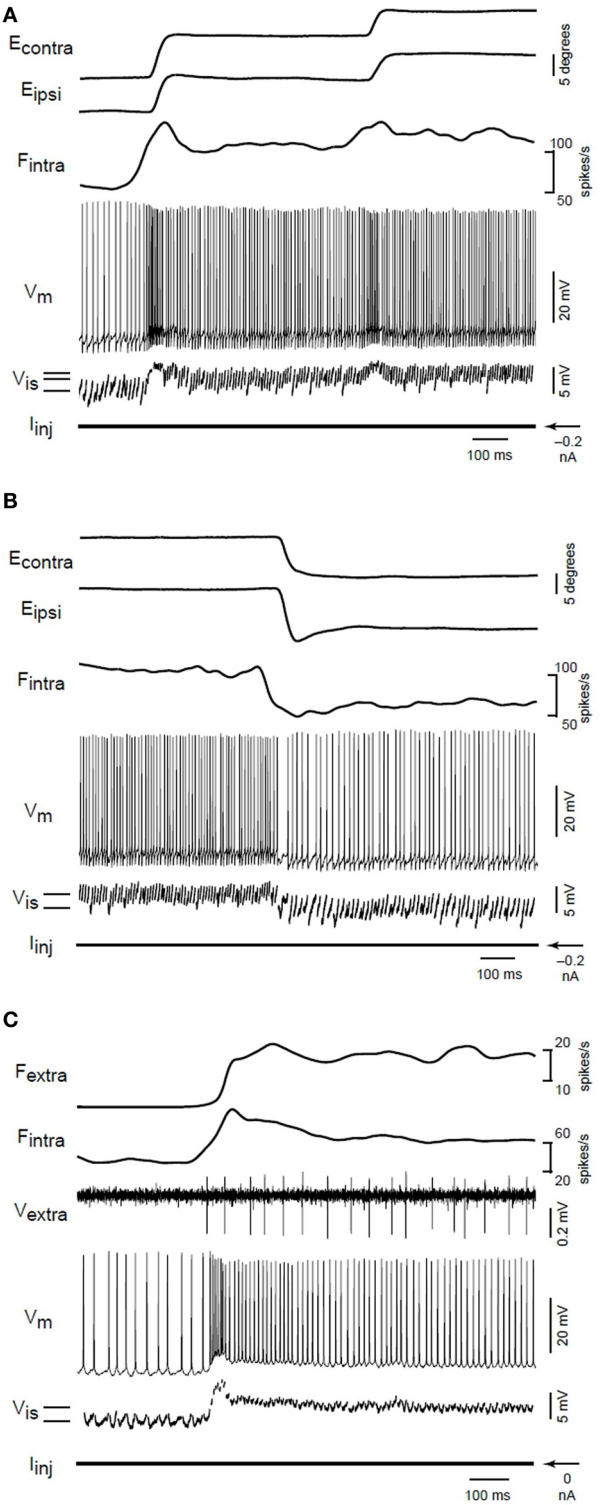


FIGURE 7 | From the original; Membrane potential changes during transitions in fixation position. (A) Eye position, firing rate of an intracellularly recorded neuron (F_{intra}), membrane potential (V_m), interspike membrane potential (V_{is}), and injected current (I_{inj}) for on-direction steps during intracellular recording. Solid lines at lower left indicate the average value of V_{is} during separate fixations. **(B)** Changes during off-direction steps *(Continued)*

FIGURE 7 | Continued
of the ipsilateral eye. This segment is taken from the end of a nasaltemporalnasal cycle that started with the transitions shown in **(A)**. **(C)** Rate and potential changes for a different neuron during an on-direction step. The firing rate (F_{extra}) of a second extracellularly recorded position neuron (V_{extra}) served as a surrogate for eye position. In this recording, the fast afterhyperpolarization following action potentials was abolished by substituting cesium for potassium in the electrode solution. From Aksay et al. (2001), copyright Nature Publishing 2001.

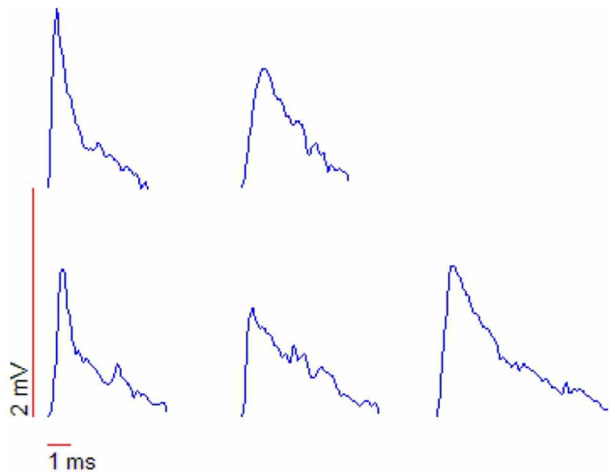


FIGURE 8 | Shapes of action potentials recorded with whole-cell patch measurements. Note the "peculiar" triangular form, typical of the leading edge of a spiral waveform. From Huang (2009), copyright The Author.

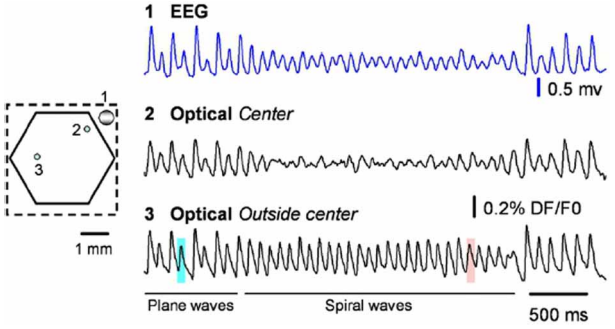


FIGURE 9 | From the original; raw data traces from the EEG (blue) and two optical detectors labeled on the left. Detector 2 (Optical spiral center) is selected from the spiral center, showing large amplitude reduction. Detector 3 is selected from a location that spiral center never swept through, showing no amplitude reduction. From Huang et al. (2010), copyright Elsevier 2010.

scaling control principle which could give rise to a visual scan path with, from the bottom up;

A. microfixations characterized by spiral dynamics (drift-tremor), which are interspersed with
B. relatively long, straight and fast flights (microsaccades), which are organized into
C. clusters of microfixations (macrofixations), which are interspersed with
D. relatively very long, straight, fast flights (macrosaccades), which in turn cluster into
E. regions of dense exploration and short saccades, interspersed by long saccades to new regions of interest?

The natural variability of human scanning patterns has been modeled at the macro-saccadic level by the imposition of a stochastic component comprising a Lévy walk (Klafter et al., 1987) upon scanpaths in a deterministic salience landscape (Brockmann and Geisel, 1999; Boccignone and Ferraro, 2012). Another possibility is that the stochastic component reflects an intrinsic probabilistic feature of the salience function, (e.g., Harel et al., 2006), rather than an imposed randomization.

4.3.2. Lévy walks in rotational and turbulent flow

Theoretical work has revealed deep links between spiral waves, turbulent flow, fractional Brownian motion, anomalous diffusion and Lévy type trajectories (Shlesinger et al., 1987; Viecelli, 1990; Metzler and Klafter, 2000). Solomon et al. (1993) observed Lévy walks of tracer particles in a physical system of effectively two dimensional rotating flow, and Solomon et al. (1994) examined in more detail behavior in periodic, chaotic and turbulent conditions. See our **Figure 10** for a reprint of Figure 6 in Solomon et al. (1994), depicting example trajectories. Long term trajectories exhibited a pattern of long, relatively direct flights in predominantly translational flow, interspersed with episodes where the particle is caught up in a spiraling curve due to capture by a vortex ("sticking"). Analysis of sticking times and flight statistics indicated a Lévy walk trajectory evolving in continuous time. Biomechanical constraints suggest that a *truncated* Lévy walk Mantegna and Stanley (1994), where maximum step lengths are finite, may be more appropriate to the biological case.

Similar dynamics in a more complex landscape consisting of multiple clusters of vortices could conceivably result in a trajectory resembling that of multiscale visual exploration. "Fixation" periods consist of clusters of mini-fixations, each of which consists of an episode of vortex sticking characterized by rotational flow. Escaping a cluster results in a relatively long step to the next cluster (i.e., a macrosaccade), followed by a sequential sampling of the new cluster. The traveling wave accompanying saccadic execution observed in the superior colliculus by Munoz et al. (1991) might be a manifestation of relatively long flights between vortex sticking visible in **Figure 10**. The curling trajectory of drift-tremor during microfixation and the curving, interrupted microsaccade trajectories reported by Zhang and Li (2012) are reminiscent of the turbulent transport scenario just outlined, though closer analysis of these data is required. Note how the microsaccadic trajectories in **Figure 4** do not always start their trajectory in the direction of their final destination. They are often curved and can have small scale variations in velocity. Looking closely at the microsaccades depicted, one may observe vertical motions exhibiting an oscillation which is damped in one horizontal dimension and amplified in the other, suggestive of transport in a potential field. Where they do travel in straight lines, this is usually on the horizontal axis, suggestive of

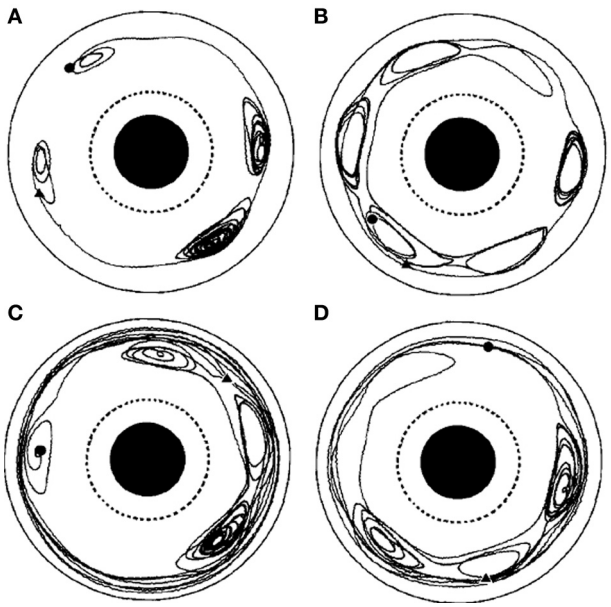


FIGURE 10 | From the original; Chaotic particle trajectories in a time-periodic flow. (A–D) Depict different runs of the apparatus with different numbers and locations of vortices. From the original; Chaotic particle trajectories in a time-periodic flow. Long sticking events can be seen in each case, and flights of length greater than one rotation about the annulus can be seen in **(C)**, **(D)**. Hyperbolic fixed points, near which the particle motion is particularly sensitive to transitions between flights and sticking events, are evident in all the trajectories. The particle motion is viewed from a reference frame that is co-rotating with the vortex chain, and the beginning of each trajectory is marked by a circle, the end by a triangle. From Solomon et al. (1994), copyright American Physical Society 1994.

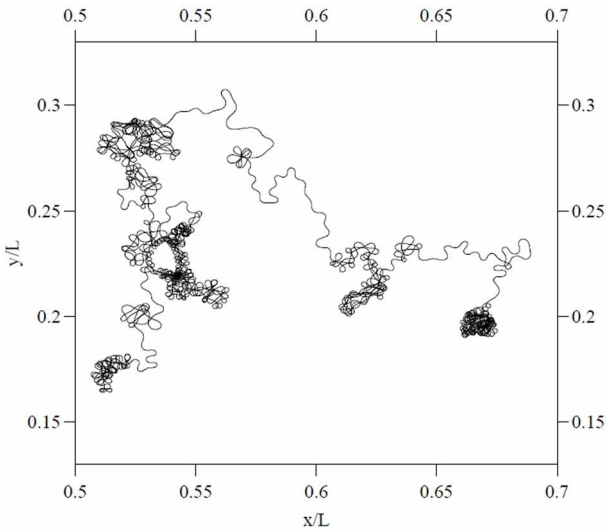


FIGURE 11 | Particle trajectory in a model of two dimensional magnetic turbulent flow. From Chiaravalloti et al. (2006), copyright the Royal Swedish Academy of Sciences 2006. Reproduced by permission of IOP publishing.

a potential field. Overall the pattern is not ballistic, but is consistent with the trajectory first escaping local vortex sticking in an unpredictable direction, followed by a trajectory dominated by translational flow with potential fields. **Figure 11** depicts an example of a trajectory in a model of anomalous transport in magnetic field turbulence, taken from Chiaravalloti et al. (2006). This provides an example, though from a different domain, of how these dynamics could generate trajectories similar to the FEM trajectories depicted in **Figure 4**. Adding a potential field might further approximate the FEM trajectories in **Figure 4**.

4.4. TURBULENT TRANSPORT AND CHAOTIC ITINERANCY

Thus there is some potential for a continuous, deterministic dynamical principle capable of generating the scale free, stochastic profile of visual scanning trajectories. This speculative proposed framework for multiscale visual exploration would imply a widespread role for traveling waves and rotational flow in brain-body hermeneutics, which may stretch the readers credulity at this stage, but there is some existing context. Breakspear (2001); Tyukin et al. (2009); Friston et al. (2012) examine traveling wave processing and self-organized instability in perception, whilst Heitmann (2013) explores traveling wave functionality in the motor context. In addition to noting a potential contribution to persistent neural activity in the sensory context, Huang et al. (2010) suggests that spiral waves, as a locally generated event, may also help a local cortical circuit to quickly disengage from globally synchronized rhythms. If traveling waves are playing functional roles in brain activity, one role of rotational flow may be to "hold the posture" of the central nervous system, while translational flow interconnects metastable postural transients. If so, this should be reflected in fixational drift-tremor and saccades, because the eye is part of the CNS. Breakspear et al. (2010) suggest that traveling wave solutions may offer optimal solutions to minimization of the free-energy in far from equilibrium initial conditions. Free energy minimization may be a rather general heuristic in nervous function (Friston, 2010). If the predictions of the singularity hypothesis of FEM drift-tremor turn out to be accurate, then the case for transient population dynamics as optimizers of behavior would gain a considerable boost.

It is interesting to speculate that attention may be related to nervous mechanisms of suppressing spiral waves. On this view, the sequential visiting of spirals in a cluster would destabilize the spirals and cause their breakup, resulting in a collateral effect resembling inhibition of return. This kind of self-destabilizing, itinerant trajectory would link action and perception into a common framework probably best described in terms of existing work on (embodied) *chaotic itinerancy* (Tsuda, 1991, 2009; Kaneko, 1992; Kaneko and Tsuda, 2003; Ikegami, 2007). Transient dynamics traversing a landscape of attractor ruins with riddled basins (Milnor, 1985) (i.e., quasi-attractors whose basin of attraction is riddled with repellent trajectories belonging to the basin of another attractor) can perform perceptual (Breakspear, 2001; Tyukin et al., 2009) and memory (Rabinovich et al., 2001) functions. This raises the possibility that the transient dynamics of embodied eye movements could play a rather sophisticated perceptual role analagous in computational description to that of neural sensory mechanisms (Tyukin et al., 2009), but at the

embodied level where the perturbation structure of the world constitutes the data set (Ikegami, 2007). Wilkinson et al. (2011) give a simple computational example of how exploratory gaze patterns structured by spiral waves can enact perception of a global property of a social scene (co-orientation), as has been observed in infants (Augusti et al., 2010; Handl et al., 2013).

5. CONCLUSION

Despite extensive study of oculomotor postural control, the generative mechanisms of fixational drift and tremor remain uncertain. We have proposed the hypothesis that these components reflect the drift and meander of spiral wave neurodynamics. Whilst speculative, the singularity hypothesis offers a parsimonious and predictive account of FEM. Though our motivations are chiefly functional, the available psychophysical and neurophysiological evidence is largely consistent with, and occasionally suggestive of, a contribution of rotational flow to the maintainence of persistent neural activity in the oculomotor system. We have laid out an argument motivating our hypothesis in terms of the existing literature, and made testable predictions which could falsify it. Our hope is that these predictions will encourage other groups working in the fields of FEM and oculomotor integration to consider looking for indicators of rotational flow when analysing data, and perhaps even motivate neuroimaging studies to examine the subthreshold spatiotemporal dynamics associated with PNA. Should empirical studies and/or analysis of existing data confirm the basic predictions of the conceptual model, future work should undertake detailed computational modeling. Further testing and development may offer a deterministic account of the stochasticity and self-similarity manifest in eye movement patterns across scales, based on the complex dynamics of anomalous transport in rotating neural flow.

ACKNOWLEDGMENTS

This research was supported by the EU project RobotDoc under 25065 from the 7th Framework Programme, Marie Curie Action ITN. We thank the reviewers for motivating significant improvements in the paper.

REFERENCES

Ahissar, E., and Arieli, A. (2001). Figuring space by time. *Neuron* 32, 185–201. doi: 10.1016/S0896-6273(01)00466-4

Ahissar, E., and Arieli, A. (2012). Seeing via miniature eye movements: a dynamic hypothesis for vision. *Front. Comput. Neurosci.* 6:89. doi: 10.3389/fncom.2012.00089

Aksay, E., Baker, R., Seung, H., and Tank, D. (2000). Anatomy and discharge properties of pre-motor neurons in the goldfish medulla that have eye-position signals during fixations. *J. Neurophysiol.* 84, 1035–1049.

Aksay, E., Baker, R., Seung, H. S., and Tank, D. W. (2003). Correlated discharge among cell pairs within the oculomotor horizontal velocity-to-position integrator. *J. Neurosci.* 23, 10852–10858.

Aksay, E., Gamkrelidze, G., Seung, H., Baker, R., and Tank, D. (2001). *In vivo* intracellular recording and perturbation of persistent activity in a neural integrator. *Nat. Neurosci.* 4, 184–193. doi: 10.1038/84023

Aksay, E., Olasagasti, I., Mensh, B. D., Baker, R., Goldman, M. S., and Tank, D. W. (2007). Functional dissection of circuitry in a neural integrator. *Nat. Neurosci.* 10, 494–504. doi: 10.1038/nn1877

Amit, D. J. (1992). *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge: Cambridge University Press.

Anastasio, T. J. (1994). The fractional-order dynamics of brainstem vestibulo-oculomotor neurons. *Biol. Cybern.* 72, 69–79. doi: 10.1007/BF00206239

Anastasio, T. J. (1998). Nonuniformity in the linear network model of the oculomotor integrator produces approximately fractional-order dynamics and more realistic neuron behavior. *Biol. Cybern.* 79, 377–391. doi: 10.1007/s004220050487

Aranson, I., Levine, H., and Tsimring, L. (1994). Controlling spatiotemporal chaos. *Phys. Rev. Lett.* 72:2561. doi: 10.1103/PhysRevLett.72.2561

Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Augusti, E.-M., Melinder, A., and Gredebäck, G. (2010). Look who's talking: preverbal infants perception of face-to-face and back-to-back social interactions. *Front. Psychol.* 1:161. doi: 10.3389/fpsyg.2010.00161

Bahramisharif, A., van Gerven, M. A., Aarnoutse, E. J., Mercier, M. R., Schwartz, T. H., Foxe, J. J., et al. (2013). Propagating neocortical gamma bursts are coordinated by traveling alpha waves. *J. Neurosci.* 33, 18849–18854. doi: 10.1523/JNEUROSCI.2455-13.2013

Bak, P. (1996). How nature works: the science of self-organized criticality. *Nature* 383, 772–773.

Bak, P., Tang, C., and Wiesenfeld, K. (1987). Self-organized criticality: an explanation of 1/f noise. *Phys. Rev. Lett.* 59, 381–384. doi: 10.1103/PhysRevLett.59.381

Barkley, D., Kness, M., and Tuckerman, L. S. (1990). Spiral-wave dynamics in a simple model of excitable media: the transition from simple to compound rotation. *Phys. Rev. A* 42, 2489–2492. doi: 10.1103/PhysRevA.42.2489

Biktashev, V., and Holden, A. (1993). Resonant drift of an autowave vortex in a bounded medium. *Phys. Lett. A* 181, 216–224. doi: 10.1016/0375-9601(93)90642-D

Biktashev, V., and Holden, A. (1995). Resonant drift of autowave vortices in two dimensions and the effects of boundaries and inhomogeneities. *Chaos Solit. Fract.* 5, 575–622. doi: 10.1016/0960-0779(93)E0044-C

Biktashev, V., and Holden, A. (1998). Deterministic brownian motion in the hypermeander of spiral waves. *Physica D* 116, 342–354. doi: 10.1016/S0167-2789(97)00304-7

Biktashev, V. N. (2007). Drift of spiral waves. *Scholarpedia* 2:1836. doi: 10.4249/scholarpedia.1836

Biktasheva, I., and Biktashev, V. (2003). Wave-particle dualism of spiral waves dynamics. *Phys. Rev. E* 67:026221. doi: 10.1103/PhysRevE.67.026221

Boccignone, G., and Ferraro, M. (2012). Gaze shift behavior on video as composite information foraging. *Sig. Process. Image Commun.* 28, 949–966. doi: 10.1016/j.image.2012.07.002

Boerlijst, M. C., and Hogeweg, P. (1991). Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. *Physica D* 48, 17–28. doi: 10.1016/0167-2789(91)90049-F

Bray, M.-A., Lin, S.-F., Aliev, R. R., Roth, B. J., and Wikswo, J. P. (2001). Experimental and theoretical analysis of phase singularity dynamics in cardiac tissue. *J. Cardiovasc. Electrophysiol.* 12, 716–722. doi: 10.1046/j.1540-8167.2001.00716.x

Breakspear, M. (2001). Perception of odors by a nonlinear model of the olfactory bulb. *Int. J. Neural Syst.* 11, 101–124. doi: 10.1142/S0129065701000564

Breakspear, M., Heitmann, S., and Daffertshofer, A. (2010). Generative models of cortical oscillations: neurobiological implications of the kuramoto model. *Front. Hum. Neurosci.* 4:190. doi: 10.3389/fnhum.2010.00190

Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., and Wiener, M. C. (2001). Geometric visual hallucinations, euclidean symmetry and the functional architecture of striate cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 299–330. doi: 10.1098/rstb.2000.0769

Brockmann, D., and Geisel, T. (1999). "Are human scanpaths levy flights?," in *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470)*, Vol. 1 (Edinburgh), 263–268.

Brody, C. D., Romo, R., and Kepecs, A. (2003). Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Curr. Opin. Neurobiol.* 13, 204–211. doi: 10.1016/S0959-4388(03)00050-3

Broer, H. W., Huitema, G. B., and Sevryuk, M. B. (1996). *Quasi-periodic motions in families of dynamical systems: order amidst chaos*. Berlin: Springer.

Burak, Y., Rokni, U., Meister, M., and Sompolinsky, H. (2010). Bayesian model of dynamic image stabilization in the visual system. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19525–19530. doi: 10.1073/pnas.1006076107

Cannon, S. C., and Robinson, D. A. (1985). An improved neural-network model for the neural integrator of the oculomotor system: more realistic neuron behavior. *Biol. Cybern.* 53, 93–108. doi: 10.1007/BF00337026

Chan, W., and Galiana, H. L. (2005). Integrator function in the oculomotor system is dependent on sensory context. *J. Neurophysiol.* 93, 3709–3717. doi: 10.1152/jn.00814.2004

Chen, C.-Y., and Hafed, Z. M. (2013). Postmicrosaccadic enhancement of slow eye movements. *J. Neurosci.* 33, 5375–5386. doi: 10.1523/JNEUROSCI.3703-12.2013

Chiaravalloti, F., Milovanov, A. V., and Zimbardo, G. (2006). Self-similar transport processes in a two-dimensional realization of multiscale magnetic field turbulence. *Phys. Scr.* 2006:79. doi: 10.1088/0031-8949/2006/T122/012

Chu, P., Milton, J. G., and Cowan, J. D. (1994). Connectivity and the dynamics of integrate-and-fire neural networks. *Int. J. Bifurcat. Chaos* 4, 237–237. doi: 10.1142/S0218127494000198

Chun-Ni, W., Jun, M., Jun, T., and Yan-Long, L. (2010). Instability and death of spiral wave in a two-dimensional array of hindmarsh–rose neurons. *Commun. Theor. Phys.* 53:382. doi: 10.1088/0253-6102/53/2/32

Codling, E. A., Plank, M. J., and Benhamou, S. (2008). Random walk models in biology. *J. R. Soc. Interface* 5, 813–834. doi: 10.1098/rsif.2008.0014

Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. *Biol. Cybern.* 93, 91–108. doi: 10.1007/s00422-005-0574-y

Cornell-Bell, A., and Finkbeiner, S. (1991). Ca^{2+} waves in astrocytes. *Cell Calcium* 12, 185–204. doi: 10.1016/0143-4160(91)90020-F

Cornsweet, T. N. (1956). Determination of the stimuli for involuntary drifts and saccadic eye movements. *JOSA* 46, 987–988. doi: 10.1364/JOSA.46.000987

Cross, M. C., and Hohenberg, P. C. (1993). Pattern formation outside of equilibrium. *Rev. Mod. Phys.* 65:851. doi: 10.1103/RevModPhys.65.851

Da-sheng, L. S.-K. Y. (1980). The spiral structure of the tropical cyclone. *Acta Meteorologica Sin.* 3.

de Dios Navarro-López, J., Alvarado, J. C., Márquez-Ruiz, J., Escudero, M., Delgado-García, J. M., and Yajeya, J. (2004). A cholinergic synaptically triggered event participates in the generation of persistent activity necessary for eye fixation. *J. Neurosci.* 24, 5109–5118. doi: 10.1523/JNEUROSCI.0235-04.2004

Delgado-Garcia, J., Vidal, P., Gomez, C., and Berthoz, A. (1989). A neurophysiological study of prepositus hypoglossi neurons projecting to oculomotor and preoculomotor nuclei in the alert cat. *Neuroscience* 29, 291–307. doi: 10.1016/0306-4522(89)90058-4

Di Stasi, L. L., McCamy, M. B., Catena, A., Macknik, S. L., Cañas, J. J., and Martinez-Conde, S. (2013). Microsaccade and drift dynamics reflect mental fatigue. *Eur. J. Neurosci.* 38, 2389–2398. doi: 10.1111/ejn.12248

du Lac, S., and Lisberger, S. G. (1995). Cellular processing of temporal information in medial vestibular nucleus neurons. *J. Neurosci.* 15, 8000–8010.

Durstewitz, D., and Deco, G. (2008). Computational significance of transient dynamics in cortical networks. *Eur. J. Neurosci.* 27, 217–227. doi: 10.1111/j.1460-9568.2007.05976.x

Durstewitz, D., and Seamans, J. (2006). Beyond bistability: biophysics and temporal dynamics of working memory. *Neuroscience* 139, 119–133. doi: 10.1016/j.neuroscience.2005.06.094

Engbert, R., and Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychol. Sci.* 15, 431–431. doi: 10.1111/j.0956-7976.2004.00697.x

Engbert, R., and Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7192–7197. doi: 10.1073/pnas.0509557103

Engbert, R., Mergenthaler, K., Sinn, P., and Pikovsky, A. (2011). An integrated model of fixational eye movements and microsaccades. *Proc. Natl. Acad. Sci. U.S.A.* 108, E765–E770. doi: 10.1073/pnas.1102730108

Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Rep. prog. Phys.* 61:353. doi: 10.1088/0034-4885/61/4/002

Fenton, F. H., Cherry, E. M., Hastings, H. M., and Evans, S. J. (2002). Multiple mechanisms of spiral wave breakup in a model of cardiac electrical activity. *Chaos* 12, 852–892. doi: 10.1063/1.1504242

Findlay, J. (1971). Frequency analysis of human involuntary eye movement. *Kybernetik* 8, 207–214. doi: 10.1007/BF00288749

Finkbeiner, S. (1992). Calcium waves in astrocytes-filling in the gaps. *Neuron* 8, 1101–1108. doi: 10.1016/0896-6273(92)90131-V

Fisher, D., Olasagasti, I., Tank, D. W., Aksay, E. R., and Goldman, M. S. (2013). A modeling framework for deriving the structural and functional architecture of a short-term memory microcircuit. *Neuron* 79, 987–1000. doi: 10.1016/j.neuron.2013.06.041

Flusberg, B. A., Cocker, E. D., Piyawattanametha, W., Jung, J. C., Cheung, E. L., and Schnitzer, M. J. (2005). Fiber-optic fluorescence imaging. *Nat. Methods* 2, 941–950. doi: 10.1038/nmeth820

Freeman, W. J. (2009). Vortices in brain activity: their mechanism and significance for perception. *Neural Netw.* 22, 491–501. doi: 10.1016/j.neunet.2009.06.050

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K., Breakspear, M., and Deco, G. (2012). Perception and self-organized instability. *Front. Computat. Neurosci.* 6:44. doi: 10.3389/fncom.2012.00044

Froese, T., Woodward, A., and Ikegami, T. (2013). Turing instabilities in biology, culture, and consciousness? On the enactive origins of symbolic material culture. *Adapt. Behav.* 21, 199–214. doi: 10.1177/1059712313483145

Goldberger, A. L., and West, B. J. (1987). Fractals in physiology and medicine. *Yale J. Biol. Med.* 60, 421.

Golding, N. L., and Spruston, N. (1998). Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal ca1 pyramidal neurons. *Neuron* 21, 1189–1200. doi: 10.1016/S0896-6273(00)80635-2

Goldman, M. S. (2009). Memory without feedback in a neural network. *Neuron* 61, 621–634. doi: 10.1016/j.neuron.2008.12.012

Goldman, M. S., Levine, J. H., Major, G., Tank, D. W., and Seung, H. (2003). Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron. *Cereb. Cortex* 13, 1185–1195. doi: 10.1093/cercor/bhg095

Gray, R. A., and Chattipakorn, N. (2005). Termination of spiral waves during cardiac fibrillation via shock-induced phase resetting. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4672–4677. doi: 10.1073/pnas.0407860102

Gray, R. A., and Jalife, J. (1996). Spiral waves and the heart. *Int. J. Bifurcat. Chaos* 6, 415–435. doi: 10.1142/S0218127496000163

Gray, R. A., Pertsov, A. M., and Jalife, J. (1998). Spatial and temporal organization during cardiac fibrillation. *Nature* 392, 75–78. doi: 10.1038/32164

Greschner, M., Bongard, M., Rujan, P., and Ammermüller, J. (2002). Retinal ganglion cell synchronization by fixational eye movements improves feature estimation. *Nat. Neurosci.* 5, 341–347. doi: 10.1038/nn821

Grinvald, A., and Hildesheim, R. (2004). Vsdi: a new era in functional imaging of cortical dynamics. *Nat. Rev. Neurosci.* 5, 874–885. doi: 10.1038/nrn1536

Hafed, Z. M., and Krauzlis, R. J. (2012). Similarity of superior colliculus involvement in microsaccade and saccade generation. *J. Neurophysiol.* 107, 1904–1916. doi: 10.1152/jn.01125.2011

Handl, A., Mahlberg, T., Norling, S., and Gredebäck, G. (2013). Facing still faces: what visual cues affect infants' observations of others? *Infant Behav. Dev.* 36, 583–586. doi: 10.1016/j.infbeh.2013.06.001

Harel, J., Koch, C., and Perona, P. (2006). Graph-based visual saliency. *Adv. Neural Inform. Process. Syst.* 19, 545–552.

Heitmann, S. A. (2013). *Principles of Encoding Motor Commands in Travelling Waves of Neural Oscillations*. PhD thesis, University of New South wales.

Heitmann, S., Boonstra, T., and Breakspear, M. (2013). A dendritic mechanism for decoding traveling waves: principles and applications to motor cortex. *PLoS Comput. Biol.* 9:e1003260. doi: 10.1371/journal.pcbi.1003260

Heitmann, S., Gong, P., and Breakspear, M. (2012). A computational role for bistability and traveling waves in motor cortex. *Front. Comput. Neurosci.* 6:67. doi: 10.3389/fncom.2012.00067

Hillyard, S. A., Vogel, E. K., and Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1257–1270. doi: 10.1098/rstb.1998.0281

Huang, X., Troy, W. C., Yang, Q., Ma, H., Laing, C. R., Schiff, S. J., et al. (2004). Spiral waves in disinhibited mammalian neocortex. *J. Neurosci.* 24, 9897–9902. doi: 10.1523/JNEUROSCI.2705-04.2004

Huang, X., Xu, W., Liang, J., Takagaki, K., Gao, X., and Wu, J.-Y. (2010). Spiral wave dynamics in neocortex. *Neuron* 68, 978–990. doi: 10.1016/j.neuron.2010.11.007

Huang, Z. (2009). *Membrane potential fluctuations in a neural integrator*. PhD thesis, Princeton University.

Idoux, E., Serafin, M., Fort, P., Vidal, P.-P., Beraneck, M., Vibert, N., et al. (2006). Oscillatory and intrinsic membrane properties of guinea pig nucleus prepositus hypoglossi neurons *in vitro*. *J. Neurophysiol.* 96, 175–196. doi: 10.1152/jn.01355.2005

Ikegami, T. (2007). Simulating active perception and mental imagery with embodied chaotic itinerancy. *J. Conscious. Stud.* 14, 111–125.

Ito, H., and Glass, L. (1991). Spiral breakup in a new model of discrete excitable media. *Phys. Rev. Lett.* 66, 671. doi: 10.1103/PhysRevLett.66.671

Jung, P., Cornell-Bell, A., Madden, K. S., and Moss, F. (1998). Noise-induced spiral waves in astrocyte syncytia show evidence of self-organized criticality. *J. Neurophysiol.* 79, 1098–1101.

Kagan, I., and Hafed, Z. M. (2013). Active vision: microsaccades direct the eye to where it matters most. *Curr. Biol.* 23, R712–R714. doi: 10.1016/j.cub.2013.07.038

Kaneko, K. (1992). Overview of coupled map lattices. *Chaos* 2, 279. doi: 10.1063/1.165869

Kaneko, K., and Tsuda, I. (2003). Chaotic itinerancy. *Chaos* 13, 926–936. doi: 10.1063/1.1607783

Kelso, J. (1995). *Dynamic Patterns: The Self Organization of Brain and Behaviour*. Cambridge, MA: The MIT Press.

Kenny, E., Coakley, D., and Boyle, G. (2013a). Biospeckle in the human sclera and impact on laser speckle correlation measurement of eye tremor. *J. Biomed. Opt.* 18, 097009–097009. doi: 10.1117/1.JBO.18.9.097009

Kenny, E., Coakley, D., and Boyle, G. (2013b). Ocular microtremor measurement using laser-speckle metrology. *J. Biomed. Opt.* 18, 016010–016010. doi: 10.1117/1.JBO.18.1.016010

Khojasteh, E., Bockisch, C. J., Straumann, D., and Hegemann, S. C. (2012). "A re-examination of the time constant of the oculomotor neural integrator in human," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* (Osaka), 4780–4783.

Kilpatrick, Z. P., and Bressloff, P. C. (2010a). Effects of synaptic depression and adaptation on spatiotemporal dynamics of an excitatory neuronal network. *Physica D* 239, 547–560. doi: 10.1016/j.physd.2009.06.003

Kilpatrick, Z. P., and Bressloff, P. C. (2010b). Spatially structured oscillations in a two-dimensional excitatory neuronal network with synaptic depression. *J. Comput. Neurosci.* 28, 193–209. doi: 10.1007/s10827-009-0199-6

Kilpatrick, Z. P., and Ermentrout, B. (2012a). Response of traveling waves to transient inputs in neural fields. *Phys. Rev. E* 85:021910. doi: 10.1103/PhysRevE.85.021910

Kilpatrick, Z. P., and Ermentrout, B. (2013). Wandering bumps in stochastic neural fields. *SIAM J. Appl. Dyn. Syst.* 12, 61–94. doi: 10.1137/120877106

Kilpatrick, Z. P., and Ermentrout, G. B. (2012b). Hallucinogen persisting perception disorder in neuronal networks with adaptation. *J. Comput. Neurosci.* 32, 25–53. doi: 10.1007/s10827-011-0335-y

Klafter, J., Blumen, A., and Shlesinger, M. F. (1987). Stochastic pathway to anomalous diffusion. *Phys. Rev. A* 35:3081. doi: 10.1103/PhysRevA.35.3081

Ko, H.-K., Poletti, M., and Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nat. Neurosci.* 13, 1549–1553. doi: 10.1038/nn.2663

Korn, H., and Faber, D. S. (2005). The mauthner cell half a century later: a neurobiological model for decision-making? *Neuron* 47, 13–28. doi: 10.1016/j.neuron.2005.05.019

Koulakov, A. A., Raghavachari, S., Kepecs, A., and Lisman, J. E. (2002). Model for a robust neural integrator. *Nat. Neurosci.* 5, 775–782. doi: 10.1038/nn893

Kuang, X., Poletti, M., Victor, J. D., and Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Curr. Biol.* 22, 510–514. doi: 10.1016/j.cub.2012.01.050

Kuramoto, Y., and Koga, S. (1981). Turbulized rotating chemical waves. *Prog. Theor. Phys.* 66, 1081–1085. doi: 10.1143/PTP.66.1081

Langham, J., and Barkley, D. (2013). Non-specular reflections in a macroscopic system with wave-particle duality: Spiral waves in bounded media. *Chaos* 23, 013134–013134. doi: 10.1063/1.4793783

Leanhardt, A., Görlitz, A., Chikkatur, A., Kielpinski, D., Shin, Y., Pritchard, D., et al. (2002). Imprinting vortices in a bose-einstein condensate using topological phases. *Phys. Rev. Lett.* 89:190403. doi: 10.1103/PhysRevLett.89.190403

Lechleiter, J., Girard, S., Peralta, E., and Clapham, D. (1991). Spiral calcium wave propagation and annihilation in xenopus laevis oocytes. *Science* 252, 123–126. doi: 10.1126/science.2011747

Li, J., and Zhang, X. (2012). "Using high-speed photography and image processing for fixational eye movements measurement," in *Imaging Systems and Techniques (IST), 2012 IEEE International Conference on* (Manchester), 28–33.

Lim, S., and Goldman, M. S. (2013). Balanced cortical microcircuitry for maintaining information in working memory. *Nat. Neurosci.* 16, 1306–1314. doi: 10.1038/nn.3492

Loewenstein, Y., and Sompolinsky, H. (2003). Temporal integration by calcium dynamics in a model neuron. *Nat. Neurosci.* 6, 961–967. doi: 10.1038/nn1109

Ma, J., Huang, L., Tang, J., Ying, H.-P., and Jin, W.-Y. (2012a). Spiral wave death, breakup induced by ion channel poisoning on regular hodgkin–huxley neuronal networks. *Commun. Nonlin. Sci. Numer. Simul.* 17, 4281–4293. doi: 10.1016/j.cnsns.2012.03.009

Ma, J., Liu, Q., Ying, H., and Wu, Y. (2012b). Emergence of spiral wave induced by defects block. *Commun. Nonlin. Sci. Numer. Simul.* 18, 1665–1675. doi: 10.1016/j.cnsns.2012.11.016

Ma, J., Tang, J., Zhang, A., and Jia, Y. (2010). Robustness and breakup of the spiral wave in a two-dimensional lattice network of neurons. *Sci. China Phys. Mech. Astron.* 53, 672–679. doi: 10.1007/s11433-010-0097-y

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* 14, 2531–2560. doi: 10.1162/089976602760407955

Major, G., and Tank, D. (2004). Persistent neural activity: prevalence and mechanisms. *Curr. Opin. Neurobiol.* 14, 675–684. doi: 10.1016/j.conb.2004.10.017

Mantegna, R. N., and Stanley, H. E. (1994). Stochastic process with ultraslow convergence to a gaussian: the truncated lévy flight. *Phys. Rev. Lett.* 73:2946. doi: 10.1103/PhysRevLett.73.2946

Matin, L., Matin, E., and Pearce, D. G. (1970). Eye movements in the dark during the attempt to maintain a prior fixation position. *Vis. Res.* 10, 837–857. doi: 10.1016/0042-6989(70)90164-1

Martinez-Conde, S. (2006). Fixational eye movements in normal and pathological vision. *Prog. Brain Res.* 154, 151–176. doi: 10.1016/S0079-6123(06)54008-7

Martinez-Conde, S., and Macknik, S. L. (2008). Fixational eye movements across vertebrates: comparative dynamics, physiology, and perception. *J. Vis.* 8:28. doi: 10.1167/8.14.28

Martinez-Conde, S., Macknik, S. L., and Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nat. Rev. Neurosci.* 5, 229–240. doi: 10.1038/nrn1348

Martinez-Conde, S., Otero-Millan, J., and Macknik, S. L. (2013). The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nat. Rev. Neurosci.* 14, 83–96. doi: 10.1038/nrn3405

McCrea, R. A., and Horn, A. K. (2006). Nucleus prepositus. *Prog. Brain Res.* 151, 205–230. doi: 10.1016/S0079-6123(05)51007-0

Mensh, B., Aksay, E., Lee, D., Seung, H., and Tank, D. (2004). Spontaneous eye movements in goldfish: oculomotor integrator performance, plasticity, and dependence on visual feedback. *Vis. Res.* 44, 711–726. doi: 10.1016/j.visres.2003.10.015

Mergenthaler, K., and Engbert, R. (2007). Modeling the control of fixational eye movements with neurophysiological delays. *Phys. Rev. Lett.* 98:138104. doi: 10.1103/PhysRevLett.98.138104

Metzler, R., and Klafter, J. (2000). The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* 339, 1–77. doi: 10.1016/S0370-1573(00)00070-3

Michalik, M. (1987). Spektralanalysen des okulären mikrotremors bei hirnstammfunktionsstörungen. *EEG. EMG. Z. Elektroenzephalogr. Elektromyogr. Verwandte Geb.* 18, 20–26.

Milnor, J. (1985). On the concept of attractor. *Commun. Math. Phys.* 99, 177–195.

Milton, J., and Jung, P. (2003). *Epilepsy as a Dynamic Disease*. Berlin: Springer. doi: 10.1007/978-3-662-05048-4

Milton, J. G. (2012). Neuronal avalanches, epileptic quakes and other transient forms of neurodynamics. *Eur. J. Neurosci.* 36, 2156–2163. doi: 10.1111/j.1460-9568.2012.08102.x

Milton, J. G., Chu, P. H., and Cowan, J. D. (1993). Spiral waves in integrate-and-fire neural networks. *Adv. Neural Inform. Process. Syst.* 5, 1001–1006.

Miri, A., Daie, K., Arrenberg, A. B., Baier, H., Aksay, E., and Tank, D. W. (2011a). Spatial gradients and multidimensional dynamics in a neural integrator circuit. *Nat. Neurosci.* 14, 1150–1159. doi: 10.1038/nn.2888

Miri, A., Daie, K., Burdine, R. D., Aksay, E., and Tank, D. W. (2011b). Regression-based identification of behavior-encoding neurons during large-scale optical imaging of neural activity at cellular resolution. *J. Neurophysiol.* 105, 964–980. doi: 10.1152/jn.00702.2010

Modolo, J., Legros, A., Thomas, A. W., and Beuter, A. (2011). Model-driven therapeutic treatment of neurological disorders: reshaping brain rhythms with neuromodulation. *Interface Focus* 1, 61–74. doi: 10.1098/rsfs.2010.0509

Molina-Terriza, G., Torres, J. P., and Torner, L. (2007). Twisted photons. *Nat. Phys.* 3, 305–310. doi: 10.1038/nphys607

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science* 319, 1543–1546. doi: 10.1126/science.1150769

Munoz, D. P., Pelisson, D., and Guitton, D. (1991). Movement of neural activity on the superior colliculus motor map during gaze shifts. *Science* 251, 1358–1360. doi: 10.1126/science.2003221

Olsen, S. R., Bortone, D. S., Adesnik, H., and Scanziani, M. (2012). Gain control by layer six in cortical circuits of vision. *Nature* 483, 47–52. doi: 10.1038/nature10835

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–972. doi: 10.1017/S0140525X01000115

Otero-Millan, J., Macknik, S. L., Langston, R. E., and Martinez-Conde, S. (2013). An oculomotor continuum from exploration to fixation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6175–6180. doi: 10.1073/pnas.1222715110

Owen, M., Laing, C., and Coombes, S. (2007). Bumps and rings in a two-dimensional neural field: splitting and rotational instabilities. *New J. Phys.* 9:378. doi: 10.1088/1367-2630/9/10/378

Pitkow, X., Sompolinsky, H., and Meister, M. (2007). A neural computation for visual acuity in the presence of eye movements. *PLoS Biol.* 5:e331. doi: 10.1371/journal.pbio.0050331

Poletti, M., Listorti, C., and Rucci, M. (2013). Microscopic eye movements compensate for nonhomogeneous vision within the fovea. *Curr. Biol.* 23, 1691–1695. doi: 10.1016/j.cub.2013.07.007

Prechtl, J., Cohen, L., Pesaran, B., Mitra, P., and Kleinfeld, D. (1997). Visual stimuli induce waves of electrical activity in turtle cortex. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7621–7626. doi: 10.1073/pnas.94.14.7621

Qu, Z., Weiss, J. N., and Garfinkel, A. (1999). Cardiac electrical restitution properties and stability of reentrant spiral waves: a simulation study. *Am. J. Physiol. Heart Circul. Physiol.* 276, H269–H283.

Rabinovich, M., Volkovskii, A., Lecanda, P., Huerta, R., Abarbanel, H., and Laurent, G. (2001). Dynamical encoding by networks of competing neuron groups: winnerless competition. *Phys. Rev. Lett.* 87, 068102. doi: 10.1103/PhysRevLett.87.068102

Roberts, J. A., Wallis, G., and Breakspear, M. (2013). Fixational eye movements during viewing of dynamic natural scenes. *Front. Psychol.* 4:797. doi: 10.3389/fpsyg.2013.00797

Rolfs, M. (2009). Microsaccades: small steps on a long way. *Vision Res.* 49, 2415–2441. doi: 10.1016/j.visres.2009.08.010

Rothman, J., Cathala, L., Steuber, V., and Silver, R. (2009). Synaptic depression enables neuronal gain control. *Nature* 457, 1015–1018. doi: 10.1038/nature07604

Rucci, M., Iovin, R., Poletti, M., and Santini, F. (2007). Miniature eye movements enhance fine spatial detail. *Nature* 447, 852–855. doi: 10.1038/nature05866

Saalmann, Y. B., and Kastner, S. (2009). Gain control in the visual thalamus during perception and cognition. *Curr. Opin. Neurobiol.* 19, 408–414. doi: 10.1016/j.conb.2009.05.007

Salinas, E., and Sejnowski, T. J. (2001). Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. *Neuroscientist* 7, 430–440. doi: 10.1177/107385840100700512

Salinas, E., and Thier, P. (2000). Gain modulation: a major computational principle of the central nervous system. *Neuron* 27, 15–21. doi: 10.1016/S0896-6273(00)00004-0

Sandstede, B., Scheel, A., and Wulff, C. (1999). Bifurcations and dynamics of spiral waves. *J. Nonlin. Sci.* 9, 439–478. doi: 10.1007/s003329900076

Sara, S. J., and Bouret, S. (2012). Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron* 76, 130–141. doi: 10.1016/j.neuron.2012.09.011

Sato, T. K., Nauhaus, I., and Carandini, M. (2012). Traveling waves in visual cortex. *Neuron* 75, 218–229. doi: 10.1016/j.neuron.2012.06.029

Savill, N. J., Rohandi, P., and Hogeweg, P. (1997). Self-reinforcing spatial patterns enslave evolution in a host-parasitoid system. *J. theor. Biol.* 188, 11–20. doi: 10.1006/jtbi.1997.0448

Schecter, D. A., Nicholls, M. E., Persing, J., Bedard A. J. Jr., and Pielke, Sr. R. A. (2008). Infrasound emitted by tornado-like vortices: basic theory and a numerical comparison to the acoustic radiation of a single-cell thunderstorm. *J. Atmos. Sci.* 65, 685–713. doi: 10.1175/2007JAS2384.1

Schiff, S. J., Jerger, K., Duong, D. H., Chang, T., Spano, M. L., and Ditto, W. L. (1994). Controlling chaos in the brain. *Nature* 370, 615–620. doi: 10.1038/370615a0

Seliger, P., Tsimring, L., and Rabinovich, M. (2003). Dynamics-based sequential memory: winnerless competition of patterns. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 67, 011905–011901. doi: 10.1103/PhysRevE.67.011905

Sendiña-Nadal, I., Alonso, S., Pérez-Muñuzuri, V., Gómez-Gesteira, M., Pérez-Villar, V., Ramírez-Piscina, L., et al. (2000). Brownian motion of spiral waves driven by spatiotemporal structured noise. *Phys. Rev. Lett.* 84:2734. doi: 10.1103/PhysRevLett.84.2734

Serafin, M., De Waele, C., Khateb, A., Vidal, P., and Mühlethaler, M. (1991). Medial vestibular nucleus in the guinea-pig. *Exp. Brain Res.* 84, 426–433. doi: 10.1007/BF00231465

Seung, H. S. (1996). How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13339–13344. doi: 10.1073/pnas.93.23.13339

Seung, H. S., Lee, D. D., Reis, B. Y., and Tank, D. W. (2000). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* 26, 259–271. doi: 10.1016/S0896-6273(00)81155-1

Shakhnovich, A., and Thomas, J. (1977). Micro-tremor of the eyes of comatose patients. *Electroencephalogr. Clin. Neurophysiol.* 42, 117–119. doi: 10.1016/0013-4694(77)90156-0

Shen, L. (1989). Neural integration by short term potentiation. *Biol. Cybern.* 61, 319–325. doi: 10.1007/BF00203180

Shlesinger, M., West, B., and Klafter, J. (1987). Lévy dynamics of enhanced diffusion: application to turbulence. *Phys. Rev. Lett.* 58, 1100. doi: 10.1103/PhysRevLett.58.1100

Solomon, T., Weeks, E. R., and Swinney, H. L. (1993). Observation of anomalous diffusion and lévy flights in a two-dimensional rotating flow. *Phys. Rev. Lett.* 71, 3975. doi: 10.1103/PhysRevLett.71.3975

Solomon, T., Weeks, E. R., and Swinney, H. L. (1994). Chaotic advection in a two-dimensional flow: Lévy flights and anomalous diffusion. *Physica D* 76, 70–84. doi: 10.1016/0167-2789(94)90251-8

Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nat. Rev. Neurosci.* 3, 952–964. doi: 10.1038/nrn986

Spauschus, A., Marsden, J., Halliday, D. M., Rosenberg, J. R., and Brown, P. (1999). The origin of ocular microtremor in man. *Exp. Brain Res.* 126, 556–562. doi: 10.1007/s002210050764

Stacey, W. (2012). Better resolution and fewer wires discover epileptic spiral waves. *Epilepsy Curr.* 12:147. doi: 10.5698/1535-7511-12.4.147

Stosiek, C., Garaschuk, O., Holthoff, K., and Konnerth, A. (2003). *In vivo* two-photon calcium imaging of neuronal networks. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7319–7324. doi: 10.1073/pnas.1232232100

Taniguchi, D., Ishihara, S., Oonuki, T., Honda-Kitahara, M., Kaneko, K., and Sawai, S. (2013). Phase geometries of two-dimensional excitable waves govern self-organized morphodynamics of amoeboid cells. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5016–5021. doi: 10.1073/pnas.1218025110

Tegnér, J., Compte, A., and Wang, X.-J. (2002). The dynamical stability of reverberatory neural circuits. *Biol. Cybern.* 87, 471–481. doi: 10.1007/s00422-002-0363-9

Teramae, J.-N., and Fukai, T. (2005). A cellular mechanism for graded persistent activity in a model neuron and its implications in working memory. *J. Comput. Neurosci.* 18, 105–121. doi: 10.1007/s10827-005-5474-6

Thiel, M., Romano, M. C., Kurths, J., Rolfs, M., and Kliegl, R. (2008). Generating surrogates from recurrences. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 366, 545–557. doi: 10.1098/rsta.2007.2109

Toomre, A. (1969). Group velocity of spiral waves in galactic disks. *Astrophys. J.* 158, 899. doi: 10.1086/150250

Tsuda, I. (1991). Chaotic itinerancy as a dynamical basis of hermeneutics in brain and mind. *World Futures J. Gen. Evol.* 32, 167–184. doi: 10.1080/02604027.1991.9972257

Tsuda, I. (2001). Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behav. Brain Sci.* 24, 793–809. doi: 10.1017/S0140525X01000097

Tsuda, I. (2009). Hypotheses on the functional roles of chaotic transitory dynamics. *Chaos* 19, 015113–015113. doi: 10.1063/1.3076393

Tyukin, I., Tyukina, T., and van Leeuwen, C. (2009). Invariant template matching in systems with spatiotemporal coding: a matter of instability. *Neural Netw.* 22, 425–449. doi: 10.1016/j.neunet.2009.01.014

Viecelli, J. (1990). Dynamics of two-dimensional turbulence. *Phys. Fluids A Fluid Dyn.* 2:2036. doi: 10.1063/1.857678

Viventi, J., Kim, D.-H., Vigeland, L., Frechette, E. S., Blanco, J. A., Kim, Y.-S., et al. (2011). Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity *in vivo*. *Nat. Neurosci.* 14, 1599–1605. doi: 10.1038/nn.2973

Werner, G. (2010). Fractals in the nervous system: conceptual implications for theoretical neuroscience. *Front. Physiol.* 1:15. doi: 10.3389/fphys.2010.00015

West, B. J. (2010). Fractal physiology and the fractional calculus: a perspective. *Front. Physiol.* 1:12. doi: 10.3389/fphys.2010.00012

West, B. J., Bassingthwaighte, J. B., and Liebovitch, L. S. (1994). *Fractal Physiology*, Vol. 2. Oxford: Oxford University Press.

Wilkinson, N., and Metta, G. (2011). A role for cortical spiral waves in visual attention? *Procedia Comput. Sci.* 7, S1–S3. doi: 10.1016/j.procs.2012.01.092

Wilkinson, N., Metta, G., and Gredeback, G. (2011). "Modelling the face-to-face effect: sensory population dynamics and active vision can contribute to perception of social context," in *Development and Learning (ICDL), 2011 IEEE International Conference on*, Vol. 2 (Frankfurt), 1–6.

Williams, J., and Holland, M. (1999). Preparing topological states of a bose–einstein condensate. *Nature* 401, 568–572. doi: 10.1038/44095

Winfree, A. T. (1967). Biological rhythms and the behavior of populations of coupled oscillators. *J. Theor. Biol.* 16, 15–42. doi: 10.1016/0022-5193(67)90051-3

Winfree, A. T. (1972). Spiral waves of chemical activity. *Science* 175, 634–636. doi: 10.1126/science.175.4022.634

Winfree, A. T. (1991). Varieties of spiral wave behavior: an experimentalists approach to the theory of excitable media. *Chaos* 1, 303–334. doi: 10.1063/1.165844

Winfree, A. T. (2001). *The Geometry of Biological Time* Vol. 12. New York, NY: Springer. doi: 10.1007/978-1-4757-3484-3

Wu, J.-Y., Huang, X., and Zhang, C. (2008). Propagating waves of activity in the neocortex: what they are, what they do. *Neuroscientist* 14, 487–502. doi: 10.1177/1073858408317066

Wulff, C. (1996). "Theory of Meandering and Drifting Spiral Waves in Reaction-Diffusion Systems," in *Doctoral thesis*. Berlin: Friei Universitat Berlin.

Xiao-Ping, Y., Jiang-Xing, C., Ye-Hua, Z., Qin, L., Lu-Lu, W., and Qian, S. (2011). Spiral wave generation in a vortex electric field. *Chin. Phys. Lett.* 28:100505. doi: 10.1088/0256-307X/28/10/100505

Xu, X., Olivas, N. D., Levi, R., Ikrar, T., and Nenadic, Z. (2010). High precision and fast functional mapping of cortical circuitry through a novel combination of voltage sensitive dye imaging and laser scanning photostimulation. *J. Neurophysiol.* 103, 2301–2312. doi: 10.1152/jn.00992.2009

Yang, H., and Yang, J. (2007). Spiral waves in linearly coupled reaction-diffusion systems. *Phys. Rev. E* 76:016206. doi: 10.1103/PhysRevE.76.016206

Yu, G., Ma, J., Jia, Y., and Tang, J. (2010). Dynamics of spiral wave in the coupled hodgkin–huxley neurons. *Int. J. Mod. Phys. B* 24, 4555–4562. doi: 10.1142/S021797921005658X

Yuan, G., Xu, L., Xu, A., Wang, G., and Yang, S. (2011). Spiral waves in excitable media due to noise and periodic forcing. *Chaos Solit. Fract.* 44, 728–738. doi: 10.1016/j.chaos.2011.06.013

Zanos, S., Zanos, T. P., Marmarelis, V. Z., Ojemann, G. A., and Fetz, E. E. (2012). Relationships between spike-free local field potentials and spike timing in human temporal cortex. *J. Neurophysiol.* 107, 1808–1821. doi: 10.1152/jn.00663.2011

Zhang, H., Hu, B., and Hu, G. (2003). Suppression of spiral waves and spatiotemporal chaos by generating target waves in excitable media. *Phys. Rev. E* 68:026134. doi: 10.1103/PhysRevE.68.026134

Zhang, H., Hu, B., Hu, G., Ouyang, Q., and Kurths, J. (2002). Turbulence control by developing a spiral wave with a periodic signal injection in the complex ginzburg-landau equation. *Phys. Rev. E* 66:046303. doi: 10.1103/PhysRevE.66.046303

Zhang, X., and Li, J. (2012). "A novel methodology for high accuracy fixational eye movements detection," in *Proc. 4th International Conference on Bioinformatics and Biomedical Technology* (Singapore), 133–140.

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: 06 December 2013; accepted: 09 February 2014; published online: 26 February 2014.

Citation: Wilkinson NM and Metta G (2014) Capture of fixation by rotational flow; a deterministic hypothesis regarding scaling and stochasticity in fixational eye movements. Front. Syst. Neurosci. ***8****:29. doi: 10.3389/fnsys.2014.00029*

This article was submitted to the journal Frontiers in Systems Neuroscience.

frontiers in Systems Neuroscience

ORIGINAL RESEARCH
published: 31 March 2016
doi: 10.3389/fnsys.2016.00028

State-Dependent Propagation of Neuronal Sub-Population in Spontaneous Synchronized Bursts

Yuichiro Yada[1,2,3], *Ryohei Kanzaki*[1,2] and *Hirokazu Takahashi*[1,2]*

[1] Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan, [2] Department of Mechano-Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan, [3] Japan Society for the Promotion of Science, Tokyo, Japan

Repeating stable spatiotemporal patterns emerge in synchronized spontaneous activity in neuronal networks. The repertoire of such patterns can serve as memory, or a reservoir of information, in a neuronal network; moreover, the variety of patterns may represent the network memory capacity. However, a neuronal substrate for producing a repertoire of patterns in synchronization remains elusive. We herein hypothesize that state-dependent propagation of a neuronal sub-population is the key mechanism. By combining high-resolution measurement with a 4096-channel complementary metal-oxide semiconductor (CMOS) microelectrode array (MEA) and dimensionality reduction with non-negative matrix factorization (NMF), we investigated synchronized bursts of dissociated rat cortical neurons at approximately 3 weeks *in vitro*. We found that bursts had a repertoire of repeating spatiotemporal patterns, and different patterns shared a partially similar sequence of sub-population, supporting the idea of sequential structure of neuronal sub-populations during synchronized activity. We additionally found that similar spatiotemporal patterns tended to appear successively and periodically, suggesting a state-dependent fluctuation of propagation, which has been overlooked in existing literature. Thus, such a state-dependent property within the sequential sub-population structure is a plausible neural substrate for performing a repertoire of stable patterns during synchronized activity.

Keywords: spontaneous synchronized burst, state-dependent activity, microelectrode array, dissociated culture, metastable dynamics

OPEN ACCESS

Edited by:
Ruben Moreno-Bote, Universidad Pompeu Fabra, Spain

Reviewed by:
Artur Luczak, University of Lethbridge, Canada
Victor De Lafuente, Universidad Nacional Autónoma de México, Mexico

***Correspondence:**
Hirokazu Takahashi
takahashi@i.u-tokyo.ac.jp

Received: *27 January 2016*
Accepted: *14 March 2016*
Published: *31 March 2016*

Citation:
Yada Y, Kanzaki R and Takahashi H (2016) State-Dependent Propagation of Neuronal Sub-Population in Spontaneous Synchronized Bursts. Front. Syst. Neurosci. 10:28. doi: 10.3389/fnsys.2016.00028

INTRODUCTION

Repeating stable spatiotemporal patterns emerge in synchronized spontaneous activity *in vivo* (Lee and Wilson, 2002; Ji and Wilson, 2007; Luczak et al., 2007; Villette et al., 2015), *in vitro* (Beggs and Plenz, 2003, 2004; Ikegaya et al., 2004), and in dissociated cultures (Segev et al., 2004; van Pelt et al., 2004; Eytan and Marom, 2006; Madhavan et al., 2007; Rolston et al., 2007; Schroeter et al., 2015). The repertoire of such patterns can serve as memory, or a reservoir (Maass et al., 2002; Sussillo and Abbott, 2009) of information, in a neuronal network; moreover, the variety of patterns may represent the memory capacity in the network (Shew et al., 2011). Furthermore, spatiotemporal patterns in spontaneous activity are often similar to those of evoked activity against external events

Abbreviations: BFM, Burst feature matrix; SPP, Sub-population pattern; SPAW, Sub-population activation weight.

(Arieli et al., 1996; Tsodyks et al., 1999; Kenet et al., 2003; Luczak et al., 2009), suggesting that the variety of spontaneous patterns constrain the processing capacity of external inputs (Luczak et al., 2009; Villette et al., 2015). Our present interest is therefore the neural mechanism required to build a repertoire of stable spatiotemporal patterns in synchronized spontaneous activities.

Both theoretical and experimental studies have demonstrated that stable patterns emerge in a sequential structure of a neuronal network, where each synaptic connection is unreliable, yet synchronized activities of a particular sub-population reliably elicit another sub-population activity (Abeles, 1991; Aertsen et al., 1996; Diesmann et al., 1999; Ikegaya et al., 2004). Nevertheless, the way in which the repertoire of patterns is built in such a sequential structure remains elusive.

To address this problem, state-dependency of neuronal activity is a plausible neural underpinning (Buonomano and Maass, 2009). Recently, cortical spontaneous activities were characterized as having multiple "metastable states" itinerating in an activity dependent manner (Mazzucato et al., 2015). A particular state should continue during a quiescent period (Dranias et al., 2013, 2015; Ju et al., 2015) because cellular and synaptic properties governing states are likely to last without explicit spiking (Buonomano and Maass, 2009). Based on these studies, we hypothesize that (i) stable spatiotemporal patterns in synchronized spontaneous activity are generated by sequential activation of sub-populations, and that (ii) these patterns are generated in a state-dependent manner, whereby multiple metastable states can be defined as a finite continuous period.

In the present study, we test our hypotheses in dissociated neuronal cultures. To date, spontaneous activities in neuronal cultures have been well characterized with a microelectrode array (MEA) (Beggs and Plenz, 2004; Eytan and Marom, 2006; Madhavan et al., 2007). However, the spatial resolution of conventional MEA is insufficient to capture the whole activity in the neuronal network, potentially causing misestimation of population properties (Gerhard et al., 2011; Ribeiro et al., 2014). To overcome this technical pitfall, we use cutting-edge complementary metal-oxide semiconductor (CMOS) microelectrode arrays (MEAs) (Berdondini et al., 2009; Frey et al., 2010; Obien et al., 2014; Müller et al., 2015), which offer excellent spatiotemporal resolution for investigating neuronal networks *in vitro* (Gandolfo et al., 2010; Bakkum et al., 2013a; Panas et al., 2015). The CMOS MEA used in this study can simultaneously measure neural activities from 4096 sites within $2.67 \times 2.67\,mm^2$ at a sampling rate of 7 kHz. The high-dimensional spatiotemporal activity patterns are then characterized by non-negative matrix factorization (NMF) (Lee and Seung, 1999; Leonard et al., 2015; Wei et al., 2015) in order to visualize whether and how sub-populations are sequentially activated in a state-dependent manner.

We demonstrate that cultured neurons obviously perform a repertoire of multiple spatiotemporal patterns in spontaneous synchronized activity, while different patterns share a partially similar sequence of sub-populations. This supports the concept that the network has invariant sequential structures of sub-populations. Additionally, similar spatiotemporal patterns appear consecutively, which suggests that pattern generation is state-dependent. Our experimental results provide compelling evidence that a repertoire of stable neural patterns is generated in a state-dependent manner.

MATERIALS AND METHODS

Cell Culture

All experimental protocols were approved by the ethical committee of the University of Tokyo and conducted in accordance with the "Guiding Principles for the Care and Use of Animals in the Field of Physiological Science" by the Japanese Physiological Society. The cell culture procedure was based on previous reports (Bakkum et al., 2013a) and was slightly modified. Cortices were dissected from E18 Wistar rats and dissociated by 0.25% trypsin-EDTA (Invitrogen) and trituration. For cell adhesion, the electrode area of the high-density CMOS MEA (3Brain, Biochip 4096S) was coated with a 20-ul drop of 0.05% polyethylenimine (Sigma) and then a 20-ul drop of 0.02 mg/ml laminin (Sigma). On the MEAs, 30,000–40,000 cells were seeded with cell plating media: 850 ul of NeuroBasal (Invitrogen) supplemented with 10% horse serum (HyClone), 2% B27 (Invitrogen), and 0.5 mM GlutaMAX (Invitrogen). After 24 h, the media were replaced with cell growth media: 850 ul of DMEM (Invitrogen) supplemented with 10% horse serum, 0.5 mM GlutaMAX, and 10 ug of sodium pyruvate. Cultures were maintained in an incubator at 37°C and 5% CO_2 humidified atmosphere. Half the media were exchanged twice a week. For avoidance of evaporation and infection, the well on the chip was covered with a custom-made lid except for the period during the medium exchange (Potter and DeMarse, 2001).

Recording with High-Density CMOS MEAs

Extracellular voltage was recorded using a commercialized high-density CMOS MEA system (3Brain). Biochip 4096S (3Brain) contains 4096 electrodes; the scale of the electrode is 21×21 um, and the distance of neighboring electrodes is also 21 um. The electrodes are squarely located in the 2.67×2.67 mm area. Extracellular signals were simultaneously captured from the 4096 electrodes through the CMOS MEA interface, BioCAM4096 (3Brain), at a sampling rate of 7 kHz. They were recorded using BrainWave (3Brain) computer software. Ten minutes of spontaneous activities of five cultures at approximately 21 days *in vitro* (DIV) were recorded; their spontaneous activities at approximately 10 DIV were also recorded for comparison. Recording was performed outside the incubator. The recording space was shielded with a blackout curtain to avoid potential effects of ambient light (Imfeld et al., 2008) and maintained at 35–36°C atmosphere.

Spike Detection

Spikes were detected from recorded data by using a precise spike timing detection (PTSD) algorithm (Maccione et al., 2009) installed in BrainWave. The parameters for the PTSD algorithm were as follows: standard deviation factor, 10.0; peak life-time period, 2.0 ms; refractory period, 1.0 ms. The timing of each spike was assigned to the timing of its negative peak. The median of

the peak amplitude of detected spikes was calculated at each electrode; spikes detected at electrodes with lower median peaks than the threshold were excluded from the following process. The threshold was the third quartile of the median peaks, which was qualified by manual inspection. Spike sorting was not performed in this experiment.

Burst Detection

Synchronized bursts (Kamioka et al., 1996) were detected from recorded spontaneous activity by a slightly modified version of the existing adaptive algorithm (Bakkum et al., 2013b). If N_{spike} spikes occurred at all electrodes in total within less than T ms, the period was defined as a burst. The threshold time, T, is adaptively determined from the inter-spike interval (ISI) of each culture. The distribution of ISI typically forms bimodal shapes; an interval that takes the minima at the valley of the distribution is chosen as T. Here, we set N_{spike} as 200 because the number of recording channels was larger than the setup used in the original paper. Additionally, *post-hoc* processing was conducted for avoidance of burst fragmentation. The original method has excellent sensitivity in detecting small sizes of bursts; however, it separated a large burst into several small bursts in some cases. Thus, if an interval between two consecutive bursts was less than 100 ms, the two bursts were merged into a single burst.

Evaluation of Burst Peak Amplitude Distributions

The maximum number of array-wide spikes in a 10-ms time bin during a burst was defined as the peak amplitude of the burst. Distribution of spontaneous burst peak amplitude was evaluated to check diversity of bursts. The distribution of bursts in dense-plated cultures showed fixed-peak-amplitude bursts or "super bursts" at approximately 10 DIV, while they showed bimodal or long-tailed ones at approximately 20 DIV (Wagenaar et al., 2006b). Kurtosis of the distribution,

$$k = \frac{E[(x - E[x])^4]}{E[(x - E[x])^2]^2} - 3,$$

where x is the value obeying the distribution, and the chi-square goodness-of-fit test for normal distribution, were used to evaluate unimodality of the distributions.

Sub-Population Pattern Extraction by Non-Negative Matrix Factorization (NMF)

The number of spikes occurred in 10-ms time bins were counted at all electrodes and a 4096 × 60,000 matrix was obtained. The matrix was defined as an observed matrix, Y. The width of the bin was chosen on the basis of the time step used in the previous report that observed multiple recursive spatiotemporal patterns in synchronized bursts (Madhavan et al., 2007). It was hypothesized that a part of the neurons in a network constitute a co-active sub-population in the temporal resolution; the sequential activation of such sub-neuronal populations generates repeatable spatiotemporal activity of neurons as synchronized bursts. In this model, the activities of sub-populations are captured as reproducible spatial patterns; we refer to them as sub-population patterns (SPPs).

It was assumed that each element of the observation matrix, $y_{i,t}$ (i= 1, 2,. . ., 4096; t= 1, 2,. . ., 60000) , was sampled from the Poisson process with parameter $s_{i,t}$, which indicates an instantaneous firing rate,

$$p\left(y_{i,t}\right) = Poisson\left(y_{i,t}|s_{i,t}\right).$$

It was additionally assumed that the firing rate of all 4096 electrodes at each bin was generated by a linear combination of D pieces of SPPs, where D is the dimension of network activity. Consequently, the instantaneous firing rate matrix, S (4096×60,000 matrix), was represented as the product of an SPP matrix, H (4096×D matrix), which contains an SPP at each column, and a sub-population activation weight (SPAW) matrix, W (D×60, 000 matrix), which contains the coefficients for linear combination,

$$S = H \times W.$$

Notably, the SPPs could have overlapping electrodes in this model. The element of the observed matrix is the number of spikes and is thus non-negative. We thus define elements of the SPP matrix and SPAW matrix as also being non-negative. With this assumption, the SPP and SPAW matrices can be derived from the observed matrix using NMF (Lee and Seung, 1999). NMF with the generalized Kullback–Leibler divergence cost function actually assumes the Poisson generative model described above. Thus, we implemented the following optimization:

$$minimize\, D_{KL}\ (Y|HW)$$

$$s.t. \forall i, \forall d, h_{i,d} \geq 0; \forall d, \forall t, w_{d,t} \geq 0,$$

where generalized Kullback–Leibler divergence D_{KL} between matrices is:

$$D_{KL}\,(A|B) = \sum_{m,n}\left(A_{m,n} log \frac{A_{m,n}}{B_{m,n}} - A_{m,n} + B_{m,n}\right).$$

The algorithm proposed by Lee and Seung (2001) was adopted to solve this problem. Open source MATLAB codes were used with slight modification[1]. The number of SPPs, D, was empirically determined to be 10 (d= 1, 2,. . ., 10). The initial values of the matrix elements were randomly set, and the iteration loop was implanted 500 times. The calculation was independently repeated ten times. The result with the minimum cost function among all trials was used for further analysis.

Burst Pattern Classification

Bursts were classified into several classes to identify multiple recursive spatiotemporal patterns. In the present study, SPAWs during a bursting period were used as a burst feature matrix (BFM) to characterize the burst. First, however, time spans of bursts were adjusted from detected burst periods to compare spatiotemporal patterns. The initiation point of a burst was

[1]https://github.com/audiofilter/nmflib

defined as the first bin where ten spikes or more were observed in the whole network in a detected burst period.

All BFMs have the same length of time before the initiation points (pre-initiation length) and the same length of time after initiation pointsı (post-initiation length). Thus, the pre-initiation length was determined to be 100 ms to avoid inclusion of previous bursts in a BFM. The post-initiation length was adjusted depending on the cultures to include the burst with the longest length from its initiation point. Then, BFMs were classified by correlation-based hierarchical clustering with some modifications from previous studies. The maximum peak of cross correlation was defined as the similarity between BFM A and BFM B to avoid the effects of the extraction of burst periods.

$$Corr(A,\,A) \;=\; \sum_{l=1}^{L}\sum_{d=1}^{D} A_{d,l}A_{d,l}$$

$$Similarity\,(A,\;B) \;=\; \max_{k}\sum_{l=1}^{L}\sum_{d=1}^{D}\frac{A_{d,l}}{Corr\,(A,\,A)^{1/2}}\frac{B_{d,l-k}}{Corr\,(B,\,B)^{1/2}}$$
$$\left(k=-L,\;L+1,\ldots..,0,\ldots,L{-}1,L\right),$$

where L is the length of the BFMs, and zero is inserted into the elements of BFMs if $l\;-\;k<0$ or $l-k>L$. If BFM A and BFM B are identical, $Similarity\,(A,\,B)=1$. Similarities between all pairs of the BFMs were calculated; the highest similar pair was grouped. An averaged BFM of the group was then used as a new BFM that represents the grouped bursts.

The above procedure was repeated until all BFMs were merged into one class group. Subsequently, the number of classes that maximize the contrast function (Beggs and Plenz, 2004) was determined as the optimal number of classes. However, if the largest class occupied more than 90% of all bursts, the next peak of the contrast function was selected as the number of the optimal class because misdetected bursts could form a small fraction in some cases.

Sub-Population Sequence Analysis

Sequences of sub-population activity between different classes of bursts were compared. The comparison was conducted because the sequences are assumed to be partially invariant regardless of the classes of overall spatiotemporal patterns if there exists stable sequential propagation between sub-populations. First, five of the ten sub-populations with the largest peaks of SPAWs in the averaged burst were selected. The other five small SPAWs were excluded as burst-unrelated, or low contributing sub-populations. Then, spatiotemporal patterns of bursts were converted into a sequence of the timing when each sub-population took the maximum SPAWs. One of the burst classes with the largest summation of SPAWs was defined as a template class. Moreover, the sub-population sequence calculated from the averaged template-class bursts was defined as a template sequence.

Here, partial similarity with respect to the template sequence was evaluated for (i) sub-population sequences in template classes, (ii) those in non-template classes, and (iii) randomized sequences. The partial similarity was measured according to two kinds of criteria, which were adopted with slight modification from a previous study of memory replay in the hippocampus (Lee and Wilson, 2002). The first criterion is the number of times of permutation of a sub-population pair order to perfectly match the template sequence. The proportion of sub-population sequences that can match the template with the same or less than N permutation was used as the index of similarity, where N is a threshold value. The second criterion is the reproducibility of the sub-population pair order (duplet) or the sub-population trio order (triplet). A pair and a trio with the highest order consistency in all synchronized bursts were used as the duplet and triplet. The probability that the duplet or triplet was replayed within sequences was used as the index of similarity. Significance of partial similarity was statistically tested according to the similarity indices of actual sub-population sequences against those of randomized sequences. The number of randomized sequences was identical to the number of total (both the template class and non-template class) bursts. The significance was evaluated by the Mann–Whitney U-test.

Sequence Randomization

In sub-population sequence analysis, as well as in burst class consecutiveness analysis, randomized sequences were used to test the statistical significance of actual data. Random real numbers between [0 1] were sampled from the uniform distribution and assigned to all elements of the sequence. The elements of the sequence were then sorted according to the assigned numbers, and this sorted sequence was compared with actual data.

Evaluation of Burst Class Consecutiveness

Consecutiveness in sequences of burst classes was evaluated by probability of burst class transition. The probability that the same burst class was generated successively in actual data was compared with that of randomized burst-class sequences. One hundred randomized sequences were generated against each culture. The statistical significance of actual data was evaluated for each data, thereby testing the null hypothesis that the median of the probabilities in the randomized sequences is equal to the probability of the actual sequence by the one-sample Wilcoxon signed-rank test.

Evaluation of Periodical Similarity in Burst Patterns

To evaluate the periodical appearance of spatiotemporal patterns in synchronized bursts, Fisher's g-statistic was used to test the significance of periodicity (Wichert et al., 2003). Fisher's g-statistic is defined as:

$$g=\frac{max_i I\,(\omega_i)}{\sum_{i=1}^{\left[N_{sample}/2\right]} I\,(\omega_i)},$$

where $I\,(\omega_i)$ is the periodogram of the signal to evaluate, N_{sample} is the sample size of the signal, and ω_i is a discrete frequency of the signal, $\omega_i = 2\pi\, i/N_{sample}\left(i=0,\,1,\,2,\ldots,\left[N_{sample}/2\right]\right)$. The

significance level of Fisher's g-statistic was determined from the distribution:

$$P(g > g^*) = \sum_{m=1}^{M} (-1)^{m-1}{}_{\left[\frac{N_{sample}}{2}\right]} C_m (1 - mg^*)^{([N_{sample}/2]-1)},$$

where M is the largest integer less than $1/g^*$. Mean burst similarity according to (i) the difference of burst indices, and (ii) the difference of burst appearance time, from which linear components were subtracted, were used as $I(\omega_i)$ in this study. For avoiding misestimation from small samples, the differences of burst indices with more than 20 pair samples and pairs of bursts appeared within 400 s were used.

RESULTS

High-density CMOS MEAs captured spontaneous neuronal activity from five cortical networks. Ten-minute recording was performed in each culture in two developmental conditions: in a developed period−−20.8 ± 2.2 (mean ± SD) DIV—and in a juvenile period— 9.8 ± 0.8 DIV—for comparison. The number of available electrodes that detected action potentials with amplitude larger than the threshold was typically around 1000 channels. Synchronized bursts of cortical neurons were observed from the juvenile periods and stably persisted through development (Kamioka et al., 1996). The bursts were detected with the adaptive algorithm (Bakkum et al., 2013b); the average number of detected bursts was 139 ± 68 (mean ± SD) in a developed period, whereas it was 88 ± 54 in a juvenile period.

Consistent with a previous study (Wagenaar et al., 2006b), juvenile cultures exhibited fixed-peak-amplitude bursts, while developed cultures exhibited variable-peak-amplitude bursts. Representative 30-s spontaneous activities of juvenile and developed cultures are shown in **Figures 1A–D**. A juvenile network shows homogeneous spatiotemporal spiking activity (**Figures 1A,B**). In a developed one, however, bursts show heterogeneous activity (**Figures 1C,D**). The distributions of burst peak amplitude—the maximum number of array-wide spikes in a 10-ms time bin during a bursting period—show a single peak in a juvenile culture (**Figure 1E**), but a bimodal shape in a developed culture (**Figure 1F**). **Figure 1G** shows kurtosis of the burst-peak-amplitude distributions. Kurtosis tends to drop and become apart from zero with development, except one culture (Culture #3), which was excluded from further analysis.

The chi-square goodness-of-fit test was performed for testing whether the burst-peak-amplitude distributions were represented with normal distribution. The developed-period distributions disobeyed normal distribution (Culture #1, $p = 3.215\times10^{-8}$, mean ± SD: 484.8 ± 225.1; Culture #2, $p = 4.656\times10^{-10}$, mean ± SD: 250.7 ± 176.2; Culture #4, $p = 0.005357$, mean ± SD: 198.7 ± 88.1; Culture #5, $p = 3.775\times10^{-16}$, mean ± SD: 1437 ± 631), while juvenile-periods ones were characterized with normal distributions (Culture #1, $p = 0.7668$, mean ± SD: 46.36 ± 6.92; Culture #2, $p = 0.5175$, mean ± SD: 46.65 ± 7.34; Culture #4, $p = 0.9448$, mean ± SD: 137.1 ± 41.5; Culture #5, $p = 0.09618$, mean ± SD: 91.02 ± 18.23). These results demonstrate that bursts in a developed period differ from those in a juvenile period in terms of characteristic peak amplitude. We thus hypothesize that the developed cultures recruit variable neuronal sub-populations to produce different patterns.

Spontaneous spiking activity of the cultured neurons was decomposed into SPPs and SPAWs. First, the frequency of spikes occurring in 10-ms bins were calculated at each electrode. Then, the NMF algorithm (Lee and Seung, 2001) decomposed a 4096-dimensional spike frequency matrix into ten SPPs, which represented the spatial patterns of reproducibly co-activated electrodes and ten-dimensional SPAWs. The number of spikes detected at each electrode at each time bin was modeled as generated by the Poisson process with a latent parameter, which corresponded to a firing probability at the time. The Kullback–Leibler divergence NMF hypothesizes that the latent parameters of the Poisson process are a linear combination of the SPPs.

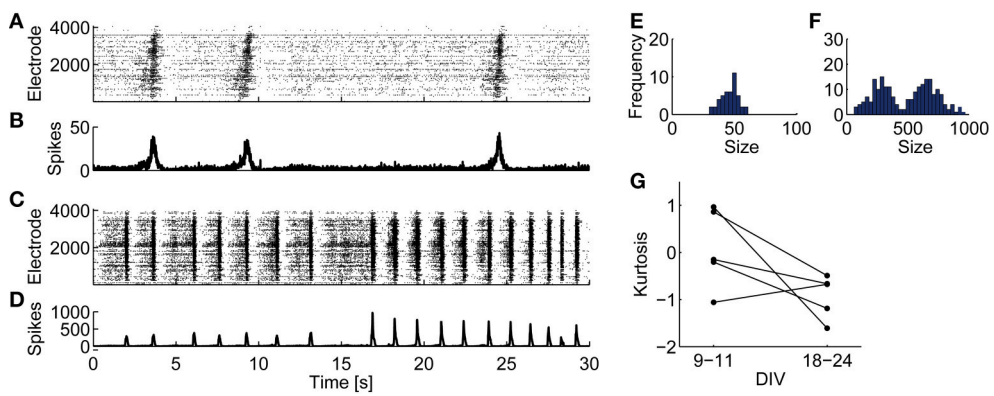


FIGURE 1 | Spontaneous spiking activities of cultured cortical neurons recorded on CMOS MEAs and distributions of burst peak amplitude. (A,B) A representative raster plot **(A)** and the number of array-wide spikes **(B)** from 30 s of spontaneous activity recorded from a culture at 10 DIV. **(C,D)** A raster plot **(C)** and the number of array-wide spikes **(D)** from recorded data from a culture at 18 DIV. **(E)** Histogram of burst peak amplitude (the maximum number of array-wide spikes in 10-ms bins during synchronized bursts) from the same recorded data shown in **(A,B)**. **(F)** Histogram of burst peak amplitude from the same recorded data shown in **(C,D)**. **(G)** Kurtosis of the burst-peak-amplitude distribution from spontaneous activities at 9–11 DIV and those at 18–24 DIV.

Non-monotonic repeating spatiotemporal patterns of synchronized bursts were observed from dimension-reduced activity obtained by NMF. The representative data in **Figure 1A** are decomposed into the temporal pattern of SPAW (**Figure 2A**) and SPPs (**Figure 2B**). The spatial distributions of co-activated electrodes in the SPPs did not always localized; some sub-populations had spatially localized activity patterns, while others had rather dispersed distributions, as illustrated in **Figure 2B**. Overall SPAW confirmed that synchronized burst patterns were reproducible both spatially and temporally. This finding is consistent with the previous study that demonstrated the stability of synchronized burst patterns in dissociated cultures (Eytan and Marom, 2006). However, the bursts in the first half (before 15 s) and those in the second half in **Figure 2A** appear to have different spatiotemporal patterns. For example, SPP #1 was recruited in the second half, but not in the first. Nevertheless, other SPPs seemed to be activated similarly in all synchronized bursts.

Spatiotemporal patterns of bursts were hierarchically clustered according to similarity of BFM and classified into several classes, as shown in the dendrogram in **Figure 2C**. The similarity matrix of bursts in **Figure 2D**, where indices of bursts are sorted according to the dendrogram in **Figure 2C**, suggests that temporal activation patterns of sub-populations in bursts are repeated. The number of classes was chosen to maximize the contrast function (**Figure 2E**) (Beggs and Plenz, 2004). The horizontal dotted line across the dendrogram in **Figure 2C** indicates the selected level of cut for the classes. Mean trajectories of SPAWs for each class are then illustrated in **Figures 2F,G**. Remarkably, temporal patterns of SPAWs in different classes seemed to be partially similar; spatiotemporal patterns in both classes were likely characterized as having

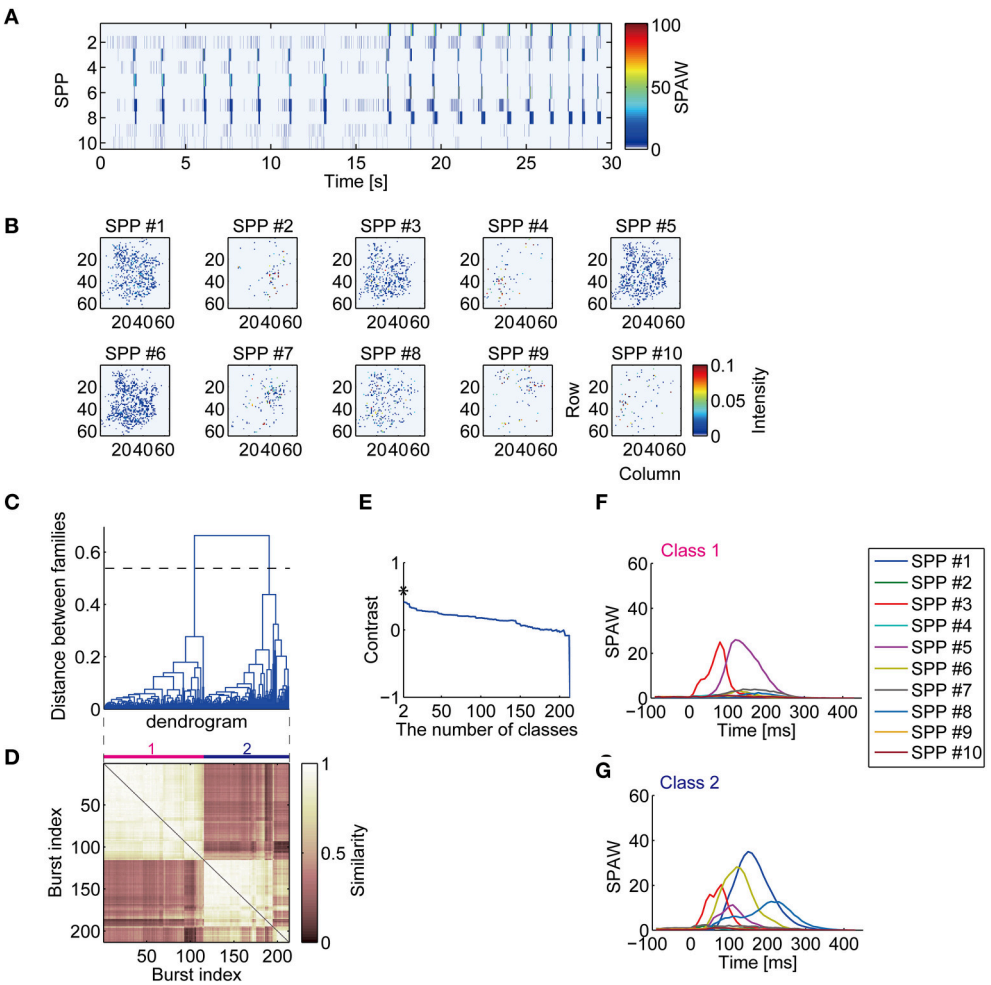


FIGURE 2 | Decomposition of high-dimensional neuronal activity and classification of synchronized burst patterns. (A) Low-dimensional activity of the neuronal network represented with sub-population activation weights (SPAWs). The same period of **Figures 1C,D** is illustrated. **(B)** Sub-population patterns (SPPs) of spontaneous activity of cultured cortical neurons obtained with NMF. The SPPs are shown as corresponding to recording electrodes configuration. **(C)** A dendrogram represents a process of hierarchical grouping of BFMs. The dotted horizontal line indicates a selected level of the grouping. **(D)** A similarity matrix of sub-population activation weights during synchronized bursts. **(E)** A contrast function for the dendrogram shown in **(C)**. Asterisk indicates a maximum peak of the function. **(F,G)** Mean SPAWs within classified burst classes.

common sub-populations with partially identical temporal orders. This suggests stable unidirectional propagation of sub-population across different classes. Such a sequential structure of neurons in spontaneous synchronized activity was suggested in previous reports (Eytan and Marom, 2006; Ham et al., 2008; Raichman and Ben-Jacob, 2008).

To evaluate partial similarity of spatiotemporal patterns among different classes, the sequences of sub-populations were compared. **Figure 3A** illustrates a schematic procedure of the analysis. Five out of ten sub-populations were selected according to the largest peaks of SPAWs in all averaged BFM to exclude effects from burst-unrelated or burst-less-related sub populations. Then, each burst was represented as a sequence of selected sub-populations. The burst class with the largest total SPAWs was defined as a template class; the sub-population sequence of averaged bursts in the template class was defined as a template sequence.

Figure 3B shows the probability that a sequence had partial similarity with the template sequence in the mean of pair permutation times. The threshold was set to two times here. Obviously, sequences in the template class had a higher probability of having partial similarity with the template sequence compared with randomly generated sequences ($p = 0.02857 < 0.05$). However, sequences in the non-template classes also showed higher probability than random sequences ($p = 0.02857 < 0.05$), although the median probability was slightly smaller than that of the template class. The findings in the permutation analysis were reconfirmed in the appearance probability of the same duplet/triplet order as shown in **Figures 3C,D** ($p = 0.02857 < 0.05$ in all comparisons). Thus,

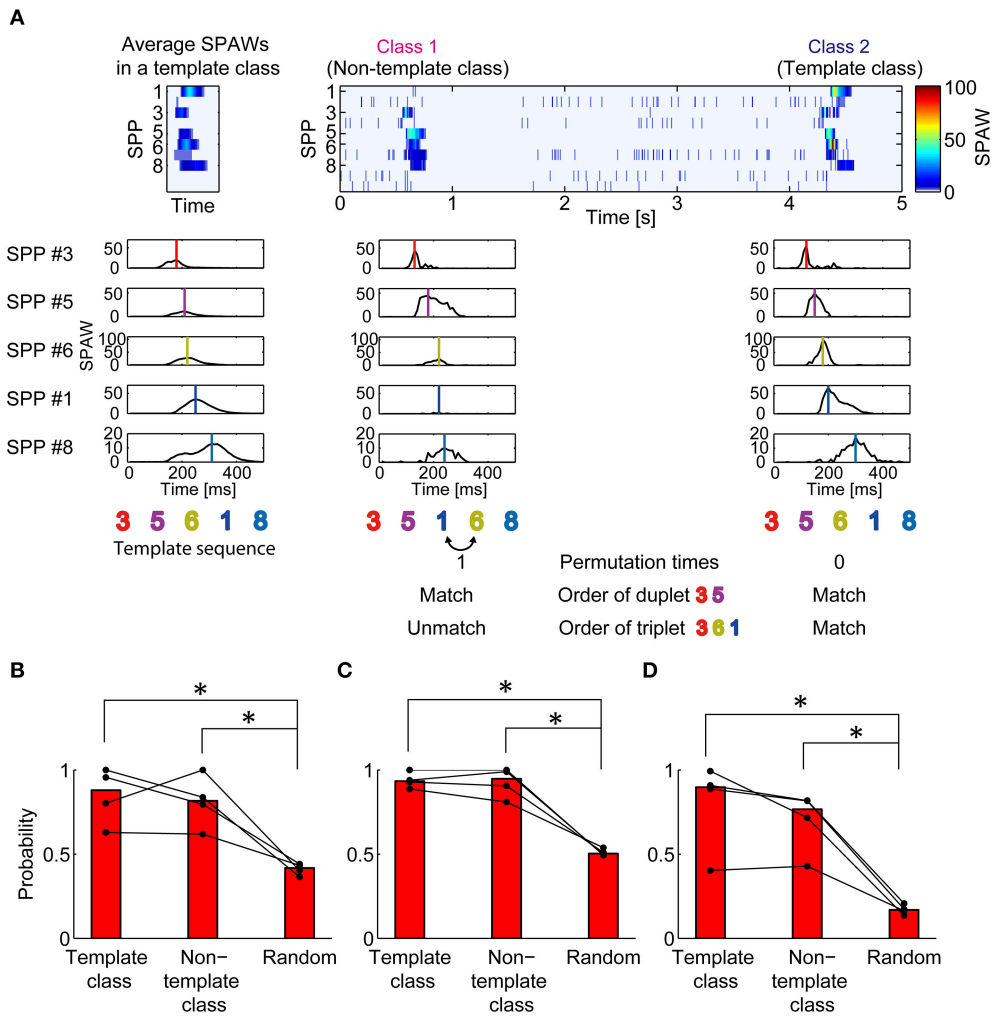


FIGURE 3 | Similarity in partial sequence of sub-population activation between synchronized burst classes. (A) Illustration of the procedure to evaluate partial similarity between sub-population sequences of bursts. SPAWs during a burst were converted into a sequence of their peaks. Sub-population sequences of bursts were compared with the template sequence. Permutation times for matching refers to the way in which many pair permutations are required to match the template sequence. Duplet/triplet order matching indicates whether the order of two/three sub-populations matches the template sequence. **(B)** Probability that permutation times for matching is two or less. **(C,D)** Probability that the duplet **(C)** or the triplet **(D)** order matched the template. The most reproduced duplet or triplet was selected for analysis. The Mann–Whitney *U*-test. $^*p < 0.05$.

different spatiotemporal patterns in synchronized bursts were likely to have partially similar sequences of sub-populations.

Next, temporal consecutiveness of these spatiotemporal patterns was investigated. If spatiotemporal patterns in synchronized bursts are an active representation of hidden states of a network (Buonomano and Maass, 2009), temporally neighboring bursts should exhibit similar patterns. **Figure 4A** shows when each class of bursts appeared through recording, clearly demonstrating that the same class of bursts appeared consecutively. The probability of remaining in the same classes of bursts was significantly higher in the experimental data than in the simulation data with randomized orders of burst classes in all the test dishes (**Figures 4B,C**; Culture #1: $p = 1.578\times10^{-30}$; Culture #2: $p = 9.466\times10^{-30}$; Culture #4: 2.157×10^{-22}; Culture #5: $p = 1.578\times10^{-30}$). In other words, a network of cortical neurons tends to maintain similar burst patterns for a certain period.

The consecutiveness in spatiotemporal patterns was also supported by the unsorted similarity matrix of bursts shown in **Figure 5A**. Similar clusters along the diagonal in the similarity matrix indicate that similar spatiotemporal patterns appeared in succession. **Figure 5B** illustrates the relationship between similarity of spatiotemporal patterns and the difference in burst appearance indices. Interestingly, not only did neighbor bursts share highly similar spatiotemporal patterns, but also the similar pairs of bursts seemed to appear periodically. From this observation, it was postulated that the combination of consecutiveness and periodicity can account for the burst similarity depending on the difference of burst appearance indices.

After subtracting the linear regression line, normalized autocorrelation of the similarity function (**Figure 5B**) was computed to visualize periodicity; then, clear periodicity was found in the autocorrelogram (**Figure 5C**). Fisher's g-statistic was calculated from the periodogram of the residual similarity function to evaluate the significance level of the largest component of the periodicity (**Figure 5D**; Wichert et al., 2003); **Figure 5D** indicates that similarity fluctuation in **Figure 5B** had 3 cycles, corresponding to 64.0 length of cycle. All cultures showed significant ($p < 0.001$) periodicity (Culture #1, $p = 1.568\times10^{-43}$; Culture #2, $p = 1.222\times10^{-5}$; Culture #4, $p = 2.270\times10^{-3}$; Culture #5, $p = 4.829\times10^{-7}$), while the length of the cycle varied between cultures (Culture #1, 64.0; Culture #2, 27.7; Culture #4, 66.0; Culture #5, 38.6). The same analysis was tested against the relationship between similarity of spatiotemporal patterns and the difference in their appearance time (**Figures 5E–G**). A mean interval of 64 bursts, i.e., a length of the cycle in **Figure 5B**, corresponds to 195.1 (±26.0, SD) s, which appeared as a peak in **Figures 5E,F**. All cultures except one showed significant ($p < 0.001$) periodicity (Culture #1, $p = 3.492\times10^{-8}$; Culture #2, $p = 1.483\times10^{-8}$; Culture #4, $p = 0.1338$; Culture #5, $p = 3.492\times10^{-8}$). Frequency of the similarity functions ranged between 0.01 and 0.02 Hz (Culture #1, 0.01 Hz; Culture #2, 0.175 Hz; Culture #5, 0.01 Hz).

Figure 5B indicates that spatiotemporal patterns may not be homogeneous during a period of the same class bursts. To address this possibility, similarity of bursts was quantified in the first two bursts in a given period (Start-Start), in the burst pairs at the start and end of a given period (Start-End), and in the successive burst pair at the transition of the period (Class

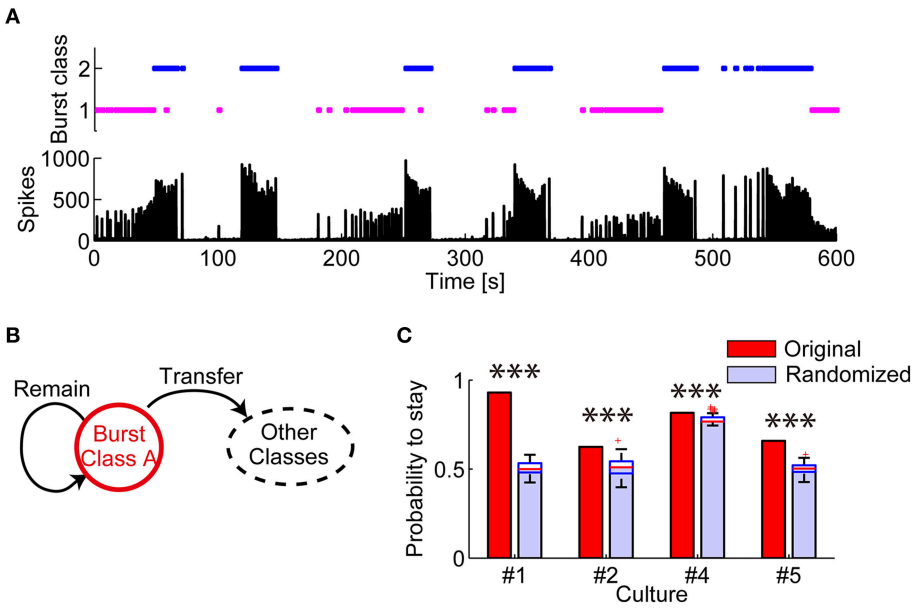


FIGURE 4 | Evaluation of consecutiveness in appearance of burst spatiotemporal patterns classes. (A) (Top) Periods in which each class of bursts appeared. (Bottom) Number of array-wide spikes through the whole recording. **(B)** Schematic illustration of the transition between multiple burst classes. The probability of "remaining" was evaluated. **(C)** Probability that the same class of bursts appeared in succession. Randomized data was generated by randomly shuffling original data. The red lines in randomized data indicate the median of the probabilities. The blue boxes are ranges from the 25th percentiles to the 75th percentiles. The whiskers are the ranges of the all probabilities excluding outliers (the red crosses). One-sample Wilcoxon signed-rank test. ***$p < 0.001$.

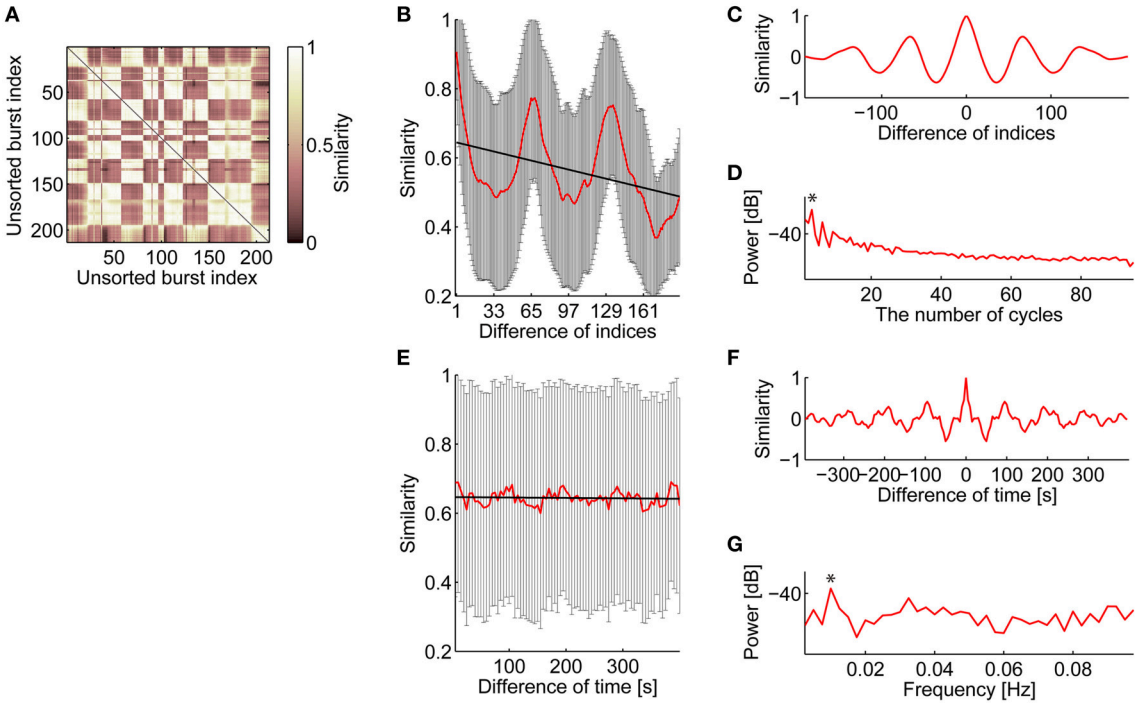


FIGURE 5 | Consecutive and periodic appearance of similar spatiotemporal patterns of bursts. (A) Similarity matrix of bursts. Bursts are listed in temporal order. **(B)** Relationship between similarity and the difference of the burst indices in temporal order. The red line indicates mean similarity against the difference of burst indices in temporal order. The error bars are SD. The thick black line is a linear regression line to the mean similarity. **(C)** Autocorrelation of the mean similarity that subtracted the regression line shown in **(B)**. **(D)** Periodogram of the mean similarity that subtracted the regression line. Asterisk indicates a maximum peak. The significance level of the maximum peak was tested using Fisher's g-statistic. **(E–G)**, The same analysis as shown in **(B–D)**, respectively, about relationship between similarity and the difference of burst appearance time.

A-Class Ā) (**Figure 6A**). Consequently, the Start pair exhibited slightly but significantly higher similarity than the Start-End pair ($p = 0.003$), while both the Start pair and the Start-End pair exhibited significantly higher similarity than the Class A-Class Ā pair ($p = 6.383 \times 10^{-7}$, $p = 3.858 \times 10^{-6}$, respectively) (**Figure 6B**). Thus, spatiotemporal patterns of the same burst class gradually change with time, yet this within-class fluctuation is much smaller than the abrupt change at the transition of the burst classes.

Consequently, these results demonstrate that spatiotemporal patterns of bursts were not stochastically generated; rather, they were consecutively and periodically generated. Our results support hypothetical ideas that spatiotemporal patterns of bursts depend on hidden internal states of the network (Buonomano and Maass, 2009), and that the internal states spontaneously and recursively fluctuate between multiple "metastable" states (Durstewitz and Deco, 2007; Mazzucato et al., 2015).

DISCUSSION

By combining high-resolution measurement with a 4096-channel CMOS MEA and dimensionality reduction with NMF, we investigated synchronized bursts of dissociated cortical neurons at approximately 3 weeks *in vitro*. We found that bursts had a repertoire of repeating spatiotemporal patterns, and different patterns shared a partially similar sequence of sub-population. These findings support the idea of propagation of neuronal sub-populations during synchronized activity (**Figure 7A**; Abeles, 1991; Ikegaya et al., 2004; Eytan and Marom, 2006). Furthermore, we found that similar spatiotemporal patterns tended to appear successively and periodically, suggesting state-dependent fluctuation of propagation (**Figure 7B**), which is overlooked in existing literature. Thus, such a state-dependent property within the sequential structure is a plausible neural substrate for performing a repertoire of stable patterns during synchronized activity.

Methodological Significances

CMOS MEA is an emerging platform for capturing electrophysiological activity of neuronal networks. It is analogous to a movie with a spatial resolution at a cellular level and temporal resolution at a single action potential (Berdondini et al., 2009). Only a small population of neurons exhibit high activity, which plausibly play crucial roles in the network (Wohrer et al., 2013); overlooking these neurons may lead to misinterpretation of results in our experiments. Thus, to avoid such sampling bias, cellular-level spatial resolution is required in the measurement (Panas et al., 2015). Furthermore, because typical burst activities of our interests last only for a few 100 ms (Eytan and Marom, 2006), the temporal resolution

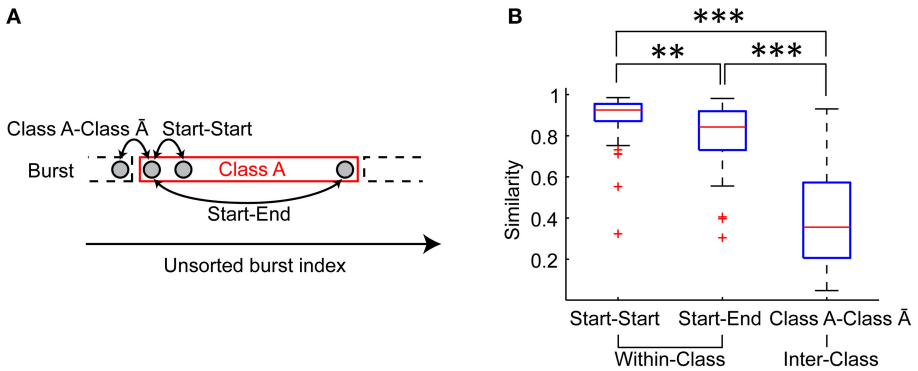


FIGURE 6 | Variation of spatiotemporal patterns during a period of the same class burst. (A) Similarity of the first two bursts in a given period (Start-Start pair), that of the burst pair at the start and end of a given period (Start-End pair) and that of the successive burst pair at the transition of the period (Class A-Class Ā pair) were evaluated. **(B)** Similarity of Start-Start pair, Start-End pair and Class A-Class Ā pair. The red lines indicate the median of the similarities. The blue boxes are ranges from the 25th percentiles to the 75th percentiles. The whiskers are the ranges of the all similarity excluding outliers (the red crosses). Wilcoxon signed-rank test. $^{**}p < 0.01$. $^{***}p < 0.001$.

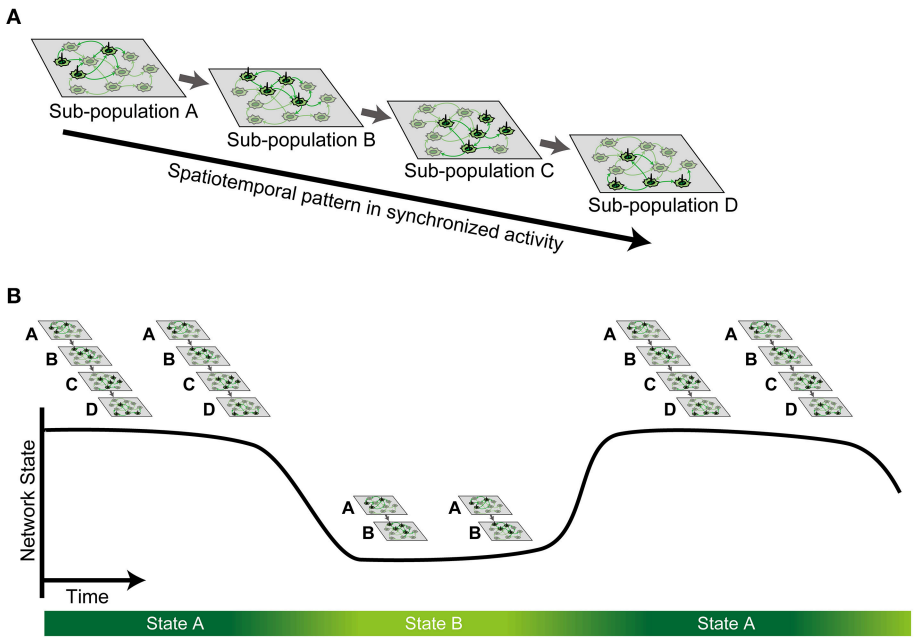


FIGURE 7 | Schematic illustration of the hypothesis. (A) Stable spatiotemporal patterns observed in synchronized spontaneous activity are generated by sequential activation of neuronal sub-population. **(B)** Such sequential activation of sub-population is state-dependent, whereby multiple metastable states can be defined as a finite continuous period.

should be on the order of ms to appropriately characterize the pattern in bursts. CMOS MEA is the only measurement device available that meets both of these spatial and temporal requirements.

The dimensions of CMOS MEA data are inherently much higher than those of the functional SPPs (Baruchi and Ben-Jacob, 2004) of our interests. Appropriate dimensionality reduction is therefore helpful in identifying functional patterns. In the present study, we employed NMF to identify stably co-activated neuronal sub-populations. Originally, NMF was developed to extract characteristic parts, such as an eye, nose, and mouth, from facial pictures (Lee and Seung, 1999). The practical advantages of NMF are that there is no need for pre-processing (Peyrache et al., 2009; Lopes-dos-Santos et al., 2013), and that non-negative components extracted from the spiking activity patterns are intuitively interpretable, just like the facial parts in facial pictures (Lee and Seung, 1999).

Repeating Spatiotemporal Patterns in Synchronized Spontaneous Activity

Consistent with the present results, previous studies showed that cortical cultures have a repertoire of repeating spatiotemporal

patterns (Segev et al., 2004; Madhavan et al., 2007; Rolston et al., 2007). A bimodal burst-peak-amplitude distribution and bimodal spatiotemporal patterns of bursts were typically observed in our experiments. In terms of burst peak amplitude, developed cultures (around 21 DIV) exhibited a bimodal distribution, whereas young cultures showed a uni-modal distribution (Eytan and Marom, 2006; Wagenaar et al., 2006b; Madhavan et al., 2007). Similarity-based clustering also demonstrated that synchronized bursts could be classified into a few patterns (Segev et al., 2004); i.e., small and large bursts (Madhavan et al., 2007). Our results extend these findings in that both large and small bursts share similar activation sequences of sub-populations.

State-Dependency and Spatiotemporal Patterns

Our results demonstrate the spontaneous itinerancy between different classes of spatiotemporal activity, suggesting that metastable states exist in innately isolated neuronal networks *in vitro*. Some intermediate states may also exist because similarity of burst patterns fluctuated continuously within a given state. Similar spontaneous transitions between metastable states were recently reported in the gustatory cortex *in vivo* (Mazzucato et al., 2015). Cortical activities were also characterized as two extreme states, i.e., desynchronized and synchronized states, with continuum of intermediate states (Harris and Thiele, 2011). These metastable states have a time scale of seconds or minutes. Thus, they are different from previously described metastable states in dissociated networks through development, which have a time scale of weeks or months and transit only unidirectionally (Pu et al., 2013).

A stable activity state of a neuronal network has been often mentioned as an attractor (Cossart et al., 2003; Wagenaar et al., 2006a). Classically, an attractor in the neural network was postulated as a memory of specific information (Hopfield, 1982). The classical attractor networks, however, are biologically implausible because the number of attractors is limited compared to the capacity of information (Maass et al., 2002) because the converging time into attractors (Maass et al., 2002; Rabinovich et al., 2008) and the effect of spontaneous activity (Kurikawa and Kaneko, 2015) are not consistent with experimental observation. Therefore, according to recent studies, it is more biologically plausible that transient metastable dynamics dominate neuronal activity (Durstewitz and Deco, 2007; Rabinovich et al., 2008).

Usually, repeating spatiotemporal activity is apprehended only as a visible sign of a metastable state (Haldeman and Beggs, 2005; Mazzucato et al., 2015) during which cellular and synaptic properties—e.g., the effect of short-term plasticity, slow inhibitory post-synaptic potentials, NMDA channel kinetics, etc.—are lasting, forming a so-called "hidden state" of neuronal networks (Buonomano and Maass, 2009). Such hidden states could account for consecutive appearances of similar bursts and can be considered an internal memory of a neuronal network. This internal memory is likely stronger than the short-term memory of external events, which is easily broken by bursts; i.e., internal memory (Dranias et al., 2013, 2015; Ju et al., 2015). Inhibitory interneurons may significantly contribute to selection of spatiotemporal patterns (Sasaki et al., 2014), depending on such hidden states.

In addition, our results demonstrate that spontaneous bursts may induce state transitions. Similarly, co-activation of some neurons trigger a state transition in the gustatory cortex (Mazzucato et al., 2015). These findings suggest that the hidden states in a neuronal network dominate spatiotemporal patterns of spontaneous activities, which in turn modulate the hidden states. Such an interaction between the hidden states and spontaneous bursts is a possible underlying mechanism of the metastable activity in neuronal networks.

Sequential Propagation Structures during Synchrony

Our results imply that multiple spatiotemporal patterns are generated by a common stable propagation structure in the network (Raichman and Ben-Jacob, 2008). Signal transmission within such a stable structure might therefore depend on the hidden states. This conceptual framework of sequential structure is compatible with previous findings of a small group of "leader neurons," which activate at burst initiation, and hierarchical structures in the dissociated neuronal networks (Eytan and Marom, 2006; Ham et al., 2008). Our results are also consistent with *in vivo* experiments in that activity bursts across states have similar spatiotemporal patterns (Luczak et al., 2013). Taken altogether, such a modified model of synfire chain with state-dependent fluctuation can account for both stability (Eytan and Marom, 2006; Panas et al., 2015) and multiple pattern generation (Segev et al., 2004; Madhavan et al., 2007; Rolston et al., 2007) in synchronized activities.

Existing models mostly overlooked state-dependent property to account for the variety of spatiotemporal patterns in the neuronal network. For example, a branching process is one of the convincing models (Beggs and Plenz, 2003). It demonstrates that cortical networks *in vitro* have metastable states, and that the critical branching process maximizes the number of metastable states (Haldeman and Beggs, 2005). Fixed propagation probabilities between neurons are postulated in these models, and spatiotemporal patterns are generated stochastically. However, this model is inconsistent with our finding that each spatiotemporal pattern does not randomly emerge; instead, it is repeated in a temporally consecutive manner. Further modeling with state-dependent properties would be one of the future directions.

Spontaneous Spatiotemporal Patterns *In vivo*

The repeating spatiotemporal patterns in spontaneous activities have been observed not only *in vitro* (Beggs and Plenz, 2004; Ikegaya et al., 2004), but also *in vivo* (Luczak et al., 2007). Neuronal networks may transmit information as "neuronal packets" (Luczak et al., 2013, 2015), i.e., activity of neuronal sub-population, generating such stable patterns, which is consistent with our results. Memory replay in the hippocampus is extensively studied as a possible mechanism

of memory consolidation during sleep (Lee and Wilson, 2002) and memory retrieval at awaking immobility (Takahashi, 2015; Villette et al., 2015). In the sensory cortex, Luczak et al. found similarity between spontaneous patterns and evoked ones. They hypothesized that a repertoire of evoked responses is a fraction of a spontaneous repertoire (Luczak et al., 2009). Spontaneous activity might be considered a prior distribution of sensory inputs (Berkes et al., 2011).

Nevertheless, in the network that experiences no external inputs, we demonstrated that repeating spatiotemporal patterns emerge in a state-dependent manner. Such network could have similar functional structures during spontaneous activity and stimulus-evoked activity (Pirino et al., 2015). Future extensive studies in both experimental and theoretical approaches are required to elucidate the functions and mechanisms of these spontaneous properties.

AUTHOR CONTRIBUTIONS

YY, RK, and HT designed research; YY performed research, composed programs and analyzed data; YY, RK, and HT wrote the paper.

ACKNOWLEDGMENTS

This works was partially supported by Japan Society for the Promotion of Science (JSPS) through Grant-in-Aid for Scientific Research (26630089), Grant-in-Aid for JSPS Fellows (26-10399) and Asahi Glass Foundation.

REFERENCES

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge: Cambridge University Press.

Aertsen, A., Diesmann, M., and Gewaltig, M. O. (1996). Propagation of synchronous spiking activity in feedforward neural networks. *J. Physiol.* 90, 243–247. doi: 10.1016/S0928-4257(97)81432-5

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. doi: 10.1126/science.273.5283.1868

Bakkum, D. J., Frey, U., Radivojevic, M., Russell, T. L., Müller, J., Fiscella, M., et al. (2013a). Tracking axonal action potential propagation on a high-density microelectrode array across hundreds of sites. *Nat. Commun.* 4, 2181. doi: 10.1038/ncomms3181

Bakkum, D. J., Radivojevic, M., Frey, U., Franke, F., Hierlemann, A., and Takahashi, H. (2013b). Parameters for burst detection. *Front. Comput. Neurosci.* 7:193. doi: 10.3389/fncom.2013.00193

Baruchi, I., and Ben-Jacob, E. (2004). Functional holography of recorded neuronal networks activity. *Neuroinformatics* 2, 333–352. doi: 10.1385/NI:2:3:333

Beggs, J. M., and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177. Available online at: http://www.jneurosci.org/content/23/35/11167.abstract

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. doi: 10.1523/JNEUROSCI.0540-04.2004

Berdondini, L., Imfeld, K., Maccione, A., Tedesco, M., Neukom, S., Koudelka-Hep, M., et al. (2009). Active pixel sensor array for high spatio-temporal resolution electrophysiological recordings from single cell to large scale neuronal networks. *Lab Chip* 9, 2644–2651. doi: 10.1039/b907394a

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331, 83–87. doi: 10.1126/science.1195870

Buonomano, D. V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125. doi: 10.1038/nrn2558

Cossart, R., Aronov, D., and Yuste, R. (2003). Attractor dynamics of network UP states in the neocortex. *Nature* 423, 283–288. doi: 10.1038/nature01614

Diesmann, M., Gewaltig, M. O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402, 529–533. doi: 10.1038/990101

Dranias, M. R., Ju, H., Rajaram, E., and VanDongen, A. M. J. (2013). Short-term memory in networks of dissociated cortical neurons. *J. Neurosci.* 33, 1940–1953. doi: 10.1523/JNEUROSCI.2718-12.2013

Dranias, M. R., Westover, M. B., Cash, S., and VanDongen, A. M. J. (2015). Stimulus information stored in lasting active and hidden network states is destroyed by network bursts. *Front. Integr. Neurosci.* 9:14. doi: 10.3389/fnint.2015.00014

Durstewitz, D., and Deco, G. (2007). Computational significance of transient dynamics in cortical networks. *Eur. J. Neurosci.* 27, 217–227. doi: 10.1111/j.1460-9568.2007.05976.x

Eytan, D., and Marom, S. (2006). Dynamics and effective topology underlying synchronization in networks of cortical neurons. *J. Neurosci.* 26, 8465–8476. doi: 10.1523/JNEUROSCI.1627-06.2006

Frey, U., Sedivy, J., Heer, F., Pedron, R., Ballini, M., Mueller, J., et al. (2010). Switch-matrix-based high-density microelectrode array in CMOS technology. *IEEE J. Solid State Circ.* 45, 467–482. doi: 10.1109/JSSC.2009.2035196

Gandolfo, M., Maccione, A., Tedesco, M., Martinoia, S., and Berdondini, L. (2010). Tracking burst patterns in hippocampal cultures with high-density CMOS-MEAs. *J. Neural Eng.* 7:056001. doi: 10.1088/1741-2560/7/5/056001

Gerhard, F., Pipa, G., Lima, B., Neuenschwander, S., and Gerstner, W. (2011). Extraction of network topology from multi-electrode recordings: is there a small-world effect? *Front. Comput. Neurosci.* 5:4. doi: 10.3389/fncom.2011.00004

Haldeman, C., and Beggs, J. M. (2005). Critical branching captures activity in living neural networks and maximizes the number of metastable states. *Phys. Rev. Lett.* 94:058101. doi: 10.1103/PhysRevLett.94.058101

Ham, M. I., Bettencourt, L. M., McDaniel, F. D., and Gross, G. W. (2008). Spontaneous coordinated activity in cultured networks: analysis of multiple ignition sites, primary circuits, and burst phase delay distributions. *J. Comput. Neurosci.* 24, 346–357. doi: 10.1007/s10827-007-0059-1

Harris, K. D., and Thiele, A. (2011). Cortical state and attention. *Nat. Pub. Group* 12, 509–523. doi: 10.1038/nrn3084

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558. doi: 10.1073/pnas.79.8.2554

Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D., et al. (2004). Synfire chains and cortical songs: temporal modules of cortical activity. *Science* 304, 559–564. doi: 10.1126/science.1093173

Imfeld, K., Neukom, S., Maccione, A., Bornat, Y., Martinoia, S., Farine, P.-A., et al. (2008). Large-scale, high-resolution data acquisition system for extracellular recording of electrophysiological activity. *IEEE Trans. Biomed. Eng.* 55, 2064–2073. doi: 10.1109/TBME.2008.919139

Ji, D., and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107. doi: 10.1038/nn1825

Ju, H., Dranias, M. R., Banumurthy, G., and VanDongen, A. M. J. (2015). Spatiotemporal memory is an intrinsic property of networks of dissociated cortical neurons. *J. Neurosci.* 35, 4040–4051. doi: 10.1523/JNEUROSCI.3793-14.2015

Kamioka, H., Maeda, E., Jimbo, Y., Robinson, H. P. C., and Kawana, A. (1996). Spontaneous periodic synchronized bursting during formation of mature patterns of connections in cortical cultures. *Neurosci. Lett.* 206, 109–112. doi: 10.1016/S0304-3940(96)12448-4

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature* 425, 954–956. doi: 10.1038/nature02078

Kurikawa, T., and Kaneko, K. (2015). Memories as bifurcations: Realization by collective dynamics of spiking neurons under stochastic inputs. *Neural Netw.* 62, 25–31. doi: 10.1016/j.neunet.2014.07.005

Lee, A. K., and Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* 36, 1183–1194. doi: 10.1016/S0896-6273(02)01096-6

Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791. doi: 10.1038/44565

Lee, D. D., and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. *Adv. Neural Inf. Process. Syst.* 13, 556–562.

Leonard, T. K., Mikkila, J. M., Eskandar, E. N., Gerrard, J. L., Kaping, D., Patel, S. R., et al. (2015). Sharp wave ripples during visual exploration in the primate hippocampus. *J. Neurosci.* 35, 14771–14782. doi: 10.1523/JNEUROSCI.0864-15.2015

Lopes-dos-Santos, V., Ribeiro, S., and Tort, A. B. L. (2013). Detecting cell assemblies in large neuronal populations. *J. Neurosci. Methods* 220, 149–166. doi: 10.1016/j.jneumeth.2013.04.010

Luczak, A., Bartho, P., and Harris, K. D. (2013). Gating of sensory input by spontaneous cortical activity. *J. Neurosci.* 33, 1684–1695. doi: 10.1523/JNEUROSCI.2928-12.2013

Luczak, A., Barthó, P., and Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* 62, 413–425. doi: 10.1016/j.neuron.2009.03.014

Luczak, A., Barthó, P., Marguet, S. L., Buzsáki, G., and Harris, K. D. (2007). Sequential structure of neocortical spontaneous activity *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 347–352. doi: 10.1073/pnas.0605643104

Luczak, A., McNaughton, B. L., and Harris, K. D. (2015). Packet-based communication in the cortex. *Nat. Rev. Neurosci.* 16, 745–755. doi: 10.1038/nrn4026

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* 14, 2531–2560. doi: 10.1162/089976602760407955

Maccione, A., Gandolfo, M., Massobrio, P., Novellino, A., Martinoia, S., and Chiappalone, M. (2009). A novel algorithm for precise identification of spikes in extracellularly recorded neuronal signals. *J. Neurosci. Methods* 177, 241–249. doi: 10.1016/j.jneumeth.2008.09.026

Madhavan, R., Chao, Z. C., and Potter, S. M. (2007). Plasticity of recurring spatiotemporal activity patterns in cortical networks. *Phys. Biol.* 4, 181–193. doi: 10.1088/1478-3975/4/3/005

Mazzucato, L., Fontanini, A., and La Camera, G. (2015). Dynamics of multistable states during ongoing and evoked cortical activity. *J. Neurosci.* 35, 8214–8231. doi: 10.1523/JNEUROSCI.4819-14.2015

Müller, J., Ballini, M., Livi, P., Chen, Y., Radivojevic, M., Shadmani, A., et al. (2015). High-resolution CMOS MEA platform to study neurons at subcellular, cellular, and network levels. *Lab Chip* 15, 2767–2780. doi: 10.1039/C5LC00133A

Obien, M. E. J., Deligkaris, K., Bullmann, T., Bakkum, D. J., and Frey, U. (2014). Revealing neuronal function through microelectrode array recordings. *Front. Neurosci.* 8:423. doi: 10.3389/fnins.2014.00423

Panas, D., Amin, H., Maccione, A., Muthmann, O., van Rossum, M., Berdondini, L., et al. (2015). Sloppiness in spontaneously active neuronal networks. *J. Neurosci.* 35, 8480–8492. doi: 10.1523/JNEUROSCI.4421-14.2015

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., and Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat. Neurosci.* 12, 919–926. doi: 10.1038/nn.2337

Pirino, V., Riccomagno, E., Martinoia, S., and Massobrio, P. (2015). A topological study of repetitive co-activation networks in *in vitro* cortical assemblies. *Phys. Biol.* 12, 1–15. doi: 10.1088/1478-3975/12/1/016007

Potter, S. M., and DeMarse, T. B. (2001). A new approach to neural cell culture for long-term studies. *J. Neurosci. Methods* 110, 17–24. doi: 10.1016/S0165-0270(01)00412-5

Pu, J., Gong, H., Li, X., and Luo, Q. (2013). Developing neuronal networks: self-organized criticality predicts the future. *Sci. Rep.* 3:1081. doi: 10.1038/srep01081

Rabinovich, M., Huerta, R., and Laurent, G. (2008). Transient dynamics for neural processing. *Science* 321, 48–50. doi: 10.1126/science.1155564

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. doi: 10.1016/j.jneumeth.2007.12.020

Ribeiro, T. L., Ribeiro, S., Belchior, H., Caixeta, F., and Copelli, M. (2014). Undersampled critical branching processes on small-world and random networks fail to reproduce the statistics of spike avalanches. *PLoS ONE* 9:e94992. doi: 10.1371/journal.pone.0094992

Rolston, J. D., Wagenaar, D. A., and Potter, S. M. (2007). Precisely timed spatiotemporal patterns of neural activity in dissociated cortical cultures. *Neuroscience* 148, 294–303. doi: 10.1016/j.neuroscience.2007.05.025

Sasaki, T., Matsuki, N., and Ikegaya, Y. (2014). Interneuron firing precedes sequential activation of neuronal ensembles in hippocampal slices. *Eur. J. Neurosci.* 39, 2027–2036. doi: 10.1111/ejn.12554

Schroeter, M. S., Charlesworth, P., Kitzbichler, M. G., Paulsen, O., and Bullmore, E. T. (2015). Emergence of rich-club topology and coordinated dynamics in development of hippocampal functional networks *in vitro*. *J. Neurosci.* 35, 5459–5470. doi: 10.1523/JNEUROSCI.4259-14.2015

Segev, R., Baruchi, I., Hulata, E., and Ben-Jacob, E. (2004). Hidden neuronal correlations in cultured networks. *Phys. Rev. Lett.* 92:118102. doi: 10.1103/PhysRevLett.92.118102

Shew, W. L., Yang, H., Yu, S., Roy, R., and Plenz, D. (2011). Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches. *J. Neurosci.* 31, 55–63. doi: 10.1523/JNEUROSCI.4637-10.2011

Sussillo, D., and Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron* 63, 544–557. doi: 10.1016/j.neuron.2009.07.018

Takahashi, S. (2015). Episodic-like memory trace in awake replay of hippocampal place cell activity sequences. *Elife* 4:e08105. doi: 10.7554/eLife.08105. Available online at: http://elifesciences.org/content/4/e08105v1

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. doi: 10.1126/science.286.5446.1943

van Pelt, J., Wolters, P. S., Corner, M. A., Rutten, W. L. C., and Ramakers, G. J. A. (2004). Long-term characterization of firing dynamics of spontaneous bursts in cultured neural networks. *IEEE Trans. Biomed. Eng.* 51, 2051–2062. doi: 10.1109/TBME.2004.827936

Villette, V., Malvache, A., Tressard, T., Dupuy, N., and Cossart, R. (2015). Internally recurring hippocampal sequences as a population template of spatiotemporal information. *Neuron* 88, 357–366. doi: 10.1016/j.neuron.2015.09.052

Wagenaar, D. A., Nadasdy, Z., and Potter, S. M. (2006a). Persistent dynamic attractors in activity patterns of cultured neuronal networks. *Phys. Rev. E* 73:051907. doi: 10.1103/PhysRevE.73.051907

Wagenaar, D. A., Pine, J., and Potter, S. M. (2006b). An extremely rich repertoire of bursting patterns during the development of cortical cultures. *BMC Neurosci.* 7:11. doi: 10.1186/1471-2202-7-11

Wei, J., Bai, W., Liu, T., and Tian, X. (2015). Functional connectivity changes during a working memory task in rat via NMF analysis. *Front. Behav. Neurosci.* 9:2. doi: 10.3389/fnbeh.2015.00002

Wichert, S., Fokianos, K., and Strimmer, K. (2003). Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics* 20, 5–20. doi: 10.1093/bioinformatics/btg364

Wohrer, A., Humphries, M. D., and Machens, C. K. (2013). Population-wide distributions of neural activity during perceptual decision-making. *Prog. Neurobiol.* 103, 156–193. doi: 10.1016/j.pneurobio.2012.09.004

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.

frontiers in
SYSTEMS NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 04 December 2013
doi: 10.3389/fnsys.2013.00101

Decomposition of spontaneous brain activity into distinct fMRI co-activation patterns

Xiao Liu*, Catie Chang and Jeff H. Duyn

Advanced MRI Section, Laboratory of Functional and Molecular Imaging, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

Edited by:
Emili Balaguer-Ballester, Bournemouth University, UK

Reviewed by:
Huo Lu, Philadelphia College of Osteopathic Medicine, USA
Daniele Marinazzo, University of Ghent, Belgium

****Correspondence:***
Xiao Liu, Advanced MRI Section, Laboratory of Functional and Molecular Imaging, National Institute of Neurological Disorders and Stroke, National Institutes of Health, 10 Center Dr., Bldg. 10, Rm. B1D-723A - MSC 1065, Bethesda, MD 20892-1065, USA
e-mail: liux15@ninds.nih.gov

Recent fMRI studies have shown that analysis of the human brain's spontaneous activity may provide a powerful approach to reveal its functional organization. Dedicated methods have been proposed to investigate co-variation of signals from different brain regions, with the goal of revealing neuronal networks (NNs) that may serve specialized functions. However, these analysis methods generally do not take into account a potential non-stationary (variable) interaction between brain regions, and as a result have limited effectiveness. To address this, we propose a novel analysis method that uses clustering analysis to sort and selectively average fMRI activity time frames to produce a set of co-activation patterns. Compared to the established networks extracted with conventional analysis methods, these co-activation patterns demonstrate novel network features with apparent relevance to the brain's functional organization.

Keywords: resting-state network, non-stationary connectivity, network dynamics, clustering, dynamic connectivity

INTRODUCTION

A growing body of neuroimaging research is reporting on the phenomenon of spontaneous neural activity occurring during rest, in the absence of overt behavior (Biswal et al., 1995; Arieli et al., 1996; Leopold and Maier, 2012). Functional magnetic resonance imaging (fMRI) (Ogawa et al., 1992) studies of the nature of this activity have revealed multiple spatial patterns of temporally correlated signal fluctuation that cover much of the brain, and often align with the established systems that support specialized brain functions (Biswal et al., 1995; Cordes et al., 2000; Hampson et al., 2002; Greicius et al., 2003; Fox et al., 2006). Based on this, it has been hypothesized that these correlational patterns of fMRI activity (often called "resting-state networks," or RSNs) (Fox and Raichle, 2007) indirectly result from spontaneous electrical activity in a number of distinct, large scale, and function-specific neuronal networks (NNs). Interpretation of RSNs in terms of the NNs supporting the major brain functions is an important goal of current neuroimaging research (Fox and Raichle, 2007; Biswal et al., 2010; Zhang and Raichle, 2010).

To tackle this challenging problem, a number of methods have been applied to analyze resting-state fMRI signals, including "seed"-based correlation analysis (Biswal et al., 1995; Fox et al., 2006), clustering based on temporal characteristics (Cordes et al., 2002; Mezer et al., 2009) and spatial or temporal independent component analysis (ICA) (Kiviniemi et al., 2003; Beckmann et al., 2005; Smith et al., 2012). These methods generally make implicit or explicit assumptions about the source signals underlying spontaneous fMRI activity, including stationarity of inter-regional interactions, and/or a statistical independence. The extent to which these assumptions reflect the nature of NNs determines how accurately they are represented by RSNs.

Recent studies have provided evidence that spontaneous brain activity may be non-stationary (Chang and Glover, 2010; Allen et al., 2012; Hutchison et al., 2012; Rack-Gomer and Liu, 2012), and in fact may be dominated by brief instances of spontaneous co-activation of brain regions (Tagliazucchi et al., 2012; Liu and Duyn, 2013; Wu et al., 2013). This has inspired a novel analysis approach that temporally decomposes conventional RSNs into multiple co-activation patterns by selective averaging of single fMRI time frames (Liu and Duyn, 2013). Here we extend this approach and perform a comprehensive analysis of a publicly accessible fMRI database to extract 30 spatial patterns of spontaneous activity (termed co-activation patterns or CAPs) that are biologically plausible and show distinct differences from networks extracted with conventional methods.

METHODS

ANALYSIS APPROACH

The proposed analysis approach is based on the notion that spontaneous activity may be dominated by brief activations and deactivations involving many (possibly overlapping) brain regions. It differs from point process analysis (PPA), which models activity underlying fMRI signals as point processes and then examines the conditional distribution of these processes given its occurrence at a specific seed region (Tagliazucchi et al., 2012; Wu et al., 2013), in that it is not specifically geared toward detecting neuronal-avalanche type activity. Rather than selectively averaging time points of activity increases in a seed region, the method proposed here classifies and averages time points with similar spatial distributions of activity using the *k*-means clustering algorithm. It thus extends the seed-based approach presented in a previous study (Liu and Duyn, 2013) to a data-driven, whole-brain analysis.

Clustering is a procedure for classifying a set of objects into different groups such that within group differences are smaller than across group differences. An approach popular with fMRI studies is the *k*-means clustering method, which has previously been applied to classify brain voxels based either on their temporal dynamics (Cordes et al., 2002; Mezer et al., 2009) or correlation profiles (Anderson et al., 2010; Kelly et al., 2012). Here, in contrast, we apply *k*-means clustering to classify fMRI time points (fMRI image volumes) based on their *spatial* similarity.

After reformatting fMRI brain volumes into a set of m-dimensional vectors $\{t_1, t_2, \ldots, t_n\}$ (where $m =$ the number of brain voxels, based on a brain mask created from the 152-brain MNI template), *k*-means clustering is applied to partition the vectors into k clusters $\boldsymbol{R} = \{R_1, R_2, \ldots, R_k\}$ (Gluck and Myers, 1997) such that the sum of within-cluster distances J (Equation 1) is minimized:

$$J = \sum_{i=1}^{k} \sum_{t_j \in R_i} d(t_j, \mu_i) \quad (1)$$

where μ_i is the mean of fMRI volumes in R_i, and $d(\bullet)$ represents the distance between two vectors, which we define here to be 1 minus their Pearson's correlation coefficient.

DATASET

The resting-state fMRI dataset used in this study was obtained from the 1000 Functional Connectomes Project (FCP) (http://www.nitrc.org/projects/fcon_1000/) (Biswal et al., 2010). It includes data from studies performed independently at three different sites (Baltimore, Berlin_Margulies, and Cambridge_Buckner) with approval from their respective ethics committees. Due to computational limitations, we focused our analysis on data from 102 subjects (mean age: 24.4 ± 6.6, range: 18-44; sex: 64 females) selected from all 3 sites. Detailed information regarding each dataset and the major MR acquisition parameters can be found at the FCP website.

PRE-PROCESSING OF RESTING-STATE fMRI SIGNALS

FCP analysis scripts (version 1.1-beta, available at http://www.nitrc.org/frs/shownotes.php?release_id=938) (Biswal et al., 2010), which employs AFNI (Cox, 1996) and FSL (http://www.fmrib.ox.ac.uk/fsl/) (Smith et al., 2004) software packages, were used to pre-process the fMRI signals (with minor modifications, described below). The typical pre-processing steps for the analysis of resting state data were applied, including image coregistration to correct for head motion, spatial smoothing with a Gaussian kernel (FWHM = 4 mm), temporal filtering with a band-pass filter (0.005–0.1 Hz), and the removal of linear and quadratic temporal trends. Additionally, the time series of ROIs in the white matter and cerebrospinal fluid (CSF), 6 affine motion parameters, as well as the brain-averaged (global) signal, were used as nuisance variables to be regressed out from the data. Given that global signal regression (GSR) may introduce artificial anti-correlation between regions (Fox et al., 2009; Murphy et al., 2009; Saad et al., 2012) our analysis was also performed without the GSR step (Figure S1).

The fMRI data was first co-registered to the high-resolution anatomical (T_1-weighted) images acquired from the same subject and then normalized to the 152-brain Montreal Neurological Institute (MNI) normalized space. Here, as departure from the original FCP scripts, the registration between the functional and anatomical images was implemented using the "align_epi_anat.py" program (Saad et al., 2009) in AFNI, which was found to provide a better registration in the superior-inferior direction. The pre-processed fMRI data were then resampled at the $3 \times 3 \times 3\,\text{mm}^3$ resolution of the MNI normalized brain space. Finally, for each voxel, the fMRI signal was temporally normalized by subtracting its mean and then dividing by its temporal standard deviation (*SD*).

EXTRACTION OF CAPs

The clustering was applied to all 13382 fMRI volumes acquired from all 102 participants. After clustering, the fMRI volumes (also referred to as "time frames") assigned to the same cluster were simply averaged, resulting in k maps that we define as CAPs. These CAPs were then normalized by the standard error (within cluster and across fMRI volumes) to generate Z-statistic maps, which quantify the degree of significance to which the CAP map values (for each voxel) deviate from zero. We also calculated three quantities for each CAP based on its raw map: (1) the occurrence rate, which was calculated by dividing the number of fMRI time frames belonging to a given CAP by the total number of time frames; (2) the within-cluster similarity, calculated as the average spatial correlation of all within-cluster volumes to their mean; and (3) the polarity, calculated as the sum of the mean of positive map values and that of negative map values, such that the resulting sign indicates whether the CAP is dominated by activation or de-activation.

To suppress the contribution of measurement noise, only the largest signal changes were considered for calculation of distances for the clustering procedure. This was achieved by applying a mask to the fMRI volumes that only admitted the 10% highest and 5% lowest signal values, and discarded regions with less than 6 inter-connected (in 3D) voxels. This masking was only performed for the clustering and not for the within-cluster averaging procedure for calculation of the grand-average CAPs. Although this masking procedure was not strictly required and did not affect the general observations in this study, it was found to result in slightly more specific CAP patterns (Figure S2).

The number of clusters k was set to 30 after comparing the outcomes of setting k equal to 20, 30, and 40. Although these three cases generated largely similar CAPs, clustering with $k = 30$ led to a few distinct CAPs that were not found with $k = 20$, while $k = 40$ led to several CAPs that were nearly indistinguishable. Thus, the choice of $k = 30$ was a compromise between extracting too many and too few distinct CAPs base on what was afforded by the data.

OCCURRENCE RATE OF CAPs

For each fMRI session, the number of fMRI time frames belonging to a specific CAP was divided by the total number of frames in order to quantify the occurrence rate of the CAP in a subject (only one session for each participant). A permutation test was applied to determine whether the occurrence rates of the CAPs were significantly different between male and female subjects. Specifically, the 102 subjects were randomly assigned to two groups with size of 38 and 64, respectively, without considering their gender, and

the difference in CAP occurrence rate between the two groups was recorded. This process was then repeated 50,000 times to build up distributions of the between-group difference for each CAP. The probabilities (p-values) of seeing differences between the male and female groups were then calculated by comparing the actual observations to these distributions, and a Bonferroni correction was applied to correct for multiple comparisons.

Head motion was also quantified for the male and female groups based on two metrics (mean translation, mean rotation) calculated from the 6 affine motion parameters. The mean translation was defined as the average displacement between any two consecutive volumes (defined as the root-mean-square of the translation parameters), and the mean rotation was calculated in similar way using the rotation parameters (Van Dijk et al., 2012).

CORRELATION MAPS

For comparison, seed-based correlation maps were calculated by correlating fMRI signals from all brain voxels to those from four seed regions ($6 \times 6 \times 6\,mm^3$ cubes) centered at the following locations (with coordinates given in MNI space): posterior cingulate cortex [PCC, (0, −53, 26)], medical prefrontal cortex [mPFC, (0, 52, -6)], left intraparietal cortex [IPS, (−24, −58, 52)], and motor cortex [(−36, −25, 57)], respectively (Van Dijk et al., 2010).

SPATIAL ICs AND TEMPORAL FUNCTIONAL MODES (TFMs)

As another methodological comparison, group-level spatial ICA with temporal concatenation was implemented with the MELODIC program in the FSL software (Beckmann et al., 2005). The number of components to be extracted was specified as 30, matching the number of the clusters specified for the clustering. The CAPs were also compared to 21 TFMs derived previously by temporal ICA (Smith et al., 2012), which are available at SumsDB (http://sumsdb.wustl.edu/sums/directory.do?id=8288032&dir_name=TFM_PNAS).

For comparison, the CAPs and ICs/TFMs that have highest spatial correlation to each other among their own category were paired up and shown in **Figure 6** and Figure S7.

RESULTS

ACTIVITY DISTRIBUTION AT SINGLE TIME FRAMES

Exemplary activity time frames (T1-T13) from a resting-state fMRI scan are shown in **Figure 1**, together with a map derived by temporally correlating signals across the brain with that from a "seed" region in the PCC. While the correlation map highlights the PCC, medial frontal cortex (MF), and bilateral parietal cortices (LPC), all recognized as primary nodes of an RSN known as the "default mode" network (DMN) (Greicius et al., 2003), single time frames often show clearly deviating patterns. For example, while T6, T11, T12, and T13 resemble DMN, T1 and T7 show co-activations at the visual cortex; T3 and T9 have high signal level specifically at the sensorimotor and insular cortex; T2, T4, T5, and T8 cover different sub-sections of the intraparietal sulcus (IPS) as well as some frontal areas, including the frontal eye fields (FEF); at T10, no apparent co-activation pattern is seen. It is worth noting that some of the patterns include strong co-activation at small but very specific brain structures, e.g., the hippocampus (HI) and posterior parahippocampal gyrus (PHG) in T11 and the ventral lateral nucleus (VL) of the thalamus in T3 (white arrows in **Figure 1**). Although some of these differences may be manifestations of experimental noise, their spatial characteristics (e.g., bilaterality and anatomical specificity) are suggestive of a neuronal origin.

SPONTANEOUS CO-ACTIVATION PATTERNS (CAPs)

To characterize stable, recurring co-activation patterns in spontaneous activity, single fMRI time-frames (13382 frames total) were clustered into 30 groups based on their spatial characteristics, after which within-group averaging was performed to calculate canonical CAPs (see Methods). These 30 CAPs were ranked by their similarity values, defined as the average correlation between each fMRI time frame within a group and their mean (Figure S3). Many of these CAPs resembled RSNs extracted with conventional analysis, suggesting network activity in the DMN, sensory regions with or without DMN involvement (see examples in Figure S3).

In various brain regions, CAPs showed interesting differences from RSNs (**Figure 2**). For example, DMN CAPs 1 and 15 selectively involved the HI, CAP 7 selectively involved the posterior PHG, and CAP 13 involved both. This distinct involvement of HI and PHG in DMN is not observed with conventional analysis. In addition, CAP 1 showed an asymmetric activation at the superior portion of the left middle frontal gyrus (MFG) even though its general pattern is fairly symmetric between hemispheres; CAP 7 has very specific activation along the superior frontal gyrus (SFG), and its LPC components localize more posteriorly compared to the others; and CAP 15 shows clear co-activations at the caudate nucleus (CN). Moreover, co-activations at the PCC/Precuneus regions in CAPs 7 and 13 appear to be patterned differently from those in CAPs 1 and 15. All these DMN-related CAPs are also associated with strong de-activations in a set of "task-positive" regions (Fox et al., 2005), although to varying spatial extent.

A second group of CAPs showed de-activations in DMN-related regions combined with strong co-activations in a set of "task-positive" regions (**Figure 3**). Their patterns appeared more distinct from one another as compared to the DMN-activated CAPs. While CAPs 2 and 3 are both related to the visual system, CAP 2 overlaps more on the high order visual area and FEF. CAPs 4 and 6 looks almost identical in some axial slices, but a close comparison reveals that CAP 4 covers the central opercular cortex (CO), parietal operculum (PO), insular, thalamus, and supplementary motor area (SMA), whereas CAP 6 involves more anterior regions, including the frontal operculum (FO) and paracingulate gyrus (PCG). Moreover, even though all these CAPs show activations in the IPS, they cover very different sub-sections. There are also CAPs showing highly lateralized patterns (Figure S4) or strong activation in the primary sensory regions (Figure S5).

Several CAPs showed focal activity in thalamic and cerebellar structures. For example, CAPs 2 and 3, both of which involve higher order visual and visual association areas, include spatially distinct thalamic nuclei, with the ones in CAP 2 located more superior, medial, and anterior than those in CAP 3 (**Figure 4A**). In both CAPs, this thalamic activity appears to occur within the pulvinar. In contrast, CAP 26, whose activity shows a preference

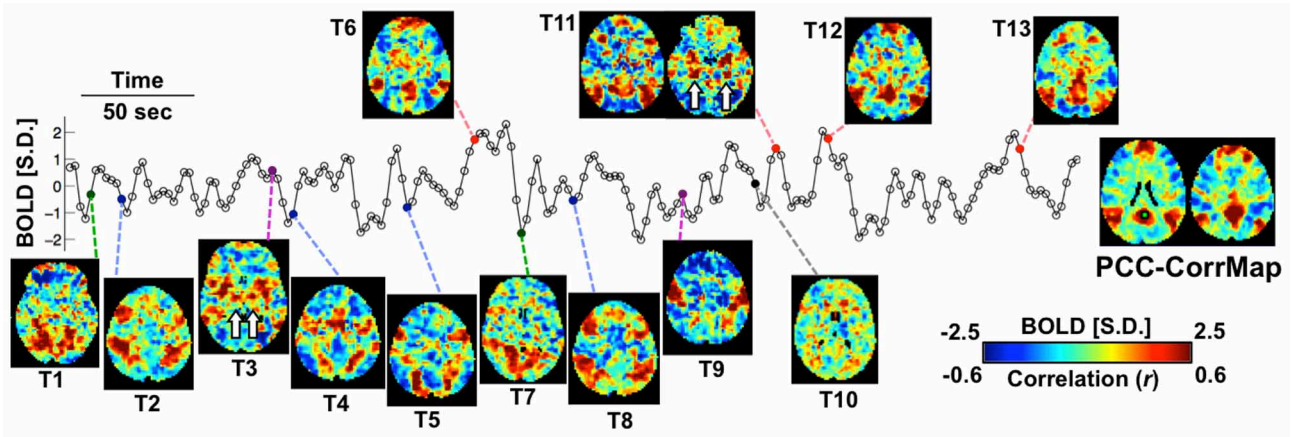


FIGURE 1 | Spontaneous activity at single fMRI time frames is suggestive of varying network involvement. Sample frames (T1-T13) from a representative fMRI scan show clear, distinct spatial patterns, some of which (red points) resemble the DMN pattern demonstrated by the PCC-based correlation map of this subject (right). The fMRI time course represents signal from the PCC seed region (green square). Small structures indicated by white arrows are the hippocampus (HI) and posterior parahippocampal gyrus (PHG) in T11 and the ventral lateral nucleus (VL) in T3. The color bar represents normalized BOLD signal for the sample frames and correlation values for the seed-based correlation map.

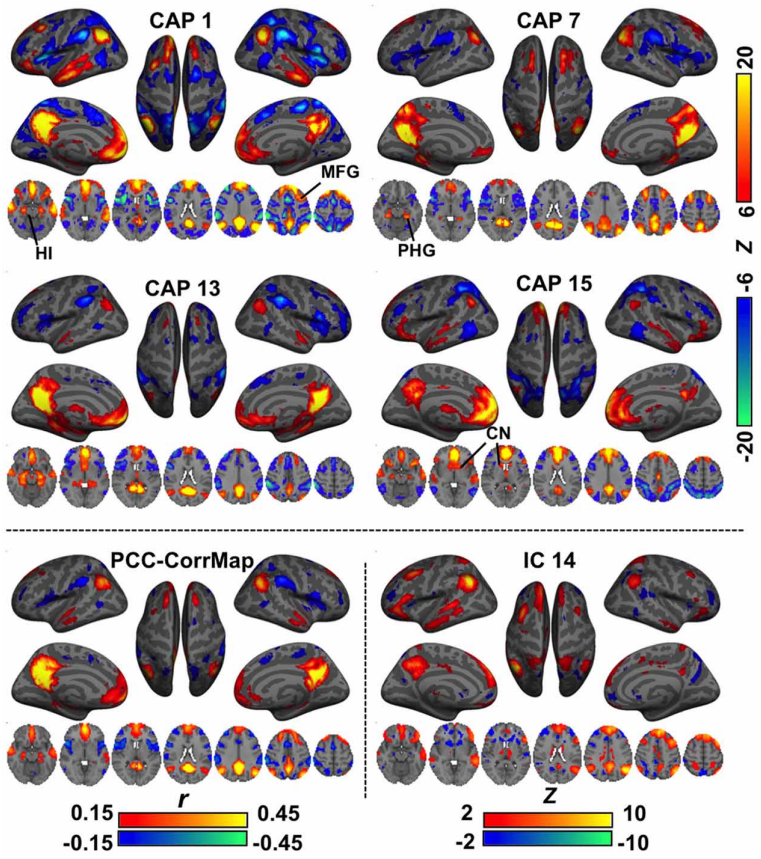


FIGURE 2 | Comparison of CAPs with the DMN derived from conventional analysis methods. CAPs 1,7,13,15 coarsely resemble the DMN pattern, but differ in detail. For example, CAP 1, 3, and 15 show specific co-activations at the hippocampus (HI), parahippocampal gyrus (PHG), and caudate nucleus (CN), respectively. For comparison, the DMN was also derived from PCC-based correlation (PCC-CorrMap, bottom left) and ICA (IC 14, having the highest spatial correlation with the PCC-CorrMap, bottom right). Both methods failed to extract multiple patterns of DMN seen in the CAPs.

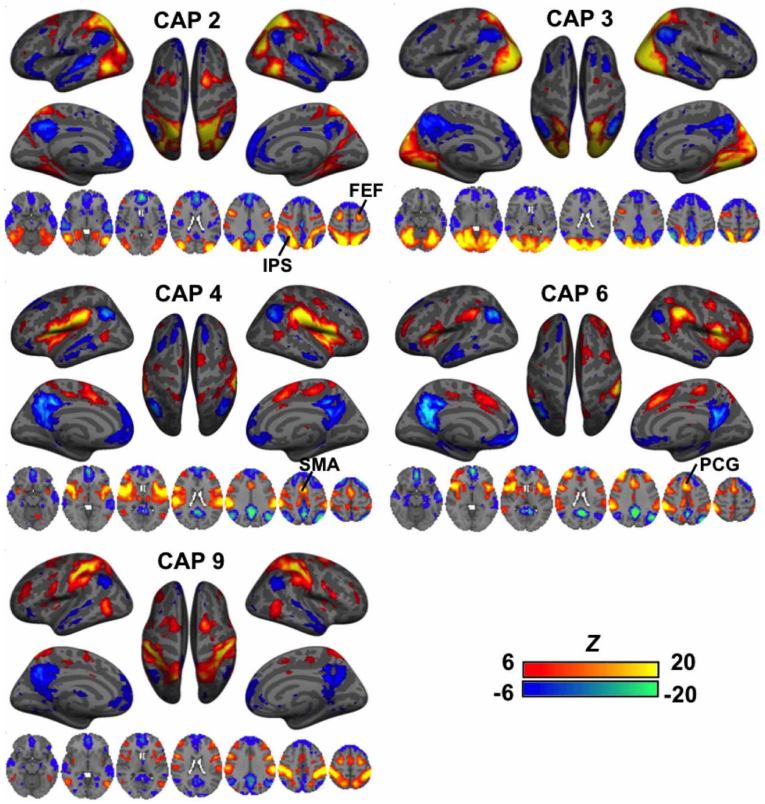


FIGURE 3 | CAPs showing distinct co-activation in "task-positive" regions (i.e., activated with common tasks) but consistent de-activation in the "task-negative" areas (i.e., de-activated with common tasks). Their patterns appeared more distinct from one another as compared to the DMN-activated CAPs. While CAPs 2 and 3 are both related to the visual system, CAP 2 overlaps more on the high order visual area and frontal eye fields (FEF). CAPs 4 and 6 looks almost identical in some axial slices, but a close inspection reveals that CAP 4 covers the central opercular cortex (CO), parietal operculum (PO), insular, thalamus, and supplementary motor area (SMA), whereas CAP 6 involves more anterior regions, including the frontal operculum (FO) and paracingulate gyrus (PCG). Moreover, even though all these CAPs show activations in the IPS, they cover very different sub-sections.

for primary visual cortex, includes bilateral focal structures that are much more inferior, lateral, and anterior than the thalamic structures seen in CAPs 2 and 3. We tentatively attribute these focal structures to the lateral geniculate nuclei (LGN). In contrast to the vision-related CAPs, thalamic nuclei in sensorimotor-related CAPs tend to show activity opposite to the cortical regions. Strong de-activations (or activations) are observed in the anterior and medial dorsal nuclei (AN/MDN, big arrows in **Figure 4B**), and in some regions surrounding posterior thalamus (small arrows), when the sensorimotor cortex shows activation (resp. de-activation) in CAP 19 (resp. CAP 8). These regions are much less distinct (for the AN/MDN) or even absent (for the thalamus surrounding areas) in maps derived by seed-based correlation or ICA (note: the display thresholds for these latter maps were adjusted, base on the whole brain, to include approximately the same spatial extent of significantly negative regions). It should also be noted that a pair of thalamic nuclei showing co-activation with cortical regions could also be clearly seen in CAP 19 when lowering the display threshold (**Figure 4C**). This pair of nuclei is attributed to the ventral posterolateral nuclei (VPL), located about 6 mm posterior to the ventral posteromedial nucleus (VL) shown in CAP 23 (**Figure 4C**).

The 30 CAPs are available at our website (http://amri.ninds.nih.gov/pub/xiao/CAPs_30_2mm.nii.gz).

OCCURRENCE RATE OF CAPs

An interesting aspect of the proposed analysis approach is that it not only provides spatial maps of co-activating brain regions, but also allows extraction of incidence rates of spatially distinct co-activations, information that is not explicitly available with conventional analysis methods. This incidence or occurrence rate of CAPs may facilitate distinction between subject populations. To illustrate this, we examined the possibility of distinguishing between spontaneous brain activity in males ($n = 38$) and females ($n = 64$) based on differences in occurrence rates of all 30 CAPs. For this purpose, the occurrence rate of a CAP was determined from the fraction of total volumes that were classified into the CAP's associated cluster.

CAP 23, which primarily covered sensorimotor areas corresponding to head regions, was found to occur more frequently ($p < 0.01$, Bonferroni corrected, permutation test) in the male members than in the female members (**Figure 5**). However, the males (mean motion: 0.048 ± 0.021 mm) had a little more ($p = 0.038$, 2-sample t-test) head motion than the females (mean

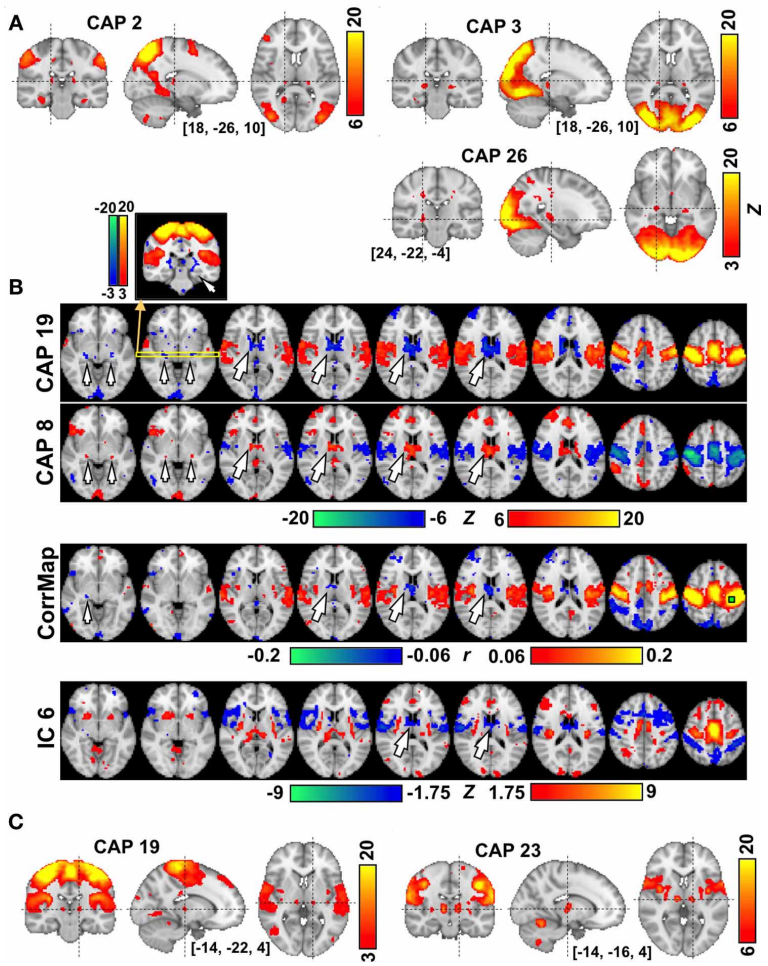


FIGURE 4 | The involvement of thalamic nuclei in CAPs. (A) CAPs in visual areas show co-activations in the LGN (CAP 26) and in different subdivisions of the Pulvinar (CAPs 2-3) each with different involvement of visual cortex. **(B)** CAPs in sensori-motor areas showing anti-phase activity at the AN/MDN of the thalamus (big arrows) and some areas surrounding the posterior thalamus (small arrows). The correlation map (CorrMap) seeded in the motor cortex (green square) and IC 6 less clearly shows anti-phase activity in these thalamic regions. Note: the display thresholds for CorrMap and IC 6 were adjusted in order to display approximately the same number of negative voxels as in CAP 19. **(C)** Sensorimotor-related CAPs also show co-activations at the thalamic nuclei, including the VPL (CAP 19) and VL (CAP 23), which are separated by only a few millimeters.

motion: 0.040 ± 0.018 mm), which may have affected this result (Van Dijk et al., 2012). To exclude this confounding factor, the same comparison was repeated between 30 males and 39 females whose mean motions exceeded 0.03 mm, which resulted in subgroups with insignificant difference in their mean motion ($p = 0.25$) (we were unable to generate subgroups with similar member counts based on upper threshold for motion). These two subgroups continued to show a significant difference in occurrence rates of CAP 23 (Figure S6).

COMPARISON TO OTHER DATA-DRIVEN APPROACHES

A more comprehensive comparison of CAPs with independent components (ICs) derived from the same dataset is presented in Figure S7 and **Figures 6A,B**. ICs typically cover a small but specific set of brain areas with very high z-statistics. As a result of the spatial independence criterion, there is little overlap of the regions having high z statistics in different ICs (Figure S7), resulting in a spatial correlation very close to zero (**Figure 6A**). In contrast, several CAPs, particularly those related to DMN or task-positive regions (Figure S3, CAPs 1,7,12,15, and CAPs 2,3,4,6,9, respectively), cover relatively large brain areas and show substantial spatial correlations with one another (**Figure 6A**). The CAPs in this category do not have clear correspondence to the patterns seen in the 30 ICs. In contrast, the CAPs covering small but specific regions, e.g., those related to primary sensory systems, tended to more closely resemble ICs (Figure S7, CAPs 16,19,23,26,29). Judging from their mutual spatial cross-correlation, at least a third of the CAPs did not pair up with any of the ICs (**Figure 6A**). Another noticeable difference between CAPs and ICs is asymmetric tails observed for the distribution of ICA map statistics (**Figure 6B**), which corresponds to the lack of significantly negative values in the ICs (Figure S7).

Recently, temporal ICA was applied to the components generated from spatial ICA in order to generate a set of temporal

functional modes (TFMs) (Smith et al., 2012). Similarity between TFMs and CAPs may be expected, since temporal independence between TFMs is, to some degree, similar to the exclusive relation of a time point to a specific CAP. To investigate this, we performed a spatial similarity analysis comparing the 30 CAPs with 21 TFMs derived in Smith et al. (2012). Although there were several CAP—TFM pairs with high similarity (CAP 2 and TFM 8, CAP 12 and TFM 19, and CAP 26 and TFM 2), most CAPs did not have a matching TFM (**Figures 6A,C**).

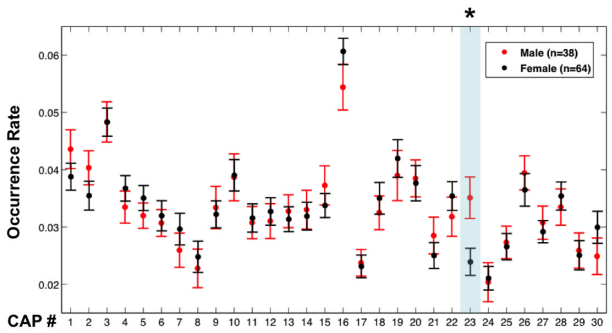


FIGURE 5 | Occurrence rates of the CAPs in males and females. Error bars represent the standard error across participants. The asterisk and light blue shadow indicates CAPs showing a significant difference in their occurrence ($p < 0.01$ with Bonferroni correction, permutation test) between these two groups.

DISCUSSION

In this report, we demonstrated a method to extract brain functional information by identifying regions that spontaneously co-activate during resting conditions. The method is based on the notion that co-activation of functionally related brain regions may not be a continual, stationary phenomenon (is implicitly assumed in conventional analysis) but rather occur sporadically over the course of a few seconds. The method also extends its previous version of seed-based analysis targeting specific networks (Liu and Duyn, 2013) to a data-driven, whole-brain approach.

By applying this approach to a large dataset of resting state fMRI studies, we identified at least 30 reproducible, spatially distinct co-activation patterns (CAPs). These CAPs in some aspects resemble canonical RSNs and TFMs extracted with ICA, but also

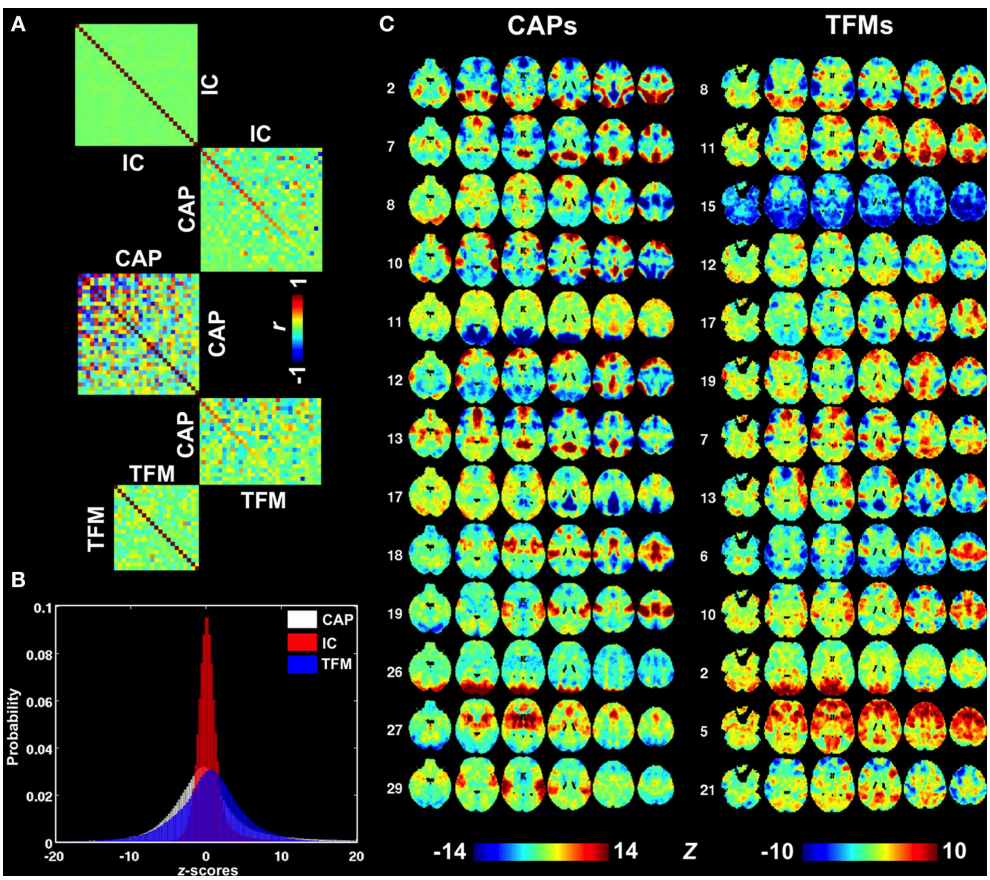


FIGURE 6 | Comparison between CAPs, ICs, and TFMs. (A) Spatial cross correlation matrices indicating substantial differences between CAPs (middle) on one hand and ICs and TFMs on the other (top and bottom, respectively). Cross-modality correlations were sorted based on correlation strength. **(B)** Distributions of map statistics (z-statistics) for CAPs (white), ICs (red), and TFMs (blue). IC statistics show a very non-Gaussian distribution with a elevated right tail, consistent with the independence assumption of ICA. **(C)** A selection of best-matched CAPs and TFMs, based on spatial similarity.

show distinct differences in various brain regions. These differences are partly attributed to methodological differences, for example the fact that the proposed method does not assume statistical independence between CAPs, as is the case for RSNs and TFMs extracted with ICA. Thus, the proposed method may lead to an alternative interpretation of the NNs underlying resting-state fMRI signals.

DISTINCT CO-ACTIVATION PATTERNS AND DYNAMIC FUNCTIONAL CONNECTIVITY

A series of recent studies has attempted to reveal dynamic changes in RSNs by examining the temporal variability in fMRI signal correlations (Chang and Glover, 2010; Allen et al., 2012; Hutchison et al., 2012; Rack-Gomer and Liu, 2012). Considerable variations in RSN characteristics were found even over the duration of a typical (several minute) resting scan, suggesting that the underlying NNs may dynamically assemble and disassemble over this time window. Such temporal variations may pose significant challenges for correlational approaches for the extraction of RSNs, as it may render them strongly dependent on the choice of analysis window. For example, the PCC-seeded correlation map within a 16.1-s window (only 7 time points) including T11 shows a DMN pattern with high correlations at the HI and the posterior PHG (Figure S8) because they co-activated at T11 (**Figure 1**); their functional connectivity to the PCC seed, however, nearly vanishes when the time window is shifted forward by 23 s to cover T12 instead. Likewise, the thalamic nucleus seen in T3 is only present in correlation maps seeded at the motor cortex for time windows including T3 (Figure S8). This dependence on position of the analysis window can be mitigated or avoided by increasing the window duration, which leads to more stable, reproducible RSN, but on the other hand can result in the loss of network information contained in single time frames.

INFORMATION IN CAPs

Using the method proposed in this study, the analysis window was effectively reduced to include only a single time point, and sensitivity was subsequently improved through selectively averaging the fMRI volumes with similar patterns. By averaging selectively rather than continually (as implicitly occurs with correlation analysis), one may extract finer detail regarding spontaneous co-activations of multiple brain regions. For example, correlation analysis with a seed in the PCC and a typical analysis window of a few minutes generally reveals a functional connection with HI and posterior PHG but little other information. In contrast, the multiple DMN-related CAPs extracted with the proposed method suggests that the two regions at times independently connect to the PCC. A possible explanation for these regions' differential involvement with DMN is their distinct roles in memory storage and retrieval (Gluck and Myers, 1997). Similarly, the varying involvement of sensory regions with the DMN apparent from CAPs (**Figure 3**) may signify that the brain may not be simply organized into two anti-correlated NNs spanning "task-positive" and "task-negative" regions, as concluded previously from correlation analysis (Fox et al., 2005).

Many CAPs included small structures in the thalamus and cerebellum that appeared to localize to specific anatomic subregions. The vision-related CAPs included thalamic nuclei that were well-separated and corresponded to established targets of functional connections with specific regions of the visual cortex that co-occurred in the CAPs. Similarly, the sensorimotor CAPs corresponding to the head (CAP 23) and lower body (CAP 19) regions showed specific co-activations at the VL and VPL, respectively, despite their limited anatomical separation of only a few millimeters. Anti-phase interaction was also observed between thalamic nuclei and sensorimotor cortex in CAPs 8 and 19. The AN/MDN in these CAPs has been shown to positively correlate with the alpha-band (8-13 Hz) electroencephalography (EEG) power, which is, in turn, negatively correlated with multiple sensory regions, including the sensorimotor cortex (Liu et al., 2012). These thalamic regions have non-specific connections to the cortex and may participate in the process of alertness and arousal (Van Der Werf et al., 2002). The second set of thalamic regions showing negative correlation to the sensorimotor cortex was around the LGN. Lowering the display threshold exposed multiple diffuse areas encompassing the posterior thalamus, which we suspects to be the thalamic reticular nucleus (TRN), partly because the anti-phase activity with sensorimotor cortex is, to some extent, consistent with the established inhibitory influence of TRN over other thalamic nuclei.

A few CAPs included isolated cortical structures that were distant from regions of major activity. For example, CAP 16, which includes areas associated with peripheral vision, showed small co-activation in the SMA, motor cortex, and medial IPS. This is plausible considering the critical role of the medial IPS in visuomotor coordinate transformation (Grefkes et al., 2004). These subtle aspects of CAPs are generally not captured with ICA or correlation analysis, indicating potential advantages of the proposed analysis approach.

Since hemodynamic response function (HRF) has been demonstrated to vary spatially (Wu et al., 2013), a potential concern is that different CAPs may actually represent the same neuronal co-activation event at different hemodynamic delays. This is, however, unlikely for the following two reasons. First, the fact that CAPs, particularly those covering similar or close cortical regions, show specific co-activations in distinct thalamic and/or cerebellar structures and that many of these CAP patterns are consistent with known organization of the brain argue strongly against their origin being attributable to vascular effects (e.g., hemodynamic delay differences between regions). Secondly, the analysis of temporal precedence of CAPs failed to find any pair of CAPs occurring in a specific order (Figure S9), which is also inconsistent with the confounding effect of the HRF delays. Conversely, co-activation in the same CAP does not guarantee the exact synchronization of neural activity at corresponding regions. Therefore, the "co-activation" in CAP refers to synchronization of fMRI signals rather than that of the underlying neural activity.

We found both co-activation and co-deactivation CAPs for some brain regions (e.g., CAP 8 and CAP 19), which are actually averages around fMRI signal peaks and troughs, respectively. Although these two types of CAPs often show very similar spatial patterns with reversed sign, a close inspection revealed that the co-activation at the fMRI peaks is significantly more synchronized than the co-deactivation at the troughs (Figure S10). The

implication of this observation remains unclear, but the peaks and troughs are clearly corresponding to distinct cortical states.

An attractive aspect of the proposed analysis is that it provides a simple quantifiable measure of CAP occurrence rate that facilitates comparison between subject populations. As example of this, we demonstrated that CAP 23, which overlaps the sensorimotor cortex, occurs more frequently in males than in females. Although far from providing a physiological explanation, this finding is consistent with a previous study reporting sex differences in the fluctuation amplitude of resting-state fMRI signals in similar brain regions (Biswal et al., 2010). We also make publicly available the 30 CAPs derived in our study, which can be directly used as templates for future studies intending to examine and compare their occurrence rates in different groups or under different conditions.

COMPARISON WITH CONVENTIONAL METHODS

As mentioned above, the extraction of CAPs based on single-time frame analysis may allow one to better capture temporally localized spontaneous co-activations as compared to conventional correlation analysis. The method presented here resembles spatial ICA in the sense that both are spatial-domain methods that regard fMRI time frames as the basic units of analysis. However, ICA assumes that the fMRI volumes comprise weighted combinations of a set of statistically ICs, the best fit of which can be found by maximizing the independence of the latter. The independence assumption yields ICs that typically cover a relatively small extent of the brain with very high statistical scores (Figure S7), because such a pattern has high non-Gaussianity (**Figure 6B**), an important factor underlying ICA based signal separation. At the same time, independence maximization yields ICs with minimal pairwise spatial correlations, as shown in **Figure 6A**. Together, these two features suggest that the ICs present a spatial parcellation of the brain rather than distinct states of functional connectivity. In addition, ICs may exhibit considerable pairwise *temporal* correlation as exemplified by the often negative correlation observed between the ICs covering task-positive and task-negative regions.

To overcome these limitations and gain more knowledge about the brain's spontaneous co-activations, temporal ICA has been further applied to spatial ICs to recombine them into a set of TFMs (Smith et al., 2012). The TFMs are perhaps more closely related to the CAPs, since the concept of temporal independence is similar to the exclusive presence of a given CAP at any given time. However, a direct comparison between TFMs and CAPs indicated substantial differences (**Figure 6**). This may not be surprising, given the rather different methodology and assumptions used to generate these network measures.

A practical advantage of the proposed method is that it involves few assumptions and transformations of the data. The classification procedure (clustering) does not perform any transformation on the data, and the CAPs are simply averages of time frames classified into groups based on spatial activity patterns. Therefore, they are easy to interpret and reflect the underlying brain activity in a rather direct way.

NEURAL ORIGIN OF CAPs

Since the co-activation patterns are clear even in single fMRI volumes, they likely reflect large-scale neural activity occurring within brief time periods. One candidate is large-scale neuronal avalanching activity, defined as spontaneous activity initiating at a specific location in the form of a brief burst, and taking with it, like an avalanche, other connected regions that are near activation threshold (Beggs and Plenz, 2003; Tagliazucchi et al., 2012). The previous analysis on spatiotemporal dynamics of spontaneous fMRI signals confirmed its scale-free property, consistent with avalanche types of activity at smaller time-scales (Tagliazucchi et al., 2012). However, it is also possible that other types of neuronal activity underlie such brief co-activations of the brain. For example, since the dataset we analyzed here was acquired during wakefulness, it is possible that neural activity associated certain conscious processes, e.g., mind wandering, may also contribute to spontaneous co-activations of the brain. The analysis of fMRI signals acquired under brain states with reduced consciousness would help to partially clarify these issues. Nevertheless, uncovering the precise neuronal origin and functional relevance of the CAPs will likely require acquisition of electrophysiological data concurrently acquired with the fMRI.

TECHNICAL LIMITATIONS

A challenge for clustering analysis is selection of k, i.e., the number of CAPs to be extracted from the data. The current value of 30 was chosen somewhat arbitrarily from a comparison of results obtained with $k = 20$, 30, and 40. We have also attempted to select k by evaluating the corresponding cluster structures with more objective indices, including the Silhouette, Calinski-Harabasz, Davies-Bouldin methods, which, however, yielded inconsistent recommendations that also tend to underestimate the number of CAPs present in the data. This may be due to the two following reasons: first, the distance (dissimilarity) among the CAPs is likely to have a very skewed distribution, with a portion of the CAPs being much closer to one another than to the others (**Figure 6A**), which increases the difficulty in finding a clear division. Secondly, the noisy fMRI volumes mentioned above will further blur the boundaries between different CAPs. Nevertheless, the CAPs in the current dataset were found to be rather insensitive to this parameter, as the majority of the CAPs shown for the case of $k = 30$ were also observed in the cases of $k = 20$ and $k = 40$.

The computational load of the k-means clustering increases quickly with the number of fMRI volumes (time points). Therefore, the application of the proposed method to fMRI datasets with very high sampling rate, e.g., those acquired with the newly developed multi-band technique, would either be limited to a relatively small population or have to require more computational resources and/or more efficient computational algorithms. This could be a challenge for future studies attempting to utilize the ultra-high temporal resolution of fMRI data for special research purposes.

ACKNOWLEDGMENTS

This research was supported (in part) by the Intramural Research Program of the NIH, NINDS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fnsys.2013.00101/abstract

REFERENCES

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., et al. (2012). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex.* doi: 10.1093/cercor/bhs352. [Epub ahead of print].

Anderson, J. S., Ferguson, M. A., Lopez-Larson, M., and Yurgelun-Todd, D. (2010). Topographic maps of multisensory attention. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20110–20114. doi: 10.1073/pnas.1011616107

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. doi: 10.1126/science.273.5283.1868

Beckmann, C. F., Deluca, M., Devlin, J. T., and Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1001–1013. doi: 10.1098/rstb.2005.1634

Beggs, J. M., and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177.

Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739. doi: 10.1073/pnas.0911855107

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Chang, C., and Glover, G. H. (2010). Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 50, 81–98. doi: 10.1016/j.neuroimage.2009.12.011

Cordes, D., Haughton, V., Carew, J. D., Arfanakis, K., and Maravilla, K. (2002). Hierarchical clustering to measure connectivity in fMRI resting-state data. *Magn. Reson. Imaging* 20, 305–317. doi: 10.1016/S0730-725X(02)00503-9

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., et al. (2000). Mapping functionally related regions of brain with functional connectivity MR imaging. *AJNR Am. J. Neuroradiol.* 21, 1636–1644.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., and Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10046–10051. doi: 10.1073/pnas.0604187103

Fox, M. D., and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* 8, 700–711. doi: 10.1038/nrn2201

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102

Fox, M. D., Zhang, D., Snyder, A. Z., and Raichle, M. E. (2009). The global signal and observed anticorrelated resting state brain networks. *J. Neurophysiol.* 101, 3270–3283. doi: 10.1152/jn.90777.2008

Gluck, M. A., and Myers, C. E. (1997). Psychobiological models of hippocampal function in learning and memory. *Annu. Rev. Psychol.* 48, 481–514. doi: 10.1146/annurev.psych.48.1.481

Grefkes, C., Ritzl, A., Zilles, K., and Fink, G. R. (2004). Human medial intraparietal cortex subserves visuomotor coordinate transformation. *Neuroimage* 23, 1494–1506. doi: 10.1016/j.neuroimage.2004.08.031

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258. doi: 10.1073/pnas.0135058100

Hampson, M., Peterson, B. S., Skudlarski, P., Gatenby, J. C., and Gore, J. C. (2002). Detection of functional connectivity using temporal correlations in MR images. *Hum. Brain Mapp.* 15, 247–262. doi: 10.1002/hbm.10022

Hutchison, R. M., Womelsdorf, T., Gati, J. S., Everling, S., and Menon, R. S. (2012). Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. *Hum. Brain Mapp.* 34, 2154–2177. doi: 10.1002/hbm.22058

Kelly, C., Toro, R., Di Martino, A., Cox, C. L., Bellec, P., Castellanos, F. X., et al. (2012). A convergent functional architecture of the insula emerges across imaging modalities. *Neuroimage* 61, 1129–1142. doi: 10.1016/j.neuroimage.2012.03.021

Kiviniemi, V., Kantola, J. H., Jauhiainen, J., Hyvarinen, A., and Tervonen, O. (2003). Independent component analysis of nondeterministic fMRI signal sources. *Neuroimage* 19, 253–260. doi: 10.1016/S1053-8119(03)00097-1

Leopold, D. A., and Maier, A. (2012). Ongoing physiological processes in the cerebral cortex. *Neuroimage* 62, 2190–2200. doi: 10.1016/j.neuroimage.2011.10.059

Liu, X., and Duyn, J. H. (2013). Time-Varying Functional Network Information Extracted from Brief Instances of Spontaneous Brain Activity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4392–4397. doi: 10.1073/pnas.1216856110

Liu, Z., De Zwart, J. A., Yao, B., Van Gelderen, P., Kuo, L. W., and Duyn, J. H. (2012). Finding thalamic BOLD correlates to posterior alpha EEG. *Neuroimage* 63, 1060–1069. doi: 10.1016/j.neuroimage.2012.08.025

Mezer, A., Yovel, Y., Pasternak, O., Gorfine, T., and Assaf, Y. (2009). Cluster analysis of resting-state fMRI time series. *Neuroimage* 45, 1117–1125. doi: 10.1016/j.neuroimage.2008.12.015

Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B., and Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage* 44, 893–905. doi: 10.1016/j.neuroimage.2008.09.036

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S.-G., Merkle, H., et al. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5951–5955. doi: 10.1073/pnas.89.13.5951

Rack-Gomer, A. L., and Liu, T. T. (2012). Caffeine increases the temporal variability of resting-state BOLD connectivity in the motor cortex. *Neuroimage* 59, 2994–3002. doi: 10.1016/j.neuroimage.2011.10.001

Saad, Z. S., Glen, D. R., Chen, G., Beauchamp, M. S., Desai, R., and Cox, R. W. (2009). A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *Neuroimage* 44, 839–848. doi: 10.1016/j.neuroimage.2008.09.037

Saad, Z. S., Gotts, S. J., Murphy, K., Chen, G., Jo, H. J., Martin, A., et al. (2012). Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. *Brain Connect.* 2, 25–32. doi: 10.1089/brain.2012.0080

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23(Suppl. 1), S208–S219. doi: 10.1016/j.neuroimage.2004.07.051

Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J., Woolrich, M. W., et al. (2012). Temporally-independent functional modes of spontaneous brain activity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3131–3136. doi: 10.1073/pnas.1121329109

Tagliazucchi, E., Balenzuela, P., Fraiman, D., and Chialvo, D. R. (2012). Criticality in large-scale brain FMRI dynamics unveiled by a novel point process analysis. *Front. Physiol.* 3:15. doi: 10.3389/fphys.2012.00015

Van Der Werf, Y. D., Witter, M. P., and Groenewegen, H. J. (2002). The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. *Brain Res. Brain Res. Rev.* 39, 107–140. doi: 10.1016/S0165-0173(02)00181-9

Van Dijk, K. R., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., and Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J. Neurophysiol.* 103, 297–321. doi: 10.1152/jn.00783.2009

Van Dijk, K. R., Sabuncu, M. R., and Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage* 59, 431–438. doi: 10.1016/j.neuroimage.2011.07.044

Wu, G. R., Liao, W., Stramaglia, S., Ding, J. R., Chen, H., and Marinazzo, D. (2013). A blind deconvolution approach to recover effective connectivity brain

networks from resting state fMRI data. *Med. Image Anal.* 17, 365–374. doi: 10.1016/j.media.2013.01.003

Zhang, D., and Raichle, M. E. (2010). Disease and the brain's dark energy. *Nat. Rev. Neurol.* 6, 15–28. doi: 10.1038/nrneurol.2009.198

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 July 2013; accepted: 15 November 2013; published online: 04 December 2013.

*Citation: Liu X, Chang C and Duyn JH (2013) Decomposition of spontaneous brain activity into distinct fMRI co-activation patterns. Front. Syst. Neurosci. **7**:101. doi: 10.3389/fnsys.2013.00101*

This article was submitted to the journal Frontiers in Systems Neuroscience.

frontiers in Systems Neuroscience

OPINION
published: 04 April 2017
doi: 10.3389/fnsys.2017.00015

Cortical Variability and Challenges for Modeling Approaches

Emili Balaguer-Ballester[1,2]*

[1] Department of Computing and Informatics, Faculty of Science and Technology, Bournemouth University, Bournemouth, UK, [2] Bernstein Center for Computational Neuroscience, Medical Faculty Mannheim and Heidelberg University, Mannheim, Germany

Keywords: metastability, excitation-inhibition balance, cortical variability, neuronal variability, recurrent networks, trial to trial variability

The functional role of the observed neuronal variability (the disparity in neural responses across multiple instances of the same experiment) is again receiving close attention in Computational and Systems Neuroscience (e.g., Durstewitz et al., 2010; Moreno-Bote et al., 2011; Oram, 2011; Beck et al., 2012; Churchland and Abbott, 2012; Brunton et al., 2013; Masquelier, 2013; Mattia et al., 2013; Balaguer-Ballester et al., 2014; Renart and Machens, 2014; Bujan et al., 2015; Lin et al., 2015; Pachitariu et al., 2015; Arandia-Romero et al., 2016; Doiron et al., 2016; McDonnell et al., 2016). Special consideration is currently given to understanding how spiking (Bujan et al., 2015; Deneve and Machens, 2016; Doiron et al., 2016; Hartmann et al., 2016; Landau et al., 2016) and phenomenological (Goris et al., 2014; Lin et al., 2015; Mochol et al., 2015; Arandia-Romero et al., 2016; Doiron et al., 2016) models account for the wide range of classical and new phenomena associated with trial-to-trial uncorrelated activity.

Specifically, it has often been proposed that a network state characterized by largely asynchronous spike times whilst maintaining slow oscillations in the firing-rates, may represent the default spontaneous cortical mode (e.g., Sanchez-Vives and Mattia, 2014; Deneve and Machens, 2016; Sancristobal et al., 2016); and similar states could also underlie observed stimulus-driven variability in rate (Litwin-Kumar and Doiron, 2012; Deneve and Machens, 2016; Hartmann et al., 2016). However, the way in which such a computationally advantageous network state for neural coding is achieved can differ substantially between modeling approaches; this challenge will be the focus of this manuscript.

PREDICTABLE COMPONENTS OF NEURONAL VARIABILITY

The view that the intrinsic stochasticity of single cell activity is the major source of variability has been questioned multiple times over the last decades by modeling (van Vreeswijk and Sompolinsky, 1996; Amit and Brunel, 1997; Shadlen and Newsome, 1998; Deneve et al., 2001; Stein et al., 2005; Faisal et al., 2008; Renart et al., 2010; Rabinovich and Varona, 2011; Masquelier, 2013; Stiefel et al., 2013; Rabinovich et al., 2014; Deneve and Machens, 2016; Hartmann et al., 2016) and empirical studies (e.g., Bryant and Segundo, 1976; Mainen and Sejnowski, 1995; Britten et al., 1996; Stein et al., 2005). It is well known that essentially deterministic networks of balanced excitation and inhibition are able to generate a weakly correlated, often chaotic attractor state which presents Poissonian statistical properties like the observed activity (van Vreeswijk and Sompolinsky, 1996; Amit and Brunel, 1997; Shadlen and Newsome, 1998; Sussillo and Abbott, 2009; Litwin-Kumar and Doiron, 2012). However, such a chaotic state is a non-mandatory modeling choice: recently, a range of models has shown that part of the observed variability may also be explained by a different class

OPEN ACCESS

Edited by:
Gilad Silberberg, Karolinska Institutet, Sweden

Reviewed by:
Maurizio Mattia, Istituto Superiore di Sanità, Italy
Arvind Kumar, Royal Institute of Technology, Sweden
Iñigo Arandia-Romero, Pompeu Fabra University, Spain

***Correspondence:**
Emili Balaguer-Ballester
eb-ballester@bournemouth.ac.uk

Received: *18 October 2016*
Accepted: *06 March 2017*
Published: *04 April 2017*

Citation:
Balaguer-Ballester E (2017) Cortical Variability and Challenges for Modeling Approaches. Front. Syst. Neurosci. 11:15. doi: 10.3389/fnsys.2017.00015

of deterministic processes (Beck et al., 2012; Mattia et al., 2013; Renart and Machens, 2014; Bujan et al., 2015; Abbott et al., 2016; Deneve and Machens, 2016; Doiron et al., 2016; Gillary and Niebur, 2016; Hartmann et al., 2016) such as the lack of specificity in top-down processing of cognitively complex tasks (Beck et al., 2012).

At the same time, empirical studies found mounting evidence of deterministic patterns for some of the trial to trial variability. For example, a range of indexes (Shadlen and Newsome, 1998; Churchland and Abbott, 2012; Marcos et al., 2013) suggest that variance is systematically reduced at the stimulus onset (Churchland et al., 2010); and in general shows a predictable trend during different events of the task (Churchland et al., 2006, 2010; Churchland and Abbott, 2012; Ledberg et al., 2012; Renart and Machens, 2014). Thus, there seems to be an increasing consensus in that at least part of the trial to trial variability shows a deterministic pattern which may play a functional role; and hence cannot be simply neglected (Balaguer-Ballester et al., 2011, 2014; Masquelier, 2013; Ecker et al., 2014; Goris et al., 2014; Renart and Machens, 2014; Lin et al., 2015; Schölvinck et al., 2015; Arandia-Romero et al., 2016; Hartmann et al., 2016).

Nevertheless, despite such recent advances, the mapping between the cognitive state and variability is still challenging. For instance, on the one hand, correlated rate variability between pairs of neurons is often reduced by top-down attentional processes (e.g., Cohen and Maunsell, 2009; Mitchell et al., 2009; Cohen and Kohn, 2011; Doiron et al., 2016). On the other hand, the opposite can be observed when attention is highly variable across trials (Roelfsema et al., 2004; Renart and Machens, 2014; Ruff and Cohen, 2014); and such noise correlation analyses show a variety of mixed results (Cohen and Kohn, 2011; Eyherabide and Samengo, 2013; Moreno-Bote et al., 2014; Ruff and Cohen, 2014, 2016; Doiron et al., 2016).

Importantly, compelling evidence suggests that a substantial portion of the spontaneous and evoked total and shared variability is attributable to global fluctuations (Ecker et al., 2014, 2016; Goris et al., 2014; Mochol et al., 2015; Pachitariu et al., 2015; Schölvinck et al., 2015; Arandia-Romero et al., 2016); and this has direct implications in neural coding in visual (Lin et al., 2015; Arandia-Romero et al., 2016; Ecker et al., 2016) and in auditory areas (Mochol et al., 2015; Pachitariu et al., 2015). For instance, high population activity in monkey V1 increases the information that a subset of neuronal ensembles carry about stimulus orientation, only the ones that show a strong multiplicative modulation. In contrast, the stimulus-decoding information of such multiplicative ensembles plummets for low global activity states; whilst information increases in the group additively-modulated neurons in the population (Arandia-Romero et al., 2016).

Global modulations could either stem from the default up/down state of ongoing activity (Mochol et al., 2015) or from fluctuations within a single state (Arandia-Romero et al., 2016). When controlled for this global co-modulations, noise correlations are often negligible (Renart et al., 2010), but not always (Pachitariu et al., 2015; Arandia-Romero et al., 2016). Moreover, stimulus-driven input statistics can also have a strong contribution to the observed evoked variability in parallel to the global network state (Oram, 2011; Bujan et al., 2015; Pachitariu et al., 2015; Doiron et al., 2016; Landau et al., 2016) and explain noise correlations dynamics (Bujan et al., 2015).

This complex variety of results has been recently analyzed using a range of phenomenological and spiking models. These recent modeling efforts aim to pin down when precisely during the course of the trial (Moreno-Bote et al., 2014; Bujan et al., 2015; Doiron et al., 2016) and in which specific network state (Arandia-Romero et al., 2016) noise correlations are informative or deleterious for neural coding (Ecker et al., 2014, 2016; Moreno-Bote et al., 2014; Lin et al., 2015; Pachitariu et al., 2015; Schölvinck et al., 2015; Arandia-Romero et al., 2016; Doiron et al., 2016).

DIVERSITY OF THEORETICAL APPROACHES

The consensus on the network origin of a substantial part of cortical variability led to the development of a multitude of models for explaining the underlying neuronal mechanisms of the asynchronous state (e.g., Boerlin et al., 2013; Deco et al., 2014; Ostojic, 2014; Barral and Reyes, 2016; Hartmann et al., 2016; Rosenbaum et al., 2017). A linking theme in these approaches is the crucial contribution of fast inhibition in recurrent networks; which is negatively correlated with excitation and strong enough to counterbalance it to different degrees (Renart et al., 2010; Deneve and Machens, 2016).

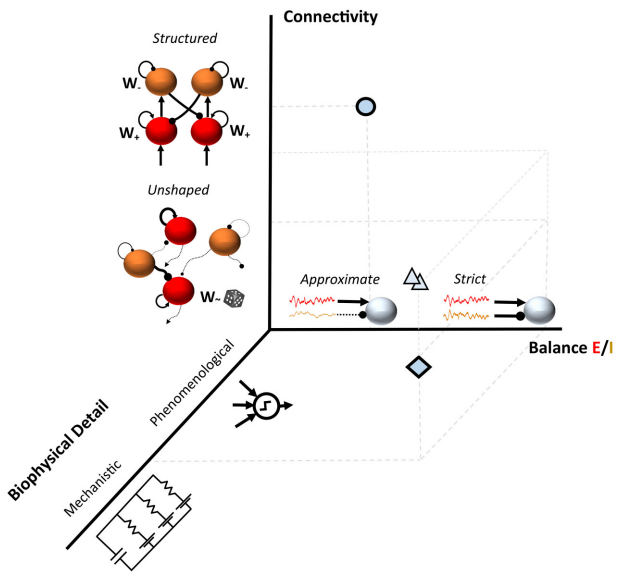


FIGURE 1 | Three *dimensions* in models of neuronal variability. The diamond shows an example of a spiking model with random connectivity and a tight Excitation/Inhibition (E/I) balance as in Renart et al., 2010; the circle represents a substantially different modeling choice such as in Hartmann et al., 2016. A range of modeling approaches typically fall between these two examples (triangles); such as semi-structured connectivity architectures which modulate the E/I balanced dynamics in realistic networks (Litwin-Kumar and Doiron, 2012; Landau et al., 2016).

This scenario is currently the subject of a lively debate; and a variety of processing architectures of spiking units have been developed to explain the observed variability phenomena from different perspectives. It has recently been proposed that a much tighter synchronization between excitation and inhibition than considered so far, at the spike level, has an even stronger experimental support and would enable the network to operate optimally by reducing the minimum coding error (Renart et al., 2010; Boerlin et al., 2013; Abbott et al., 2016; Deneve and Machens, 2016). The precise way in which the asynchronous state is achieved however, is not unique. For example, connectivity weights are specifically learnt in the design termed spike-coding network (Boerlin et al., 2013; Schwemmer et al., 2015; Abbott et al., 2016; Deneve and Machens, 2016); whilst connectivity is clustered in Litwin-Kumar and Doiron (2012), shaped by plasticity in Vogels et al. (2011) and Landau et al. (2016) and much less structured in many other dense (Renart et al., 2010; Abbott et al., 2016) or sparse (Ostojic, 2014) networks.

In contrast, other recent approaches rely on non-closely balanced excitatory-inhibitory dynamics in networks of simplified units. For instance, Hartmann et al. (2016) proposed a fully deterministic approach to describe spontaneous and stimulus evoked variability; consisting of an architecture of schematic noise-free units. In this approach, excitatory connectivity is specifically set by plasticity and homeostasis; and is not necessarily balanced. In Deco et al. (2014), the spontaneous state also stems from not necessarily tightly balanced architectures, where elements are field equations derived from spiking units with background input noise. In this and related models, connectivity is also set; but in the latter case local inhibition is regulated by a different homeostatic control. In addition, it has been recently shown that most of the evoked variability could be accounted for by essentially feedforward architectures (Bujan et al., 2015; Doiron et al., 2016).

Asymmetry and slightly unbalanced configurations are also considered to promote the so called metastable state (Mattia et al., 2013; Tognoli and Kelso, 2014a,b; Deco and Kringelbach, 2016), in which high dimensional ensembles flexibly re-organize, synchronize and disengage, possibly by changing their role in a high-dimensional setting (Lapish et al., 2015; Fusi et al., 2016). Such state also exhibits advantageous computational properties (Hellyer et al., 2015; Deco and Kringelbach, 2016). In an instantiation of such ideas, essentially deterministic structures of simplified units generate itinerancy through robust transient states. These states enable the model to process cognitive entities without the compelling need for attractors (Rabinovich et al., 2008, 2014; Varona and Rabinovich, 2016). Moreover, combinations of attractor-based and transient-based computations could also underlie motor plans (Mattia et al., 2013). Other state-dependent computational ingredients such as neuromodulation could also play a major role in shaping the observed variability (Mattia et al., 2013; Lapish et al., 2015; Doiron et al., 2016).

FUTURE CHALLENGES FOR MODELS

This non-exhaustive summary of few recent examples suggests the availability of a plethora of recurrent and feedforward network models for understanding the source of variability during cognitive processing and in the resting state. These configurations often differ at least in the level of detail of the computational units, in the connectivity structure and in the degree of balance between excitation and inhibition (**Figure 1**).

This challenging scenario perhaps compels to the development of novel approaches for probing the networks in order to identify the suitable architecture or architectures for each specific cognitive process and cortical area. However, the question remains how to effectively dissect a recurrent network, beyond the linearization of the network dynamics, in order to investigate the components originating the asynchronous state (Sussillo, 2014). Recently, Doiron et al. (2016) proposed a framework to identify the physiological processes underlying decorrelation in feedforward circuits by analysing state-dependent correlations in different time windows. However, applying this approach is more problematic in recurrent circuits when coupling is not weak and is highly nonlinear; as is often the case in models.

Hence, inferring which level of detail and architecture are mostly probably responsible for the neuronal variability phenomena is perhaps one of the major challenges for the next years, which possibly requires the development of novel theoretical tools for scrutinizing network behavior.

AUTHOR CONTRIBUTIONS

The author conceived and wrote the manuscript.

FUNDING

This paper has been funded by Bournemouth University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The author wants to thank very much the three reviewers for their valuable comments.

REFERENCES

Abbott, L. F., DePasquale, B., and Memmesheimer, R.-M. M. (2016). Building functional networks of spiking model neurons. *Nat. Neurosci.* 19, 350–355. doi: 10.1038/nn.4241

Amit, D. J., and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* 7, 237–252. doi: 10.1093/cercor/7.3.237

Arandia-Romero, I., Tanabe, S., Drugowitsch, J., Kohn, A., and Moreno-Bote, R. (2016). Multiplicative and additive modulation of neuronal tuning with

population activity affects encoded information. *Neuron* 89, 1305–1316. doi: 10.1016/j.neuron.2016.01.044

Balaguer-Ballester, E., Lapish, C., Seamans, J., and Durstewitz, D. (2011). Attracting dynamics of frontal cortex ensembles during memory guided decision making. *PLoS Comput. Biol.* 7:e1002057. doi: 10.1371/journal.pcbi.1002057

Balaguer-Ballester, E., Tabas-Diaz, A., and Budka, M. (2014). Can we identify non-stationary dynamics of trial-to-trial variability? *PLoS ONE* 9:e95648. doi: 10.1371/journal.pone.0095648

Barral, J., and Reyes, A. (2016). Synaptic scaling rule preserves excitatory-inhibitory balance and salient neuronal network dynamics. *Nat. Neurosci.* 19, 1690–1696. doi: 10.1038/nn.4415

Beck, J., Ma, W. J., Pitkow, X., Latham, P. E., and Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron* 74, 33–39. doi: 10.1016/j.neuron.2012.03.016

Boerlin, M., Machens, C. K., and Deneve, S. (2013). Predictive coding of dynamical variables in balanced spiking networks. *PLoS Comput. Biol.* 9:e1003258. doi: 10.1371/journal.pcbi.1003258

Britten, K., Newsome, W., Shadlen, M., Celebrini, S., and Movshon, J. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque mt. *Vis. Neurosci.* 13, 87–100. doi: 10.1017/S095252380000715X

Brunton, B., Botvinick, M., and Brody, C. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science* 340, 95–98. doi: 10.1126/science.1233912

Bryant, H. L., and Segundo, J. P. (1976). Spike initiation by transmembrane current: a white-noise analysis. *J. Physiol.* 260, 279–314. doi: 10.1113/jphysiol.1976.sp011516

Bujan, A. F., Aertsen, A., and Kumar, A. (2015). Role of input correlations in shaping the variability and noise correlations of evoked activity in the neocortex. *J. Neurosci.* 35, 8611–8625. doi: 10.1523/JNEUROSCI.4536-14.2015

Churchland, M. M., and Abbott, L. F. (2012). Two layers of neural variability. *Nat. Neurosci.* 15, 1472–1474. doi: 10.1038/nn.3247

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378. doi: 10.1038/nn.2501

Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G., and Shenoy, K. V. (2006). Neural variability in premotor cortex provides a signature of motor preparation. *J. Neurosci.* 26, 3697–3712. doi: 10.1523/JNEUROSCI.3762-05.2006

Cohen, M. R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14, 811–819. doi: 10.1038/nn.2842

Cohen, M. R., and Maunsell, J. H. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 12, 1594–1600. doi: 10.1038/nn.2439

Deco, G., and Kringelbach, M. L. (2016). Metastability and coherence: extending the communication through coherence hypothesis using a whole-brain computational perspective. *Trends Neurosci.* 39, 125–135. doi: 10.1016/j.tins.2016.01.001

Deco, G., Ponce-Alvarez, A., Hagmann, P., Romani, G. L., Mantini, D., and Corbetta, M. (2014). How local excitation–inhibition ratio impacts the whole brain dynamics. *J. Neurosci.* 34, 7886–7898. doi: 10.1523/JNEUROSCI.5068-13.2014

Deneve, S., Latham, P. E., and Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nat. Neurosci.* 4, 826–831. doi: 10.1038/90541

Deneve, S., and Machens, C. K. (2016). Efficient codes and balanced networks. *Nat. Neurosci.* 19, 375–382. doi: 10.1038/nn.4243

Doiron, B., Litwin-Kumar, A., Rosenbaum, R., Ocker, G. K., and Josic, K. (2016). The mechanics of state-dependent neural correlations. *Nat. Neurosci.* 19, 383–393. doi: 10.1038/nn.4242

Durstewitz, D., Vittoz, N. M., Floresco, S. B., and Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* 66, 438–448. doi: 10.1016/j.neuron.2010.03.029

Ecker, A., Berens, P., Cotton, R., Subramaniyan, M., Denfield, G., Cadwell, C., et al. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron* 82, 235–248. doi: 10.1016/j.neuron.2014.02.006

Ecker, A. S., Denfield, G. H., Bethge, M., and Tolias, A. S. (2016). On the structure of neuronal population activity under fluctuations in attentional state. *J. Neurosci.* 36, 1775–1789. doi: 10.1523/JNEUROSCI.2044-15.2016

Eyherabide, H. G., and Samengo, I. (2013). When and why noise correlations are important in neural decoding. *J. Neurosci.* 33, 17921–17936. doi: 10.1523/JNEUROSCI.0357-13.2013

Faisal, A. A., Selen, L. P., and Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.* 9, 292–303. doi: 10.1038/nrn2258

Fusi, S., Miller, E. K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* 37, 66–74. doi: 10.1038/nrn2258

Gillary, G., and Niebur, E. (2016). The edge of stability: response times and delta oscillations in balanced networks. *PLoS Comput. Biol.* 12:e1005121. doi: 10.1371/journal.pcbi.1005121

Goris, R. L., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nat. Neurosci.* 17, 858–865. doi: 10.1038/nn.3711

Hartmann, C., Lazar, A., Nessler, B., and Triesch, J. (2016). Where is the noise? key features of spontaneous activity and neural variability arise through learning in a deterministic network. *PLoS Comput. Biol.* 11:e1004640. doi: 10.1371/journal.pcbi.1004640

Hellyer, P. J., Scott, G., Shanahan, M., Sharp, D. J., and Leech, R. (2015). Cognitive flexibility through metastable neural dynamics is disrupted by damage to the structural connectome. *J. Neurosci.* 35, 9050–9063. doi: 10.1523/JNEUROSCI.4648-14.2015

Landau, I. D., Egger, R., Dercksen, V. J., Oberlaender, M., and Sompolinsky, H. (2016). The impact of structural heterogeneity on excitation-inhibition balance in cortical networks. *Neuron* 92, 1106–1121. doi: 10.1016/j.neuron.2016.10.027

Lapish, C. C., Balaguer-Ballester, E., Seamans, J. K., Phillips, A. G., and Durstewitz, D. (2015). Amphetamine exerts dose-dependent changes in prefrontal cortex attractor dynamics during working memory. *J. Neurosci.* 35, 10172–10187. doi: 10.1523/JNEUROSCI.2421-14.2015

Ledberg, A., Montagnini, A., Coppola, R., and Bressler, S. L. (2012). Reduced variability of ongoing and evoked cortical activity leads to improved behavioral performance. *PLoS ONE* 7:e43166. doi: 10.1371/journal.pone.0043166

Lin, I.-C., Okun, M., Carandini, M., and Harris, K. D. (2015). The nature of shared cortical variability. *Neuron* 87, 644–656. doi: 10.1016/j.neuron.2015.06.035

Litwin-Kumar, A., and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* 15, 1498–1505. doi: 10.1038/nn.3220

Mainen, Z. F., and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science (New York, N.Y.)* 268, 1503–1506. doi: 10.1126/science.7770778

Marcos, E., Pani, P., Brunamonti, E., Deco, G., Ferraina, S., and Verschure, P. (2013). Neural variability in premotor cortex is modulated by trial history and predicts behavioral performance. *Neuron* 78, 249–255. doi: 10.1016/j.neuron.2013.02.006

Masquelier, T. (2013). Neural variability, or lack thereof. *Front. Comput. Neurosci.* 7:7. doi: 10.3389/fncom.2013.00007

Mattia, M., Pani, P., Mirabella, G., Costa, S., Del Giudice, P., and Ferraina, S. (2013). Heterogeneous attractor cell assemblies for motor planning in premotor cortex. *J. Neurosci.* 33, 11155–11168. doi: 10.1523/JNEUROSCI.4664-12.2013

McDonnell, M. D., Goldwyn, J. H., and Lindner, B. (2016). Editorial: neuronal stochastic variability: influences on spiking dynamics and network activity. *Front. Comput. Neurosci.* 10:38. doi: 10.3389/fncom.2016.00038

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. doi: 10.1016/j.neuron.2009.09.013

Mochol, G., Hermoso-Mendizabal, A., Sakata, S., Harris, K. D., and de la Rocha, J. (2015). Stochastic transitions into silence cause noise correlations in cortical circuits. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3529–3534. doi: 10.1073/pnas.1410509112

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nat. Neurosci.* 17, 1410–1417. doi: 10.1038/nn.3807

Moreno-Bote, R., Knill, D. C., and Pouget, A. (2011). Bayesian sampling in visual perception. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12491–12496. doi: 10.1073/pnas.1101430108

Oram, M. W. (2011). Visual stimulation decorrelates neuronal activity. *J. Neurophysiol.* 105, 942–957. doi: 10.1152/jn.00711.2009

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nat. Neurosci.* 17, 594–600. doi: 10.1038/nn.3658

Pachitariu, M., Lyamzin, D. R., Sahani, M., and Lesica, N. A. (2015). State-dependent population coding in primary auditory cortex. *J. Neurosci.* 35, 2058–2073. doi: 10.1523/JNEUROSCI.3318-14.2015

Rabinovich, M., Huerta, R., and Laurent, G. (2008). Transient dynamics for neural processing. *Science* 321, 48–50. doi: 10.1126/science.1155564

Rabinovich, M. I., and Varona, P. (2011). Robust transient dynamics and brain functions. *Front. Comput. Neurosci.* 5:24. doi: 10.3389/fncom.2011.00024

Rabinovich, M. I., Varona, P., Tristan, I., and Afraimovich, V. S. (2014). Chunking dynamics: heteroclinics in mind. *Front. Comput. Neurosci.* 8:22. doi: 10.3389/fncom.2014.00022

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. doi: 10.1126/science.1179850

Renart, A., and Machens, C. (2014). Variability in neural activity and behavior. *Curr. Opin. Neurobiol.* 25, 211–220. doi: 10.1016/j.conb.2014.02.013

Roelfsema, P. R., Lamme, V. A., and Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nat. Neurosci.* 7, 982–991. doi: 10.1038/nn1304

Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E., and Doiron, B. (2017). The spatial structure of correlated neuronal variability. *Nat. Neurosci.* 20, 107–114. doi: 10.1038/nn.4433

Ruff, D. A., and Cohen, M. R. (2014). Attention can either increase or decrease spike count correlations in visual cortex. *Nat. Neurosci.* 17, 1591–1597. doi: 10.1038/nn.3835

Ruff, D. A., and Cohen, M. R. (2016). Attention increases spike count correlations between visual cortical areas. *J. Neurosci.* 36, 7523–7534. doi: 10.1523/JNEUROSCI.0610-16.2016

Sanchez-Vives, M. V., and Mattia, M. (2014). Slow wave activity as the default mode of the cerebral cortex. *Arch. Ital. Biol.* 152, 147–155. doi: 10.12871/000298292014239

Sancristobal, B., Rebollo, B., Boada, P., Sanchez-Vives, M. V., and Garcia-Ojalvo, J. (2016). Collective stochastic coherence in recurrent neuronal networks. *Nat. Phys.* 12, 881–887. doi: 10.1038/nphys3739

Schölvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D., and Carandini, M. (2015). Cortical state determines global variability and correlations in visual cortex. *J. Neurosci.* 35, 170–178. doi: 10.1523/JNEUROSCI.4994-13.2015

Schwemmer, M. A., Fairhall, A. L., Deneve, S., and Shea-Brown, E. T. (2015). Constructing precisely computing networks with biophysical spiking neurons. *J. Neurosci.* 35, 10112–10134. doi: 10.1523/JNEUROSCI.4951-14.2015

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.

Stein, R. B., Gossen, E. R., and Jones, K. E. (2005). Neuronal variability: noise or part of the signal? *Nat. Rev. Neurosci.* 6, 389–397. doi: 10.1038/nrn1668

Stiefel, K. M., Englitz, B., and Sejnowski, T. J. (2013). Origin of intrinsic irregular firing in cortical interneurons. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7886–7891. doi: 10.1073/pnas.1305219110

Sussillo, D. (2014). Neural circuits as computational dynamical systems. *Curr. Opin. Neurobiol.* 25, 156–163. doi: 10.1016/j.conb.2014.01.008

Sussillo, D., and Abbott, L. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron* 63, 544–557. doi: 10.1016/j.neuron.2009.07.018

Tognoli, E., and Kelso, S. A. (2014a). Enlarging the scope: grasping brain complexity. *Front. Syst. Neurosci.* 8:122. doi: 10.3389/fnsys.2014.00122

Tognoli, E., and Kelso, S. A. (2014b). The metastable brain. *Neuron* 81, 35–48. doi: 10.1016/j.neuron.2013.12.022

van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726. doi: 10.1126/science.274.5293.1724

Varona, P., and Rabinovich, M. I. (2016). Hierarchical dynamics of informational patterns and decision-making. *Proc. R. Soc. Lond. B Biol. Sci.* 283:20160475. doi: 10.1098/rspb.2016.0475

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* 334, 1569–1573. doi: 10.1126/science.1211095

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article

Frontiers
EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on