Wanpracha Chaovalitwongse
Panos M. Pardalos
Petros Xanthopoulos (Editors)

Computational Neuroscience

Springer

COMPUTATIONAL NEUROSCIENCE

Springer Optimization and Its Applications

VOLUME 38

Managing Editor
Panos M. Pardalos (University of Florida)

Editor–Combinatorial Optimization
Ding-Zhu Du (University of Texas at Dallas)

Advisory Board
J. Birge (University of Chicago)
C.A. Floudas (Princeton University)
F. Giannessi (University of Pisa)
H.D. Sherali (Virginia Polytechnic and State University)
T. Terlaky (McMaster University)
Y. Ye (Stanford University)

Aims and Scope
Optimization has been expanding in all directions at an astonishing rate during the last few decades. New algorithmic and theoretical techniques have been developed, the diffusion into other disciplines has proceeded at a rapid pace, and our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in all areas of applied mathematics, engineering, medicine, economics and other sciences.

The Springer *Optimization and Its Applications* series publishes undergraduate and graduate textbooks, monographs and state-of-the-art expository works that focus on algorithms for solving optimization problems and also study applications involving such problems. Some of the topics covered include nonlinear optimization (convex and nonconvex), network flow problems, stochastic optimization, optimal control, discrete optimization, multiobjective programming, description of software packages, approximation techniques and heuristic approaches.

For other titles in this series, go to www.springer.com/series/7393

Computational Neuroscience

By

WANPRACHA CHAOVALITWONGSE
Rutgers University, Piscataway, NJ, USA

PANOS M. PARDALOS
University of Florida, Gainesville, FL, USA

PETROS XANTHOPOULOS
University of Florida, Gainesville, FL, USA

Springer

Editors

Wanpracha Chaovalitwongse
Department of Industrial and
  Systems Engineering
Rutgers State University of
  New Jersey
96 Frelinghuysen Rd.
Piscataway NJ 08854
USA
wchaoval@rci.rutgers.edu

Panos M. Pardalos
Department of Industrial and
  Systems Engineering
University of Florida
303 Weil Hall
Gainesville FL 32611-6595
USA
pardalos@ufl.edu

Petros Xanthopoulos
Department of Industrial and
  Systems Engineering
University of Florida
303 Weil Hall P.O.Box 116595
Gainesville FL 32611-6595
USA
petrosx@ufl.edu

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To our parents

Preface

ΔΥΟ ΓΑΡ ΕΠΙΣΤΗΜΗ ΤΕ ΚΑΙ ΔΟΞΑ ΩΝ ΤΟ ΜΕΝ ΕΠΙΣΘΑΣΘΑΙ ΠΟΙΕΕΙ ΤΟ ΔΕ ΑΓΝΟΕΕΙΝ.

ΙΠΠΟΚΡΑΤΗΣ (460 π.X-360 π.X.)

There are in fact two things, science and opinion; the former begets knowledge, the latter ignorance.

Hippocrates (460BC-360BC)

This book represents a collection of recent advances in computational studies in neuroscience research that practically applies to a collaborative and integrative environment in engineering and medical domains. This work has been designed to address the explosion of interest by academic researchers and practitioners in highly-effective coordination between computational models and tools and quantitative investigation of neuroscientific data. To bridge the vital gap between science and medicine, this book brings together diverse research areas ranging from medical signal processing, image analysis, and data mining to neural network modeling, regulation of gene expression, and brain dynamics.

We hope that this work will also be of value to investigators and practitioners in academic institutions who become involved in computational modeling as an aid in translating information in neuroscientific data to their colleagues in medical domain. This volume will be very appealing to graduate (and advanced undergraduate) students, researchers, and practitioners across a wide range of industries (e.g., pharmaceutical, chemical, biological sciences), who require a detailed overview of the practical aspects of computational modeling in real-life neuroscience problems. For this reason, our audience is assumed to be very diverse and heterogenous, including:

- researchers from engineering, computer science, statistics, and mathematics domains as well as medical and biological scientists;
- physicians working in scientific research to understand how basic science can be linked with biological systems.

The book presents a collection of papers, several of which have been presented at DIMACS Conference on Computational Neuroscience that took place at the University of Florida on February 20 – 21, 2008. It is consisted of three major research themes in this book: data mining and medical data processing, brain modeling, and analysis of brain dynamics and neural synchronization. Each theme addresses the answer to a classical, yet extremely important, question in neuroscience, "How do we go from the mathematical modeling and computational techniques to the practical investigations of neuroscience problems?"

The first theme includes six chapters focused on data mining and medical data processing. The first chapter, by Paiva et al. lay down the platform of this book by presenting a complete methodological framework based on optimization for reproducing Hilbert spaces of spike trains. In the second chapter, Anderson et al. propose graph-theoretic models to investigate functional cooperation in the human brain. Not only can these models be applied to cognitive studies, they may also be used in diagnosis studies. In the third chapter, Sakkalis and Zervakis propose a framework for extracting time frequency features from electroencephalographic (EEG) recordings through the use of wavelet analysis. In the fourth chapter Chih-I Hung et al. present an application of independent component analysis (ICA) transformation into Creutzfeldt–Jakob disease. In the fifth chapter, Ramezani and Fatemizadeh discuss a comparison study of classification methods using various data preprocessing procedures applied to functional magnetic resonance imaging (fMRI) data for the detection of brain activation. In the sixth chapter, Fan et al. discuss the most well-known methods in biclustering applied to a neuroscientific application in evaluating the therapeutic intervention using vagus nerve stimulation treatment for patients with epilepsy. In the seventh chapter, Achler and Amir propose a genetic classifier used in the study of gene expression regulation.

The second theme includes five chapters that provide reviews and challenges in brain modeling in respect of human behavior and brain disease. In the eighth chapter, Ramírez et al. provide a review of the inverse source localization problem for neuroelectromagnetic source imaging of brain dynamics. In the ninth chapter, Wu et al. propose an approach based on the queuing theory and reinforcement learning for modeling the brain function and interpreting the human behavior. In the tenth and eleventh chapters, Cutsuridis suggests deterministic mathematical model for modeling neural networks of voluntary single-joint movement organization in normal subjects as well as patients with Parkinson's disease. In the twelfth chapter, Kawai et al. propose a parametric model for optical time series data of the respiratory neural network in the brainstem. In the thirteenth chapter, Leondopulos and Micheli-Tzanakou give an overview of the closed-loop deep brain stimulation technology and in the fourteenth chapter, Garzon and Neel present a novel approach to build fine grain models of the human brain with a large number of neurons inspired by recent advances in computing based on DNA modecules.

The third theme includes six chapters that focus on quantitative analyses of EEG recordings to investigate the brain dynamics and neural synchronization. In the fifteenth chapter, Sabesan et al. investigate the synchronization in the neural networks based on information flow, measured by the metric of network transfer entropy, among different brain areas. In the sixteenth chapter, Pardalos et al. describe an optimization-based model for estimating all Lyapunov exponents to characterize the dynamics of EEG recordings. In the seventeenth chapter, Faith et al. report the potential use of nonlinear dynamics for analyzing EEG recordings to evaluate the efficacy of antiepileptic drugs. In the eighteenth chapter, Kammerdiner and Pardalos study the synchronization of EEG recordings using the measures of phase synchronization and cointergrated VAR. In the nineteenth chapter, Liu et al. use the concept of mutual information to measure the coupling strength of EEG recordings in order to evaluate the efficacy of antiepileptic drugs in a very rare brain disease. In the last chapter, Sackellares et al. propose a seizure monitoring and alert system to be used in an intensive care unit based on statistical analyses of EEG recordings.

The completion of this issue would not have been possible without the assistance of many of our colleagues. We wish to express our gratitude to the authors for submitting and revising their work. We wish to express our sincere appreciation to anonymous referees for their careful reviewing. Their constructive comments contributed greatly to the quality of the issue. We cannot thank them enough for their time, efforts, and dedication to make this volume successful. The experience has been challenging, yet extremely rewarding. We truly hope that the reader will find the presented fundamental research and application papers presented as stimulating and valuable as we did.

USA,																									*Wanpracha Chaovalitwongse*
July 2009																								*Panos M. Pardalos*
																														Petros Xanthopoulos

Contents

Part II Modeling

Part III Brain Dynamics/Synchronization

List of Contributors

Tsvi Achler
Department of Computer Science, University of Illinois Urbana-Champaign,
Urbana, IL 61801, USA,
e-mail: `achler@uiuc.edu`

Eyal Amir
Department of Computer Science, University of Illinois Urbana-Champaign,
Urbana, IL 61801, USA,
e-mail: `eyal@cs.uiuc.edu`

Michael L. Anderson
Department of Psychology, Franklin and Marshall College, Lancaster, PA 17604,
USA; Institute for Advanced Computer Studies, University of Maryland, College
Park, MD 20742, USA,
e-mail: `michael.anderson@fandm.edu`

Sylvain Baillet
MEG Program, Department of Neurology, Medical College of Wisconsin
and Froedtert Hospital, Milwaukee, WI, USA,
e-mail: `sbaillet@mcw.edu`

Marc Berman
Department of Psychology, Department of Industrial and Operations Engineering,
University of Michigan-Ann Arbor, MI, USA,
e-mail: `bermanm@umich.edu`

Nikita Boyko
Department of Industrial and Systems Engineering, Center for Applied
Optimization, University of Florida, Gainesville, FL, USA,
e-mail: `nikita@ufl.edu`

Joan Brumbaugh
Department of Psychology, Franklin and Marshall College, Lancaster, PA 17604,
USA

Altannar Chinchuluun
Department of Industrial and Systems Engineering, Center for Applied
Optimization, University of Florida, Gainesville, FL, USA,
e-mail: altannar@ufl.edu

Vassilis Cutsuridis
Centre for Memory and Brain, Boston University, Boston, MA, USA,
e-mail: vcut@bu.edu

Aaron Faith
The Harrington Department of Bioengineering, Arizona State University, Tempe,
AZ 85287, USA,
e-mail: atfaith@asu.edu

Neng Fan
Department of Industrial and Systems Engineering, Center for Applied
Optimization, University of Florida, Gainesville, FL, USA,
e-mail: andynfan@ufl.edu

Emad Fatemizadeh
Biomedical Image and Signal Processing Laboratory (BiSIPL), School of Electrical
Engineering, Sharif University of Technology, Tehran, Iran,
e-mail: Fatemizadeh@sharif.edu

Max H. Garzon
Department of Computer Science, The University of Memphis, Memphis, TN,
USA,
e-mail: mgarzon@memphis.edu

Jen-Chuen Hsieh
Integrated Brain Research Laboratory, Department of Medical Research
and Education, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; Institute
of Brain Science, National Yang-Ming University, Taipei, Taiwan, ROC,
e-mail: jchsieh@vghtpe.gov.tw

Chih-I Hung
Department of Biomedical Imaging and Radiological Sciences, National Yang-
Ming University, Taipei, Taiwan, ROC; Integrated Brain Research Laboratory,
Department of Medical Research and Education, Taipei Veterans General Hospital,
Taipei, Taiwan, ROC

Leon Iasemidis
The Harrington Department of Bioengineering, Arizona State University, Tempe,
AZ, USA; Mayo Clinic, Phoenix, AZ, USA; Department of Electrical Engineering,
Arizona State University, Tempe, AZ, USA,
e-mail: leon.iasemidis@asu.edu

Makio Ishiguro
The Institute of Statistical Mathematics, Minato-ku, Tokyo, Japan,
e-mail: `ishiguro@ism.ac.jp`

Alla R. Kammerdiner
Department of Industrial and Systems Engineering, University of Florida,
Gainesville, FL, USA,
e-mail: `alla.ua@gmail.com`

Shigeharu Kawai
The Graduate University for Advanced Studies, Minato-ku, Tokyo, Japan,
e-mail: `kawai@ism.ac.jp`

Kazutaka Kobayashi
Department of Neurological Surgery Nihon University School of Medicine,
Tokyo, Japan; Division of Applied System Neuroscience, Department of Advanced
Medical, Science Nihon University School of Medicine, Tokyo, Japan

Stathis S. Leondopulos
Rutgers University, NJ, USA,
e-mail: `stathis@ece.rutgers.edu`

Chang-Chia Liu
J. Crayton Pruitt Family Department of Biomedical Engineering, University
of Florida, Gainesville, FL, USA,
e-mail: `iamjeff@ufl.edu`

Yili Liu
Department of Industrial and Operations Engineering, University of Michigan, Ann
Arbor, MI, USA,
e-mail: `yililiu@umich.edu`

Alexandre Messo
Department of Optimization, Kungliga Tekniska Högskolan, Stockholm, Sweden,
e-mail: `alex.messo@gmail.com`

Evangelia Micheli-Tzanakou
Rutgers University, NJ, USA,
e-mail: `etzanako@rci.rutgers.edu`

Fumikazu Miwakeichi
Chiba University, Inage-ku, Chiba, Japan,
e-mail: `miwake1@faculty.chiba-u.jp`

Andrew J. Neel
Department of Computer Science, The University of Memphis, Memphis, TN,
USA,
e-mail: `aneel@memphis.edu`

Yasumasa Okada
Keio University Tsukigase Rehabilitation Center, Izu, Shizuoka, Japan,
e-mail: `yasumasaokada@1979.jukuin.keio.ac.jp`

Yositaka Oku
Hyogo College of Medicine, Nishinomiya, Hyogo, Japan,
e-mail: yoku@hyo-med.ac.jp

António R. C. Paiva
Department of Electrical and Computer Engineering, University of Florida,
Gainesville, FL, USA,
e-mail: arpaiva@cnel.ufl.edu

Panos M. Pardalos
Department of Industrial and Systems Engineering, Center for Applied
Optimization, University of Florida, Gainesville, FL, USA; J. Crayton Pruitt Family
Department of Biomedical Engineering, University of Florida, Gainesville, FL
32611, USA; The Evelyn F. and William L. McKnight Brain Institute, University
of Florida, Gainesville, FL 32611, USA,
e-mail: pardalos@ufl.edu

Il Park
Pruitt Family Department of Biomedical Engineering, University of Florida,
Gainesville, FL, USA,
e-mail: memming@cnel.ufl.edu

José C. Príncipe
Department of Electrical and Computer Engineering, University of Florida,
Gainesville, FL, USA,
e-mail: principe@cnel.ufl.edu

Mahdi Ramezani
Biomedical Image and Signal Processing Laboratory (BiSIPL), School of Electrical
Engineering, Sharif University of Technology, Tehran, Iran,
e-mail: Ramezani@ee.sharif.edu

Rey R. Ramírez
MEG Program, Department of Neurology, Medical College of Wisconsin
and Froedtert Hospital, Milwaukee, WI, USA,
e-mail: rrramirez@mcw.edu

Shivkumar Sabesan
The Harrington Department of Bioengineering, Arizona State University, Tempe,
AZ USA; Barrow Neurological Institute, Phoenix, AZ, USA,
e-mail: ssabesa@asu.edu

J. Chris Sackellares
Optima Neuroscience, Inc., Gainesville, FL, USA,
e-mail: csackellares@optimaneuro.com

Vangelis Sakkalis
Department of Electronic and Computer Engineering, Technical University
of Crete, Chania, Greece,
e-mail: sakkalis@ics.forth.gr

Deng-Shan Shiau
Optima Neuroscience, Inc., Gainesville, FL, USA,
e-mail: dshiau@optimaneuro.com

Joseph Sirven
Mayo Clinic, Phoenix, AZ 85054, USA,
e-mail: joseph.sirven@mayo.edu

Bing-Wen Soong
The Neurological Institute, Taipei Veterans General Hospital, Taiwan, ROC;
Department of Neurology, National Yang-Ming University School of Medicine,
Taipei, Taiwan, ROC

Andreas Spanias
Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA,
e-mail: spanias@asu.edu

Aysu Şuben
Department of Psychology, Franklin and Marshall College, Lancaster, PA, USA

Yoshiyasu Tamura
The Institute of Statistical Mathematics, Minato-ku, Tokyo, Japan,
e-mail: tamura@ism.ac.jp

Shin Teng
Department of Biomedical Imaging and Radiological Sciences, National Yang-
Ming University, Taipei, Taiwan, ROC; Integrated Brain Research Laboratory,
Department of Medical Research and Education, Taipei Veterans General Hospital,
Taipei, Taiwan, ROC

Vera Tomaino
Bioinformatics Laboratory, Experimental and Clinical Medicine Department,
Magna Græcia University, viale Europa 88100, Catanzaro, Italy; Department of
Industrial and Systems Engineering, Center for Applied Optimization, University
of Florida, Gainesville, FL, USA,
e-mail: vera.tomaino@gmail.com

David Treiman
Barrow Neurological Institute, Phoenix, AZ, USA,
e-mail: dtreiman@chw.edu

Konstantinos Tsakalis
Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA,
e-mail: tsakalis@asu.edu

Basim M. Uthman
Department of Neurology, University of Florida, Gainesville, FL 32611, USA;
Department of Neuroscience, University of Florida, Gainesville, FL, USA;
The Evelyn F. and William L. McKnight Brain Institute, University of Florida,

Gainesville, FL, USA; Neurology Services, North Florida/South Georgia Veterans
Health System, Gainesville, FL, USA,
e-mail: `basim.uthman@med.va.gov`

Po-Shan Wang
The Neurological Institute, Taipei Veterans General Hospital, Taiwan, ROC;
Department of Neurology, National Yang-Ming University School of Medicine,
Taipei, Taiwan, ROC; The Neurological Institute, Taipei Municipal Gan-Dau
Hospital, Taipei, Taiwan, ROC

Norman Wang
Barrow Neurological Institute, Phoenix, AZ, USA

David Wipf
Biomagnetic Imaging Laboratory, University of California San Francisco, San
Francisco, CA, USA,
e-mail: `david.wipf@mrsc.ucsf.edu`

Changxu Wu
Department of Industrial and Systems Engineering, State University of New York
(SUNY), Buffalo, NY, USA,
e-mail: `Changxu@buffalo.edu`

Yu-Te Wu
Department of Biomedical Imaging and Radiological Sciences, National Yang-
Ming University, Taipei, Taiwan, ROC; Institute of Brain Science, National
Yang-Ming University, Taipei, Taiwan, ROC,
e-mail: `yute.wu@msa.hinet.net`

Petros Xanthopoulos
Department of Industrial and Systems Engineering, Center for Applied
Optimization, University of Florida, Gainesville, FL, USA,
e-mail: `petrosx@ufl.edu`

Vitaliy A. Yatsenko
Department of Industrial and Systems Engineering, Center for Applied
Optimization, University of Florida, Gainesville, FL, USA,
e-mail: `yatsenko@ufl.edu`

Michalis Zervakis
Department of Electronic and Computer Engineering, Technical University
of Crete, Chania, Greece,
e-mail: `michalis@display.tuc.gr`

Part I
Data Mining

Chapter 1
Optimization in Reproducing Kernel Hilbert Spaces of Spike Trains

António R. C. Paiva, Il Park, and José C. Príncipe

Abstract This chapter presents a framework based on reproducing kernel Hilbert spaces (RKHS) for optimization with spike trains. To establish the RKHS for optimization we start by introducing kernels for spike trains. It is shown that spike train kernels can be built from ideas of kernel methods or from the intensity functions underlying the spike trains. However, the later approach shall be the main focus of this study. We introduce the memoryless cross-intensity (mCI) kernel as an example of an inner product of spike trains, which defines the RKHS bottom-up as an inner product of intensity functions. Being defined in terms of the intensity functions, this approach toward defining spike train kernels has the advantage that points in the RKHS incorporate a statistical description of the spike trains, and the statistical model is explicitly stated. Some properties of the mCI kernel and the RKHS it induces will be given to show that this RKHS has the necessary structure for optimization. The issue of estimation from data is also addressed. We finalize with an example of optimization in the RKHS by deriving an algorithm for principal component analysis (PCA) of spike trains.

António R. C. Paiva

Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA, e-mail: `arpaiva@cnel.ufl.edu`

Il Park

Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611, USA, e-mail: `memming@cnel.ufl.edu`

José C. Príncipe

Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA, e-mail: `principe@cnel.ufl.edu`

1.1 Introduction

A spike train $s \in \mathcal{S}(\mathcal{T})$ is a sequence of ordered spike times $s = \{t_m \in \mathcal{T} : m = 1, \ldots, N\}$ corresponding to the time instants in the interval $\mathcal{T} = [0, T]$ at which a neuron fires. In a different perspective, spike trains are realizations of stochastic point processes. Spike trains can be observed whenever studying either real or artificial neurons. In neurophysiological studies, spike trains result from the activity of multiple neurons in single-unit recordings by ignoring the stereotypical shape of action potentials [5]. And, more recently, there has also been a great interest in using spike trains for biologically inspired computation paradigms such as the liquid-state machine (LSM) [13, 12] or spiking neural networks (SNN) [3, 12]. Regardless of the nature of the process giving rise to the spike trains, the ultimate goal is to filter or classify the spike trains to manipulate or extract the encoded information.

Filtering, eigendecomposition, clustering, and classification are often formulated in terms of a criterion to be optimized. However, formulation of a criterion and/or optimization directly with spike trains is not a straightforward task. The most widely used approach is to bin the spike trains, obtained by segmenting the spike train in small intervals and counting the number of spikes within each interval [5]. The advantage of this approach is that the randomness in time is mapped to randomness in amplitude of a discrete-time random process, and, therefore, our usual statistical signal processing and machine learning techniques can be applied. It is known that if the bin size is large compared to the average inter-spike interval this transformation provides a rough estimate of the instantaneous rate. However, the discretization of time introduced by binning leads to low resolution.

The caveats associated with binned spike trains have motivated alternative methodologies involving the spike times directly. For example, to deal with the problem of classification, Victor and Purpura [36, 37] defined a distance metric between spike trains resembling the edit distance in computer science. An alternative distance measure was proposed by van Rossum [34]. Using spike train distances for classification simplifies the problem to that of finding a threshold value. However, for more general problems the range of applications that can be solved directly using distances is limited since these metrics do not lend themselves to optimization. The reason is that although distances are useful concepts in classification and pattern analysis they do not provide a general framework for statistical signal processing and machine learning. Recent attempts were also made to develop a mathematical theory from simple principles [4, 31], such as the definition of an inner product and an associated kernel, but these developments are mainly associated with the earlier proposed distance measures [37, 34].

The framework described in this chapter is different in the sense that it does not attempt to propose a distance or criterion directly. Rather, we propose to define first inner product kernel functions[1] for spike trains. These kernels induce

---

[1] Throughout this document we will refer to inner products and kernels indistinguishably since they represent the same concept. However, stated more correctly, kernels denote inner products in a reproducing kernel Hilbert space of functions on the arguments of the kernel.

reproducing kernel Hilbert spaces (RKHS) of functions on spike trains, which provide the needed mathematical structure to easily define and optimize criteria for a diverse range of problems. Another advantage of this approach is that many of the difficulties found in manipulating spike trains which lead to the use of binning are implicitly taken care of through the mapping to the RKHS. In this chapter we exemplify the construction of an RKHS by defining an inner product of spike trains called *memoryless cross-intensity (mCI) kernel*. This spike train kernel defines the RKHS bottom-up as an inner product of intensity functions and thus incorporates a statistical description of the spike trains. As will be showed later, this particular kernel is related to the *generalized cross-correlation* (GCC) [18] but provides a more principled and broader perspective on many spike train methods reported in the literature.

For continuous and discrete random processes, RKHS theory has already been proven essential in a number of applications, such as statistical signal processing [20, 23] and detection [9, 11, 10], as well as statistical learning theory [29, 35, 38]. Indeed, Parzen showed that several statistical signal processing algorithms can be stated as optimization problems in the RKHS and easily solved [20, 23]. For instance, the cross-correlation function used throughout statistical analysis and signal processing, including the celebrated Wiener filter [8], is a valid kernel and induces an RKHS space [20]. Although frequently overlooked, RKHS theory plays a pivotal role in kernel methods [29, 35] because it is the reason for the famed kernel trick which allows for the otherwise seemingly intractable task of deriving and applying kernel techniques.

In the following, we introduce how to define spike train kernels and present some examples. A systematic approach which builds the RKHS from the ground up is followed by defining inner products for spike trains. The main advantage in this path is a general and mathematically precise methodology which, nevertheless, can easily be interpreted intuitively by analyzing the definition of the inner product or, conversely, defining the inner product to match our understanding of a given problem. In this study we present the mCI kernel as an example, since it incorporates a statistical description of the spike trains and the statistical model is clearly stated, but the ideas can be easily extended. A number of properties are proved for the mCI kernel, and the relationships between the RKHS and the congruent spaces are discussed for additional insight. The issue of estimation from data is also addressed. Finally, the usefulness of an RKHS framework for optimization is demonstrated through the derivation of an algorithm for principal component analysis (PCA) of spike trains.

1.2 Some Background on RKHS Theory

In this section, some basic concepts of kernel methods and RKHS theorem necessary for the understanding of the next sections are reviewed. The notation was purposely chosen to be different from the one used later since the presentation here is meant to be as general and introductory as possible.

The fundamental result in RKHS theory is the famed *Moore–Aronszajn theorem* [1, 17]. Let K denote a generic symmetric and positive definite function of two variables defined on some space E. That is, a function $K(\cdot,\cdot) : E \times E \to \mathbb{R}$ such that it verifies:

(i) Symmetry: $K(x,y) = K(y,x), \quad \forall x,y \in E$.
(ii) Positive definiteness: for any finite number of l ($l \in \mathbb{N}$) points $x_1, x_2, \ldots, x_l \in E$ and any corresponding coefficients $c_1, c_2, \ldots, c_l \in \mathbb{R}$,

$$\sum_{m=1}^{l} \sum_{n=1}^{l} c_m c_n K(x_m, x_n) \geq 0. \tag{1.1}$$

These are sometimes called the Mercer conditions [16] in the kernel methods literature. Then, the Moore–Aronszajn theorem [1, 17] guaranties that there exists a unique Hilbert space $\mathcal{H}$ of real valued functions defined on E such that, for every $x \in E$,

(i) $K(x, \cdot) \in \mathcal{H}$ and
(ii) for any $f \in \mathcal{H}$

$$f(x) = \langle f(\cdot), K(x, \cdot) \rangle_{\mathcal{H}}. \tag{1.2}$$

The identity on Equation (1.2) is called the *reproducing property* of K and, for this reason, $\mathcal{H}$ is said to be an RKHS with reproducing kernel K.

Two essential corollaries of the theorem just described can be observed. First, since both $K(x, \cdot)$ and $K(y, \cdot)$ are in $\mathcal{H}$, we get from the reproducing property that

$$K(x,y) = \langle K(x, \cdot), K(y, \cdot) \rangle_{\mathcal{H}}. \tag{1.3}$$

Hence, K evaluates the inner product in this RKHS. This identity is the *kernel trick*, well known in kernel methods, and is the main tool for computation in this space. Second, a consequence of the previous properties and which can be seen easily in the kernel trick is that, given any point $x \in E$, the representer of evaluation in the RKHS is $\Psi_x(\cdot) = K(x, \cdot)$. Notice that the *functional transformation* Ψ from the input space E into the RKHS $\mathcal{H}$ evaluated for a given x, and in general any element of the RKHS, is a real function defined on E.

A quite interesting perspective to RKHS theory is provided by Parzen's work [22]. In his work, Parzen proved that for *any* symmetric and positive definite function there exists a space of Gaussian distributed random variables defined in the input space of the kernel for which this function is the covariance function [20]. Notice that, assuming stationarity and ergodicity, this space might just as well be thought of as a space of random processes. That is to say that any kernel inducing an RKHS denotes simultaneously an inner product in the RKHS and a covariance operator in another space. Furthermore, it is established that there exists an isometric inner product-preserving mapping, a *congruence*, between these two spaces. Consequently, the RKHS $\mathcal{H}$ induced by the kernel and the space of random variables where this kernel is a covariance function are said to be *congruent*. This is an important result as it sets up a correspondence between the inner product due to a

kernel in the RKHS to our intuitive understanding of the covariance function and the associated linear statistics. In other words, due to the congruence between the two spaces an algorithm can be derived and interpreted in any of the spaces.

1.3 Inner Product for Spike Times

Denote the mth spike time in a spike train indexed by i as $t_m^i \in \mathcal{T}$, with $m \in \{1, 2, \ldots, N_i\}$ and N_i the number of spike times in the spike train. To simplify the notation, however, the spike train index will be omitted if is irrelevant for the presentation or obvious from the context.

The simplest inner product that can be defined for spike trains operates with only two spike times at a time as observed by Carnell and Richardson [4]. In the general case, such an inner product can be defined in terms of a kernel function defined on $\mathcal{T} \times \mathcal{T}$ into the reals, with $\mathcal{T}$ the interval of spike times. Let κ denote such a kernel. Conceptually, this kernel operates in the same way as the kernels operating on data samples in kernel methods [29] and information theoretic learning [24]. Although it operates only with two spike times, it will play a major role whenever we operate with complete realizations of spike trains. Indeed, as the next sections show, the estimators for one of the kernels we define on spike trains rely on this kernel as an elemental operation for computation.

To take advantage of the framework for statistical signal processing provided by RKHS theory, κ is required to be a symmetric positive definite function. By the Moore–Aronszajn theorem [1], this ensures that an RKHS $\mathcal{H}_\kappa$ must exist for which κ is a reproducing kernel. The inner product in $\mathcal{H}_\kappa$ is given as

$$\kappa(t_m, t_n) = \langle \kappa(t_m, \cdot), \kappa(t_n, \cdot) \rangle_{\mathcal{H}_\kappa} = \langle \Phi_m, \Phi_n \rangle_{\mathcal{H}_\kappa}, \qquad (1.4)$$

where Φ_m is the element in $\mathcal{H}_\kappa$ corresponding to t_m (that is, the transformed spike time).

Since the kernel operates directly on spike times and is, typically, undesirable to emphasize events in this space, κ is further required to be *shift-invariant*; that is, for any $\theta \in \mathbb{R}$,

$$\kappa(t_m, t_n) = \kappa(t_m + \theta, t_n + \theta), \quad \forall t_m, t_n \in \mathcal{T}. \qquad (1.5)$$

In other words, the kernel is only sensitive to the difference of the arguments and, consequently, we may also write $\kappa(t_m, t_n) = \kappa(t_m - t_n)$.

For any symmetric, shift-invariant, and positive definite kernel, it is known that $\kappa(0) \geq |\kappa(\theta)|$.[2] This is important in establishing κ as a similarity measure between spike times. In other words, as usual, an inner product should intuitively measure some form of inter-dependence between spike times. However, notice that the conditions posed do not restrict this study to a single kernel. Quite on the contrary, any

---

[2] This is a direct consequence of the fact that symmetric positive definite kernels denote inner products that obey the Cauchy–Schwarz inequality.

kernel satisfying the above requirements is theoretically valid and understood under the framework proposed here although, obviously, the practical results may vary.

An example of a family of kernels that can be used (but not limited to) is the radial basis functions [2],

$$\kappa(t_m, t_n) = \exp(-|t_m - t_n|^p), \quad t_m, t_n \in \mathcal{T}, \tag{1.6}$$

for any $0 < p \leq 2$. Some well-known kernels, such as the widely used Gaussian and Laplacian kernels, are special cases of this family for $p = 2$ and $p = 1$, respectively.

It is interesting to notice that shift-invariant kernels result in a natural norm induced by the inner product with the following property:

$$\|\Phi_m\| = \sqrt{\kappa(0)}, \quad \forall \Phi_m \in \mathcal{H}_\kappa. \tag{1.7}$$

Since the norm of the transformed spike times in $\mathcal{H}_\kappa$ is constant, all the spike times are mapped to the surface of a hypersphere in $\mathcal{H}_\kappa$. The set of transformed spike times is called the manifold of $\mathcal{S}(\mathcal{T})$. Moreover, this shows in a different perspective why the kernel used needs to be nonnegative. Furthermore, the *geodesic distance* corresponding to the length of the smallest path contained within this manifold (in this case, the hypersphere) between two functions in this manifold, Φ_m and Φ_n, is given by

$$\begin{aligned} d(\Phi_m, \Phi_n) &= \|\Phi_m\| \arccos\left(\frac{\langle \Phi_m, \Phi_n \rangle}{\|\Phi_m\| \|\Phi_n\|}\right) \\ &= \sqrt{\kappa(0)} \arccos\left[\frac{\kappa(t_m, t_n)}{\kappa(0)}\right]. \end{aligned} \tag{1.8}$$

Put differently, from the geometry of the transformed spike times, the kernel function is proportional to the cosine of the angle between two transformed spike times in $\mathcal{H}_\kappa$. Because the kernel is nonnegative, the maximum angle is $\pi/2$, which restricts the manifold of transformed spike times to a small area of the hypersphere. With the kernel inducing the above metric, the manifold of the transformed points forms a *Riemannian space*. However, this space is *not* a linear space. Fortunately, its span is obviously a linear space. In fact, it equals the RKHS associated with the kernel. Although this is not a major problem, computing with the transformed points will almost surely yield points outside of the manifold of transformed spike times. This means that such points cannot be mapped back to the input space directly. Depending on the aim of the application this may not be necessary, but if required, it may be solvable through a projection to the manifold of transformed input points.

1.4 Inner Product for Spike Trains

Although any kernel verifying the conditions discussed in the previous section induces an RKHS and therefore is of interest on itself, the fact that it only operates

with two spike times at a time limits its practical use. In particular, spike trains are *sets* of spike times but we have not yet addressed the problem of how to combine the kernel for all spike times. One immediate approach is to utilize the linearity of the RKHS [4]. If the *m*th spike time is represented in the RKHS by Φ_m, then the spike train can be represented in the RKHS as the sum of the transformed spike times,

$$\Psi = \sum_{m=1}^{N} \Phi_m. \tag{1.9}$$

Notice that if a spike time is represented by a given function, say, an impulse, the spike train will be a sum of time-shifted impulses centered at the spike times. Then Equation (1.9) implies that the mapping of the spike train into the RKHS induced by the spike time kernel is linear. Using the linearity of the RKHS it results that the inner product of spike trains is

$$\left\langle \Psi_{s_i}, \Psi_{s_j} \right\rangle_{\mathcal{H}_\kappa} = \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \left\langle \Phi_m^i, \Phi_n^j \right\rangle_{\mathcal{H}_\kappa} = \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \kappa(t_m^i, t_n^j). \tag{1.10}$$

It must be remarked that Equation (1.10) is only one example of a spike train kernel from inner products on spike times. Indeed, as is commonly done in kernel methods, more complex spike train kernels can be defined utilizing the kernel on spike times as a building block equating the nonlinear relationship between the spike times. On the other hand, the main disadvantage in this approach toward spike train analysis is that the underlying model assumed for the spike train is not clearly stated. This is important in determining and understanding the potential limitations of a given spike train kernel for data analysis.

Rather than utilizing this direct approach, an alternative construction is to define first a general inner product for the spike trains from the fundamental statistical descriptors. In fact, it will be seen that the inner product for spike trains builds upon the kernel on single spike times. This bottom-up construction of the kernel for spike trains is unlike the previous approach and is rarely taken in machine learning, but it exposes additional insight on the properties of the kernel and the RKHS it induces for optimization and data analysis.

A spike train is a realization of an underlying stochastic point process [33]. In general, to completely characterize a point process, the conditional intensity function must be used. The Poisson process is a special case because it is memoryless and, therefore, the intensity function (or rate function) is sufficient [33, Chapter 2]. Spike trains in particular have been found to be reasonably well modeled as realizations of Poisson processes [28, Chapter 2]. Hence, for the remaining of this study only Poisson spike trains are considered.

Consider two spike trains, $s_i, s_j \in \mathcal{S}(\mathcal{T})$, with $i, j \in \mathbb{N}$. Denote the intensity of the underlying Poisson processes by $\lambda_{s_i}(t)$ and $\lambda_{s_j}(t)$, respectively, where $t \in \mathcal{T} = [0, T]$ denotes the time coordinate. Note that the dependence of the intensity function on t indicates that the Poisson processes considered may be inhomogeneous (i.e., nonstationary). For any practical spike train and for finite T, we have that

$$\int_{\mathcal{T}} \lambda_{s_i}^2(t)dt < \infty. \tag{1.11}$$

As a consequence, the intensity functions of spike trains are valid elements of $L_2(\mathcal{T}) \subset L_2$. Moreover, in this space, we can define an inner product of intensity functions as the usual inner product in L_2,

$$I(s_i, s_j) = \left\langle \lambda_{s_i}, \lambda_{s_j} \right\rangle_{L_2(\mathcal{T})} = \int_{\mathcal{T}} \lambda_{s_i}(t)\lambda_{s_j}(t)dt. \tag{1.12}$$

We shall refer to $I(\cdot, \cdot)$ as the memoryless cross-intensity (mCI) kernel. Notice that the mCI kernel incorporates the statistics of the processes directly and treats seamlessly even the case of inhomogeneous Poisson processes.

Furthermore, the definition of inner product naturally induces a norm in the space of the intensity functions,

$$\|\lambda_{s_i}(\cdot)\|_{L_2(\mathcal{T})} = \sqrt{\langle \lambda_{s_i}, \lambda_{s_i} \rangle_{L_2(\mathcal{T})}} = \sqrt{\int_{\mathcal{T}} \lambda_{s_i}^2(t)dt} \tag{1.13}$$

which is very useful for the formulation of optimization problems.

It is insightful to compare the mCI kernel definition in Equation (1.12) with the so-called *generalized cross-correlation* (GCC) [18],

$$\begin{aligned} C_{AB}(\theta) &= E\left\{\lambda_A(t)\lambda_B(t+\theta)\right\} \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \lambda_A(t)\lambda_B(t+\theta)dt. \end{aligned} \tag{1.14}$$

Although the GCC was proposed directly as a more general form of cross-correlation of spike trains, one verifies that the two ideas are fundamentally equivalent. Nevertheless, the path toward the definition of mCI is more principled. More importantly, this path suggests alternative spike train kernel definitions which may not require a Poisson assumption, or, if the Poisson model is assumed, extract more information in the event of deviations from the model.

1.5 Properties and Estimation of the Memoryless Cross-Intensity Kernel

1.5.1 Properties

In this section some relevant properties of the mCI kernel are presented. In addition to the knowledge they provide, they are necessary for a clear understanding of the following sections.

Property 1.1. The mCI kernel is a symmetric, nonnegative, and linear operator in the space of the intensity functions.

Because the mCI kernel operates on elements of $L_2(\mathcal{T})$ and corresponds to the usual dot product from L_2, this property is a direct consequence of the properties inherited from L_2. More specifically, Property 1.1 guaranties the mCI kernel is a valid inner product.

Property 1.2. For any set of $n \geq 1$ spike trains, the mCI kernel matrix

$$
V = \begin{bmatrix}
I(s_1, s_1) & I(s_1, s_2) & \ldots & I(s_1, s_n) \\
I(s_2, s_1) & I(s_2, s_2) & \ldots & I(s_2, s_n) \\
\vdots & \vdots & \ddots & \vdots \\
I(s_n, s_1) & I(s_n, s_2) & \ldots & I(s_n, s_n)
\end{bmatrix}
$$

is symmetric and nonnegative definite.

The proof is given in the appendix. Through the work of Moore [17] and due to the Moore–Aronszajn theorem [1], the following two properties result as corollaries of Property 1.2.

Property 1.3. The mCI kernel is a symmetric and positive definite kernel. Thus, by definition, for any set of $n \geq 1$ point processes and corresponding n scalars $a_1, a_2, \ldots, a_n \in \mathbb{R}$,

$$
\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j I(s_i, s_j) \geq 0. \tag{1.15}
$$

Property 1.4. There exists a Hilbert space for which the mCI kernel is a reproducing kernel.

Actually, Property 1.3 can be obtained explicitly by verifying that the inequality of Equation (1.15) is implied by Equations (1.44) and (1.45) in the proof of Property 1.2 (in the appendix).

Properties 1.2, 1.3, and 1.4 are equivalent in the sense that any of these properties implies the other two. The most important consequence of these properties, explicitly stated through Property 1.4, is that the *mCI kernel induces an unique RKHS*, henceforth denoted by $\mathcal{H}_I$.

Property 1.5. The mCI kernel verifies the Cauchy–Schwarz inequality,

$$
I^2(s_i, s_j) \leq I(s_i, s_i) I(s_j, s_j) \qquad \forall s_i, s_j \in \mathcal{S}(\mathcal{T}). \tag{1.16}
$$

The proof is given in the appendix. The Cauchy–Schwarz inequality is important since the triangle inequality results as an immediate consequence and it induces a correlation coefficient-like measure very useful for matching spike trains. Indeed, the Cauchy–Schwarz inequality is the concept behind the spike train measure proposed by Schreiber et al. [32]. However, our proof in appendix verifies that all it is required is a spike train kernel inducing an RKHS, and, therefore, the idea by Schreiber and colleagues is easily extendible.

Property 1.6. For any two point processes $s_i, s_j \in \mathcal{S}(\mathcal{T})$ the triangle inequality holds. That is,

$$\left\|\lambda_{s_i} + \lambda_{s_j}\right\| \leq \left\|\lambda_{s_i}\right\| + \left\|\lambda_{s_j}\right\|.$$

As before, the proof is given in the appendix.

1.5.2 Estimation

As previous stated, spike trains are realizations of underlying point processes, but the memoryless cross-intensity kernel as presented so far is a deterministic operator on the point processes rather than on the observed spike trains. Using a well-known methodology for the estimation of the intensity function we now derive an estimator for the memoryless cross-intensity kernel. One of the advantages of this route is that the conceptual construction of spike train kernel is dissociated from the problem of estimation from data. Put differently, in this way it is possible to have a clear statistical interpretation while later approaching the problem from a practical point of view. The connection between the mCI kernel and κ will now become obvious.

A well-known method for intensity estimation from *a single spike train* is kernel smoothing [5, 26]. Accordingly, given a spike train s_i comprising of spike times $\{t_m^i \in \mathcal{T} : m = 1, \ldots, N_i\}$ the estimated intensity function is

$$\hat{\lambda}_{s_i}(t) = \sum_{m=1}^{N_i} h(t - t_m^i), \tag{1.17}$$

where h is the smoothing function. This function must be nonnegative and integrate to one over the real line (just like a probability distribution function (pdf)). Commonly used smoothing functions are the Gaussian, Laplacian, and α-functions, among others.

From a filtering perspective, Equation (1.17) can be seen as a linear convolution between the filter impulse response given by $h(t)$ and the spike train given as a sum of Dirac functionals centered at the spike times. In particular, binning is nothing but a special case of this procedure in which the spike times are first quantized according to the binsize and h is a rectangular window [5]. Moreover, compared with pdf estimation with Parzen windows [21], we immediately observe that intensity estimation as shown above is directly related to the problem of pdf estimation except for a normalization term, a connection made clear by Diggle and Marron [6].

Consider spike trains $s_i, s_j \in \mathcal{S}(\mathcal{T})$ with estimated intensity functions $\hat{\lambda}_{s_i}(t)$ and $\hat{\lambda}_{s_j}(t)$ according to Equation (1.17). Substituting the estimated intensity functions in the definition of the mCI kernel (Equation (1.12)) yields

$$\hat{I}(s_i, s_j) = \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \kappa(t_m^i - t_n^j), \tag{1.18}$$

where κ is the 'kernel' obtained by the autocorrelation of the smoothing function h. Notice that ultimately the obtained estimator linearly combines and weights the contribution of a kernel operating on a pair of event coordinates. Moreover, this estimator operates directly on the event coordinates of the whole realization without loss of resolution and in a computationally efficient manner since it takes advantage of the, typically, sparse occurrence of events.

If the kernel κ is chosen such that it satisfies the requirements in Section 1.3, then the mCI kernel corresponds to a summation of all pairwise inner products between spike times of the spike trains, evaluated by kernel on the spike time differences. Put in this way, we can now clearly see how the mCI inner product on spike trains builds upon the inner product on spike times denoted by κ and the connection to Equation (1.10). The later approach, however, clearly states the underlying point process model.

1.6 Induced RKHS and Congruent Spaces

Some considerations about the RKHS space $\mathcal{H}_I$ induced by the mCI kernel and congruent spaces are made in this section. The relationship between $\mathcal{H}_I$ and its congruent spaces provides alternative perspectives and a better understanding of the mCI kernel. Figure 1.1 provides a diagram of the relationships among the various spaces discussed next.
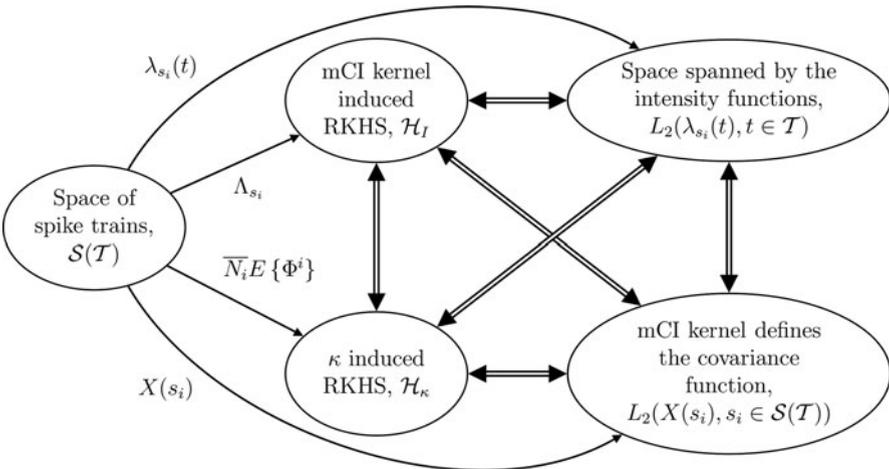

Fig. 1.1: Relation between the original space of spike trains $\mathcal{S}(\mathcal{T})$ and the various Hilbert spaces. The *double-line* bi-directional connections denote congruence between spaces.

1.6.1 Space Spanned by Intensity Functions

In the introduction of the mCI kernel the usual dot product in $L_2(\mathcal{T})$, the space of square integrable intensity functions defined on $\mathcal{T}$, was utilized. The definition of the inner product in this space provides an intuitive understanding to the reasoning involved. $L_2(\lambda_{s_i}(t), t \in \mathcal{T}) \subset L_2(\mathcal{T})$ is clearly a Hilbert space with inner product and norm defined in Equations (1.12) and (1.13). Notice that the span of this space contains also elements (functions) that may not be valid intensity functions since, by definition, intensity functions are always nonnegative. However, since our interest is mainly on the evaluation of the inner product this is of no consequence. The key limitation, however, is that $L_2(\lambda_{s_i}(t), t \in \mathcal{T})$ is *not* an RKHS. This should be clear because elements in this space are functions defined on $\mathcal{T}$, whereas elements in the RKHS $\mathcal{H}_I$ must be functions defined on $\mathcal{S}(\mathcal{T})$.

Despite the differences, the spaces $L_2(\lambda_{s_i}(t), t \in \mathcal{T})$ and $\mathcal{H}_I$ are closely related. In fact, $L_2(\lambda_{s_i}(t), t \in \mathcal{T})$ and $\mathcal{H}_I$ are congruent. This congruence can be verified explicitly since there is clearly a one-to-one mapping,

$$\lambda_{s_i}(t) \in L_2(\lambda_{s_i}(t), t \in \mathcal{T}) \quad \longleftrightarrow \quad \Lambda_{s_i}(s) \in \mathcal{H}_I,$$

and, by definition of the mCI kernel,

$$I(s_i, s_j) = \left\langle \lambda_{s_i}, \lambda_{s_j} \right\rangle_{L_2(\mathcal{T})} = \left\langle \Lambda_{s_i}, \Lambda_{s_j} \right\rangle_{\mathcal{H}_I}. \tag{1.19}$$

A direct implication of the basic congruence theorem is that the two spaces have the same dimension [20].

1.6.2 Induced RKHS

In Section 1.5.1 it was shown that the mCI kernel is symmetric and positive definite (Properties 1.1 and 1.3, respectively) and consequently, by the Moore–Aronszajn theorem [1], there exists a Hilbert space $\mathcal{H}_I$ in which the mCI kernel evaluates the inner product and is a reproducing kernel (Property 1.4). This means that $I(s_i, \cdot) \in \mathcal{H}_I$ for any $s_i \in \mathcal{S}(\mathcal{T})$ and, for any $\xi \in \mathcal{H}_I$, the reproducing property holds

$$\langle \xi, I(s_i, \cdot) \rangle_{\mathcal{H}_I} = \xi(s_i). \tag{1.20}$$

As a result the kernel trick follows:

$$I(s_i, s_j) = \left\langle I(s_i, \cdot), I(s_j, \cdot) \right\rangle_{\mathcal{H}_I}. \tag{1.21}$$

Written in this form, it is easy to verify that the point in $\mathcal{H}_I$ corresponding to a spike train $s_i \in \mathcal{S}(\mathcal{T})$ is $I(s_i, \cdot)$. In other words, given any spike train $s_i \in \mathcal{S}(\mathcal{T})$, this spike train is mapped to $\Lambda_{s_i} \in \mathcal{H}_I$, given explicitly (although unknown in closed form) as $\Lambda_{s_i} = I(s_i, \cdot)$. Then Equation (1.21) can be restated in the more usual form

$$I(s_i, s_j) = \left\langle \Lambda_{s_i}, \Lambda_{s_j} \right\rangle_{\mathcal{H}_I}. \tag{1.22}$$

It must be remarked that $\mathcal{H}_I$ is in fact a functional space. More specifically, that points in $\mathcal{H}_I$ are functions of spike trains defined on $\mathcal{S}(\mathcal{T})$. This is a key difference between the space of intensity functions $L_2(\mathcal{T})$ explained above and the RKHS $\mathcal{H}_I$, in that the latter allows for statistics of the transformed spike trains to be estimated *as functions of spike trains*. The usefulness of an RKHS for optimization and general computation with spike trains can be appreciated, for example, in the derivation of principal component analysis in Section 1.7.

1.6.3 mCI Kernel and the RKHS Induced by κ

The mCI kernel estimator in Equation (1.18) shows the evaluation written in terms of elementary kernel operations on the spike times. This fact alone provides a different perspective on how the mCI kernel uses the statistics of the spike times. To see this more clearly, if κ is chosen according to Section 1.3 as symmetric positive definite, then it can be substituted by its inner product (Equation (1.4)) in the mCI kernel estimator, yielding

$$
\begin{aligned}
\hat{I}(s_i, s_j) &= \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \left\langle \Phi_m^i, \Phi_n^j \right\rangle_{\mathcal{H}_\kappa} \\
&= \left\langle \sum_{m=1}^{N_i} \Phi_m^i, \sum_{n=1}^{N_j} \Phi_n^j \right\rangle_{\mathcal{H}_\kappa}.
\end{aligned}
\tag{1.23}
$$

When the number of samples approaches infinity (so that the intensity functions and, consequently the mCI kernel, can be estimated exactly) the mean of the transformed spike times approaches the expectation. Hence, Equation (1.23) results in

$$I(s_i, s_j) = \overline{N_i}\, \overline{N_j} \left\langle E\left\{\Phi^i\right\}, E\left\{\Phi^j\right\} \right\rangle_{\mathcal{H}_\kappa}, \tag{1.24}$$

where $E\left\{\Phi^i\right\}, E\left\{\Phi^j\right\}$ denotes the expectation of the transformed spike times and $\overline{N_i}, \overline{N_j}$ are the expected number of spikes in spike trains s_i and s_j, respectively.

Equation (1.23) explicitly shows that the mCI kernel can be computed as an inner product of the expectation of the transformed spike times in the RKHS $\mathcal{H}_\kappa$ induced by κ. In other words, there is a congruence $\mathscr{G}$ between $\mathcal{H}_\kappa$ and $\mathcal{H}_I$ in this case given explicitly by the expectation of the transformed spike times, $\mathscr{G}(\Lambda_{s_i}) = \overline{N_i} E\left\{\Phi^i\right\}$, such that

$$\left\langle \Lambda_{s_i}, \Lambda_{s_j} \right\rangle_{\mathcal{H}_I} = \left\langle \mathscr{G}(\Lambda_{s_i}), \mathscr{G}(\Lambda_{s_j}) \right\rangle_{\mathcal{H}_\kappa} = \left\langle \overline{N_i} E\left\{\Phi^i\right\}, \overline{N_j} E\left\{\Phi^j\right\} \right\rangle_{\mathcal{H}_\kappa}. \tag{1.25}$$

Recall that the transformed spike times form a manifold (the subset of a hypersphere) and, since these points have constant norm, the kernel inner product depends

only on the angle between points. This is typically not true for the average of these points, however. Observe that the circular variance [14] of the transformed spike times of spike trains s_i is

$$
\begin{aligned}
\mathrm{var}(\Phi^i) &= E\left\{\left\langle \Phi^i_m, \Phi^i_m \right\rangle_{\mathcal{H}_\kappa}\right\} - \left\langle E\left\{\Phi^i\right\}, E\left\{\Phi^i\right\}\right\rangle_{\mathcal{H}_\kappa} \\
&= \kappa(0) - \left\|E\left\{\Phi^i\right\}\right\|^2_{\mathcal{H}_\kappa}.
\end{aligned}
\tag{1.26}
$$

So, the norm of the mean transformed spike times is inversely proportional to the variance of the elements in $\mathcal{H}_\kappa$. This means that the inner product between two spike trains depends also on the dispersion of these average points. This fact is important because data reduction techniques rely heavily on optimization with the data variance. For instance, kernel principal component analysis [30] directly maximizes the variance expressed by Equation (1.26) [19].

1.6.4 mCI Kernel as a Covariance Kernel

In Section 1.5.1 it was shown that the mCI kernel is indeed a symmetric positive definite kernel. As mentioned in Section 1.2, Parzen [22] showed that any symmetric and positive definite kernel is also a covariance function of a random process defined in the original space of the kernel (see also Wahba [38, Chapter 1]). In the case of the mCI kernel, this means the random processes are defined on $\mathcal{S}(\mathcal{T})$.

Let X denote this random process. Then, for any $s_i \in \mathcal{S}(\mathcal{T})$, $X(s_i)$ is a random variable on a probability space $(\Omega, \mathcal{B}, P)$ with measure P. As proved by Parzen, this random process is Gaussian distributed with zero mean and covariance function

$$
I(s_i, s_j) = E_\omega\left\{X(s_i)X(s_j)\right\}.
\tag{1.27}
$$

Notice that the expectation is over $\omega \in \Omega$ since $X(s_i)$ is a random variable defined on Ω, a situation which can be written explicitly as $X(s_i, \omega)$, $s_i \in \mathcal{S}(\mathcal{T})$, $\omega \in \Omega$. This means that X is actually a doubly stochastic random process. An intriguing perspective is that, for any given ω, $X(s_i, \omega)$ is an ordered and almost surely nonuniform sampling of $X(\cdot, \omega)$. The space spanned by these random variables is $L_2(X(s_i), s_i \in \mathcal{S}(\mathcal{T}))$ since X is obviously square integrable (that is, X has finite covariance).

The RKHS $\mathcal{H}_I$ induced by the mCI kernel and the space of random functions $L_2(X(s_i), s_i \in \mathcal{S}(\mathcal{T}))$ are clearly congruent. This fact is a consequence of the basic congruence theorem [22] since the two spaces have the same dimension or, alternatively, by verifying that the congruence mapping between the two spaces exists. For this reason we may consider the mCI kernel also as a covariance measure of random variables directly dependent on the spike trains with well-defined statistical properties. Allied to our familiarity and intuitive knowledge of the use of covariance (which is nothing but cross-correlation between centered random variables) this concept can be of great importance in optimization and design of optimal

learning algorithms that work with spike trains. This is because linear methods are known to be optimal for Gaussian distributed random variables.

1.7 Principal Component Analysis

To exemplify the importance of the developments shown here, in the following we derive the algorithm to perform principal component analysis (PCA) of spike trains. The PCA algorithm will be derived from two different perspectives to show the generality of an RKHS framework for optimization with spike trains.

First, PCA will be derived directly in the RKHS induced by the mCI kernel. This approach highlights that optimization with spike trains is possible by the definition of an inner product, and more specifically through the mathematical structure provided by the RKHS. This is also the traditional approach in the functional analysis literature [25] and has the advantage of being completely general, regardless of the spike train kernel definition. A well-known example of discrete PCA done in an RKHS is kernel PCA [30].

In the second approach we will derive PCA in the space spanned by the intensity functions utilizing the inner product defined in this space. Since the RKHS is congruent to this space and, therefore, the inner products in the two spaces are isometric the outcome will be found to be the same. However, this approach has the advantage that the eigenfunctions are explicitly available. In general, the eigenfunctions are not available in the RKHS because the transformation to the RKHS is unknown. However, this approach is possible here due to the linearity of the space spanned by the intensity functions with the inner product we defined.

1.7.1 Optimization in the RKHS

Suppose we are given a set of spike trains, $\{s_i \in \mathcal{S}(\mathcal{T}), i = 1, \ldots, N\}$, for which we wish to determine the principal components. Computing the principal components of the spike trains directly is not feasible because we would not know how to define a principal component (PC), however, this is a trivial task in an RKHS.

Let $\{\Lambda_{s_i} \in \mathcal{H}_I, i = 1, \ldots, N\}$ be the set of elements in the RKHS $\mathcal{H}_I$ corresponding to the given spike trains. Denote the mean of the transformed spike trains as

$$\bar{\Lambda} = \frac{1}{N} \sum_{i=1}^{N} \Lambda_{s_i}, \tag{1.28}$$

and the centered transformed spike trains (i.e., with the mean removed) can be obtained as

$$\tilde{\Lambda}_{s_i} = \Lambda_{s_i} - \bar{\Lambda}. \tag{1.29}$$

PCA finds an orthonormal transformation providing a compact description of the data. Determining the principal components of spike trains in the RKHS can be formulated as the problem of finding the set of orthonormal vectors in the RKHS such that the projection of the centered transformed spike trains $\{\tilde{\Lambda}_{s_i}\}$ has the *maximum variance*. This means that the principal components can be obtained by solving an optimization problem in the RKHS. A function $\xi \in \mathcal{H}_I$ (i.e., $\xi : \mathcal{S}(\mathcal{T}) \longrightarrow \mathbb{R}$) is a principal component if it maximizes the cost function

$$J(\xi) = \sum_{i=1}^{N} \left[\operatorname{Proj}_{\xi} (\tilde{\Lambda}_{s_i}) \right]^2 - \rho \left(\|\xi\|^2 - 1 \right), \tag{1.30}$$

where $\operatorname{Proj}_{\xi} (\tilde{\Lambda}_{s_i})$ denotes the projection of the ith centered transformed spike train onto ξ, and ρ is the Lagrange multiplier to the constraint $\left(\|\xi\|^2 - 1 \right)$ imposing that the principal components have unit norm. To evaluate this cost function one needs to be able to compute the projection and the norm of the principal components. However, in an RKHS, an inner product is the projection operator and the norm is naturally defined (see Equation (1.13)). Thus, the above cost function can be expressed as

$$J(\xi) = \sum_{i=1}^{N} \langle \tilde{\Lambda}_{s_i}, \xi \rangle_{\mathcal{H}_I}^2 - \rho \left(\langle \xi, \xi \rangle_{\mathcal{H}_I} - 1 \right). \tag{1.31}$$

Because in practice we always have a finite number of spike trains, ξ is restricted to the subspace spanned by the centered transformed spike trains $\{\tilde{\Lambda}_{s_i}\}$. Consequently, there exist coefficients $b_1, \ldots, b_N \in \mathbb{R}$ such that

$$\xi = \sum_{j=1}^{N} b_j \tilde{\Lambda}_{s_j} = \boldsymbol{b}^T \tilde{\boldsymbol{\Lambda}} \tag{1.32}$$

where $\boldsymbol{b}^T = [b_1, \ldots, b_N]$ and $\tilde{\boldsymbol{\Lambda}}(t) = \left[\tilde{\Lambda}_{s_1}(t), \ldots, \tilde{\Lambda}_{s_N}(t) \right]^T$. Substituting in Equation (1.31) yields

$$\begin{aligned}
J(\xi) &= \sum_{i=1}^{N} \left(\sum_{j=1}^{N} b_j \langle \tilde{\Lambda}_{s_i}, \tilde{\Lambda}_{s_j} \rangle \right) \left(\sum_{k=1}^{N} b_k \langle \tilde{\Lambda}_{s_i}, \tilde{\Lambda}_{s_k} \rangle \right) \\
&\quad + \rho \left(1 - \sum_{j=1}^{N} \sum_{k=1}^{N} b_j b_k \langle \tilde{\Lambda}_{s_i}, \tilde{\Lambda}_{s_k} \rangle \right) \\
&= \boldsymbol{b}^T \tilde{\boldsymbol{I}}^2 \boldsymbol{b} + \rho \left(1 - \boldsymbol{b}^T \tilde{\boldsymbol{I}} \boldsymbol{b} \right),
\end{aligned} \tag{1.33}$$

where $\tilde{\boldsymbol{I}}$ is the Gram matrix of the centered spike trains; that is, the $N \times N$ matrix with elements

$$
\begin{aligned}
\tilde{\boldsymbol{I}}_{ij} &= \left\langle \tilde{\Lambda}_{s_i}, \tilde{\Lambda}_{s_j} \right\rangle \\
&= \left\langle \Lambda_{s_i} - \bar{\Lambda}, \Lambda_{s_j} - \bar{\Lambda} \right\rangle \\
&= \left\langle \Lambda_{s_i}, \Lambda_{s_j} \right\rangle - \frac{1}{N} \sum_{l=1}^{N} \left\langle \Lambda_{s_i}, \Lambda_{s_l} \right\rangle - \frac{1}{N} \sum_{l=1}^{N} \left\langle \Lambda_{s_l}, \Lambda_{s_j} \right\rangle + \frac{1}{N^2} \sum_{l=1}^{N} \sum_{n=1}^{N} \left\langle \Lambda_{s_l}, \Lambda_{s_n} \right\rangle .
\end{aligned}
\tag{1.34}
$$

In matrix notation,

$$
\tilde{\boldsymbol{I}} = \boldsymbol{I} - \frac{1}{N} (\mathbf{1}_N \boldsymbol{I} + \boldsymbol{I} \mathbf{1}_N) + \frac{1}{N^2} \mathbf{1}_N \boldsymbol{I} \mathbf{1}_N,
\tag{1.35}
$$

where $\boldsymbol{I}$ is the Gram matrix of the inner product of spike trains $\boldsymbol{I}_{ij} = \left\langle \Lambda_{s_i}, \Lambda_{s_j} \right\rangle$, and $\mathbf{1}_N$ is the $N \times N$ matrix with all ones. This means that $\tilde{\boldsymbol{I}}$ can be computed directly in terms of $\boldsymbol{I}$ without the need to explicitly remove the mean of the transformed spike trains.

From Equation (1.33), finding the principal components simplifies to the problem of estimating the coefficients $\{b_i\}$ that maximize $J(\xi)$. Since $J(\xi)$ is a quadratic function its extrema can be found by equating the gradient to zero. Taking the derivative with regard to $\boldsymbol{b}$ (which characterizes ξ) and setting it to 0 results in

$$
\frac{\partial J(\xi)}{\partial \boldsymbol{b}} = 2\tilde{\boldsymbol{I}}^2 \boldsymbol{b} - 2\rho \tilde{\boldsymbol{I}} \boldsymbol{b} = 0
\tag{1.36}
$$

and thus corresponds to the eigendecomposition problem[3]

$$
\tilde{\boldsymbol{I}} \boldsymbol{b} = \rho \boldsymbol{b}.
\tag{1.37}
$$

This means that any eigenvector of the centered Gram matrix is a solution of Equation (1.36). Thus, the eigenvectors determine the coefficients of Equation (1.32) and characterize the principal components. It is easy to verify that, as expected, the variance of the projections onto each principal component equals the corresponding eigenvalue. So, the ordering of ρ specifies the relevance of the principal components.

To compute the projection of a given input spike train s onto the kth principal component (corresponding to the eigenvector with the kth largest eigenvalue) we need only to compute in the RKHS the inner product of Λ_s with ξ_k. That is,

---

[3] Note that the simplification in the eigendecomposition problem is valid regardless if the Gram matrix is invertible or not, since $\tilde{\boldsymbol{I}}^2$ and $\tilde{\boldsymbol{I}}$ have the same eigenvectors and the eigenvalues of $\tilde{\boldsymbol{I}}^2$ are the eigenvalues of $\tilde{\boldsymbol{I}}$ squared.

$$\mathrm{Proj}_{\xi_k}(\Lambda_s) = \langle \Lambda_s, \xi_k \rangle_{\mathcal{H}_I}$$

$$= \sum_{i=1}^{N} b_{ki} \langle \Lambda_s, \tilde{\Lambda}_{s_i} \rangle \tag{1.38}$$

$$= \sum_{i=1}^{N} b_{ki} \left(I(s, s_i) - \frac{1}{N} \sum_{j=1}^{N} I(s, s_j) \right).$$

1.7.2 Optimization in the Space Spanned by the Intensity Functions

As before, let $\{s_i \in \mathcal{S}(\mathcal{T}), i = 1, \ldots, N\}$ denote the set of spike trains for which we wish to determine the principal components, and $\{\lambda_{s_i}(t), t \in \mathcal{T}, i = 1, \ldots, N\}$ the corresponding intensity functions. The mean intensity function is

$$\bar{\lambda}(t) = \frac{1}{N} \sum_{i=1}^{N} \lambda_{s_i}(t), \tag{1.39}$$

and, therefore, the centered intensity functions are

$$\tilde{\lambda}_{s_i}(t) = \lambda_{s_i}(t) - \bar{\lambda}(t). \tag{1.40}$$

Again, the problem of finding the principal components of a set of data can be stated as the problem of finding the eigenfunctions of unit norm such that the projections have maximum variance. This can be formulated in terms of the following optimization problem. A function $\zeta(t) \in L_2(\lambda_{s_i}(t), t \in \mathcal{T})$ is a principal component if it maximizes the cost function

$$J(\zeta) = \sum_{i=1}^{N} \left[\mathrm{Proj}_{\zeta}(\tilde{\lambda}_{s_i}) \right]^2 - \gamma \left(\|\zeta\|^2 - 1 \right)$$

$$= \sum_{i=1}^{N} \left\langle \tilde{\lambda}_{s_i}, \zeta \right\rangle_{L_2}^2 - \gamma \left(\|\zeta\|^2 - 1 \right), \tag{1.41}$$

where γ is the Lagrange multiplier constraining ζ to have unit norm. It can be shown that $\zeta(t)$ lies in the subspace spanned by the intensity functions $\{\tilde{\lambda}_{s_i}(t), i = 1, \ldots, N\}$. Therefore, there exist coefficients $b_1, \ldots, b_N \in \mathbb{R}$ such that

$$\zeta(t) = \sum_{j=1}^{N} b_j \tilde{\lambda}_{s_j}(t) = \boldsymbol{b}^T \tilde{\boldsymbol{r}}(t). \tag{1.42}$$

with $\boldsymbol{b}^T = [b_1, \ldots, b_N]$ and $\tilde{\boldsymbol{r}}(t) = \left[\tilde{\lambda}_{s_1}(t), \ldots, \tilde{\lambda}_{s_N}(t) \right]^T$. Substituting in Equation (1.31) yields

$$J(\zeta) = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} b_j \left\langle \tilde{\lambda}_{s_i}, \tilde{\lambda}_{s_j} \right\rangle \right) \left(\sum_{k=1}^{N} b_k \left\langle \tilde{\lambda}_{s_i}, \tilde{\lambda}_{s_k} \right\rangle \right)$$

$$+ \gamma \left(1 - \sum_{j=1}^{N} \sum_{k=1}^{N} b_j b_k \left\langle \tilde{\lambda}_{s_i}, \tilde{\lambda}_{s_k} \right\rangle \right) \tag{1.43}$$

$$= \boldsymbol{b}^T \tilde{\boldsymbol{I}}^2 \boldsymbol{b} + \gamma \left(1 - \boldsymbol{b}^T \tilde{\boldsymbol{I}} \boldsymbol{b} \right),$$

where $\tilde{\boldsymbol{I}}$ is the Gram matrix of the centered intensity functions (i.e., $\tilde{\mathrm{I}}_{ij} = \left\langle \tilde{\lambda}_{s_i}, \tilde{\lambda}_{s_j} \right\rangle_{L_2}$).

As expected, since the inner product is the same and the two spaces are congruent, this cost function yields the same solution. However, unlike the previous, this presentation has the advantage that it shows the role of the eigenvectors of the Gram matrix and, most importantly, how to obtain the principal component functions in the space of intensity functions. From Equation (1.42), the coefficients of the eigenvectors of the Gram matrix provide a weighting for the intensity functions of each spike trains and, therefore, express how important a spike train is to represent others. In a different perspective, this suggests that the principal component functions should reveal general trends in the intensity functions.

1.7.3 Results

To illustrate the algorithm just derived we performed a simple experiment. We generated two template spike trains comprising of 10 spikes uniformly random distributed over an interval of 0.25 s. In a specific application these template spike trains could correspond, for example, to the average response of a culture of neurons to two distinct but fixed input stimuli. For the computation of the coefficients of the eigendecomposition ("training set"), we generated a total of 50 spike trains, half for each template, by randomly copying each spike from the template with probability 0.8 and adding zero mean Gaussian distributed jitter with standard deviation 3 ms. For testing of the obtained coefficients, 200 spike trains were generated following the same procedure. The simulated spike trains are shown in Fig. 1.2.

According to the PCA algorithm derived previously, we computed the eigendecomposition of the matrix $\tilde{\boldsymbol{I}}$ as given by Equation (1.35) so that it solves Equation (1.37). The evaluation of the mCI kernel was estimated from the spike trains according to Equation (1.12) and computed with a Gaussian kernel with size 2 ms. The eigenvalues $\{\rho_l, l = 1, \ldots, 100\}$ and first two eigenvectors are shown in Fig. 1.3. The first eigenvalue alone accounts for more than 26% of the variance of the dataset in the RKHS space. Although this value is not impressive, its importance is clear since it is nearly four times higher than the second eigenvalue (6.6%). Furthermore, notice that the first eigenvector clearly shows the separation between spike trains generated from different templates (Fig. 1.3b). This again can be seen in the first principal component function, shown in Fig. 1.4, which reveals the location of the spike times used to generate the templates while discriminating between them with
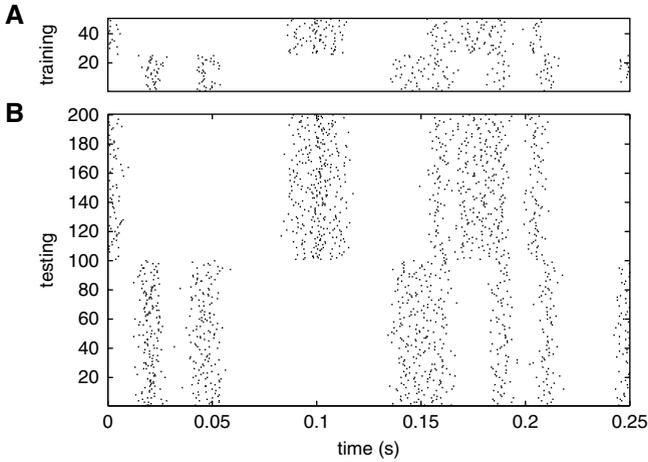
Fig. 1.2: Spike trains used for evaluation of the eigendecomposition coefficients of PCA algorithm (**a**) and for testing of the result (**b**). In either case, the first half of spike trains corresponds to the first template and the remaining to the second template.

(a) Eigenvalues $\{\rho_l, l = 1, \ldots, 100\}$ in decreasing order.

(b) First two eigenvectors.
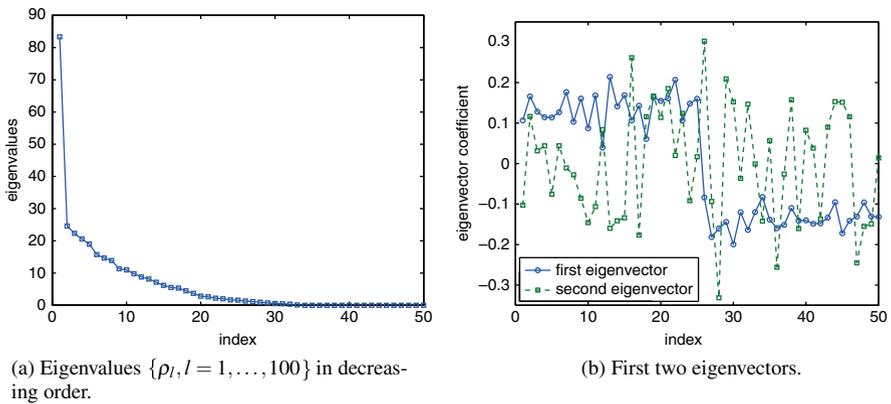
Fig. 1.3: Eigendecomposition of the spike trains Gram matrix $\tilde{I}$.

opposite signs. Around periods of time where the spikes from both templates overlap, the first principal component is 0. As can be seen from the second principal component function, the role of the second eigenvector is to account for the dispersion in the data capable of differentiate spike trains generated from different templates.

Both datasets, for evaluation and testing, were projected onto the first two principal components. Figure 1.5 shows the projected spike trains. As noted from the difference between the first and second eigenvalues, the first principal component
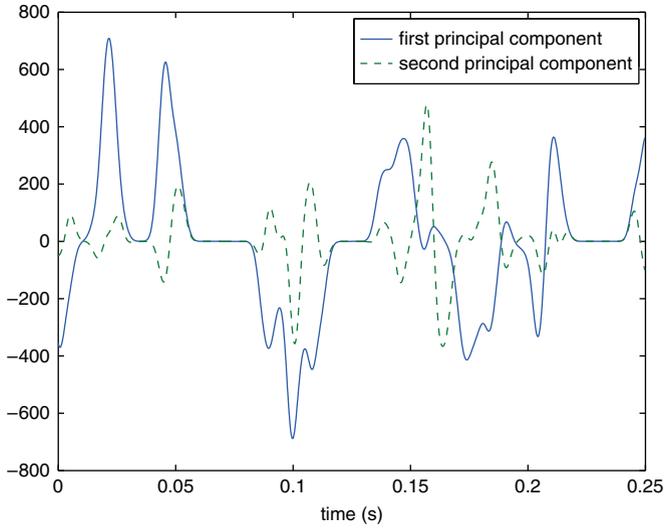
Fig. 1.4: First two principal component functions (i.e., eigenfunctions) in the space of intensity functions. They are computed by substituting the coefficients of the first two eigenvectors of the Gram matrix in Equation (1.42).

(a) Projection of the spike trains in the training set.

(b) Projection of the spike trains in the testing set.
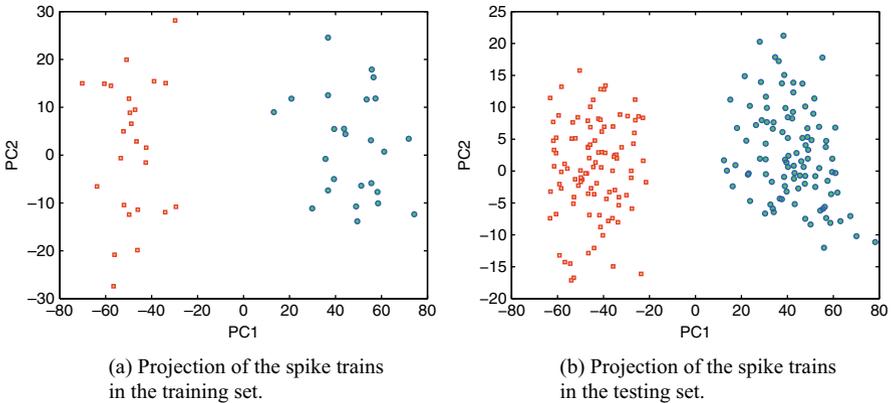
Fig. 1.5: Projection of spike trains onto the first two principal components. The different point marks differentiate between spike trains corresponding to each one of the classes.

is the main responsible for the dispersion between classes of the projected spike trains. This happens because the direction of maximum variance is the one that passes through both clusters of points in the RKHS due to the small dispersion within class. The second principal component seems to be responsible for dispersion due to the jitter noise introduced in the spike trains and suggests that other principal components play a similar role.

A more specific understanding can be obtained from the considerations done in Section 1.6.3. There, the congruence between the RKHS induced by the mCI kernel, $\mathcal{H}_I$, and the RKHS induced by κ, $\mathcal{H}_\kappa$, was utilized to show that the mCI kernel is inversely related to the variance of the transformed spike times in $\mathcal{H}_\kappa$. In this dataset and for the kernel size utilized, this guaranties that the value of the mCI kernel within class is always smaller than interclass. This is a reason why in this scenario the first principal component always suffices to project the data in a way that distinguishes between spike trains generated each of the templates.

Conventional PCA was also applied to this dataset by binning the spike trains. Although cross-correlation is an inner product for spike trains and, therefore, the above algorithm could have been used, for comparison, the conventional approach was followed [27, 15]. That is, to compute the covariance matrix with each binned spike train taken as a data vector. This means that the dimensionality of the covariance matrix is determined by the number of bins per spike train, which may be problematic if long spike trains are used or small bin sizes are needed for high temporal resolution.

The results of PCA using bin size of 5 ms are shown in Figs. 1.6 and 1.7. The bin size was chosen to provide a good compromise between temporal resolution and smoothness of the eigenfunctions (important for interpretability). Comparing these results the ones using the mCI kernel, the distribution of the eigenvalues is quite similar and the first eigenfunction does reveal somewhat of the same trend as in Fig. 1.4. The same is not true for the second eigenfunction, however, which looks much more "jaggy." In fact, as Fig. 1.7 shows, in this case the projections along the first two principal directions are not orthogonal. This means that the covariance matrix does not fully express the structure of the spike trains. It is noteworthy that this is not only because the covariance matrix is being estimated with a small number of data vectors. In fact, even if the binned cross-correlation was utilized directly in the above algorithm as the inner product the same effect was observed, meaning that the *binned cross-correlation does not characterize the spike train structure in*

(a) Eigenvalues in decreasing order.

(b) First two eigenvectors/eigenfunctions.
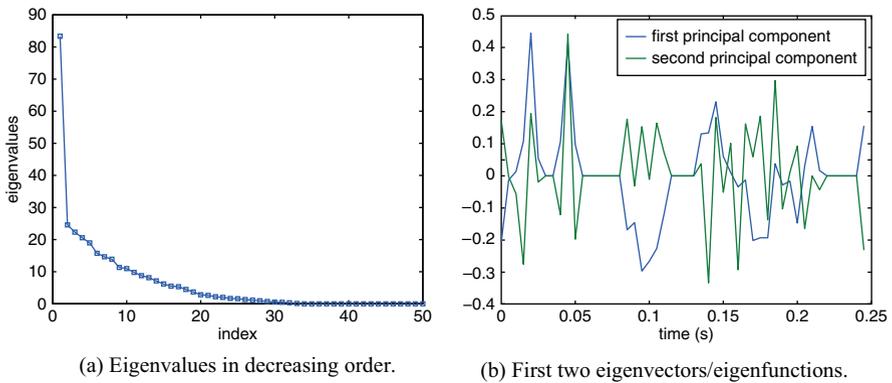
Fig. 1.6: Eigendecomposition of the binned spike trains covariance matrix.

(a) Projection of the spike trains in the training set. (b) Projection of the spike trains in the testing set.
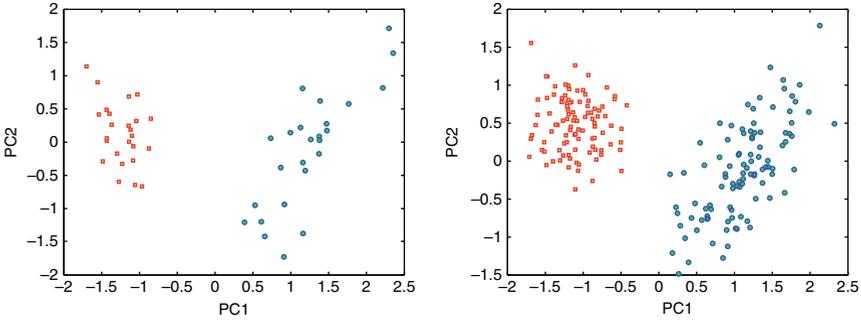
Fig. 1.7: Projection of spike trains onto the first two principal components of the covariance matrix of binned spike trains. The different point marks differentiate between spike trains corresponding to each one of the classes.

sufficient detail. Since the binned cross-correlation and the mCI kernel are conceptually equivalent apart from the discretization introduced by binning, this shows the ill effects of this preprocessing step for analysis and computation with spike train, and point process realizations in general.

1.8 Conclusion

A reproducing kernel Hilbert space (RKHS) framework for optimization with spike trains is introduced. Although the application of kernel methods to spike trains without binning is not entirely novel [4, 31], a more general view of the problem is presented. Instead of a top-down approach often taken in kernel methods, the mCI kernel was built bottom-up from the concept of intensity functions which are basic statistical descriptors of spike trains. Indeed, intensity functions are the core concept of the statistical analysis of spike trains and is perhaps one of reasons why binning is such a well-established technique, at any timescale of interest [28, 5]. Kernel methods applied before to spike trains seemed to have no connection to intensity estimation. This chapter, however, bridges these two perspectives seamlessly. In one perspective, the mCI kernel approximates our intuitive understanding regarding intensity functions as functional descriptors of point processes. On the other hand, the evaluation (or estimation) of the mCI kernel for given spike trains easily links to other methodologies in the literature. Most importantly, the approach taken lends itself to generalization to other point process models and spike train kernels nonlinear in the space of intensity functions taking advantage of the RKHS mathematical structure and without sacrifice in rigor.

In addition to this enlightening connection of point of view, the rigorous yet general mathematical approach toward the problem of optimization for manipulating

spike trains clarifies exactly from basic principles which kernels can be used and what are the general properties of the mCI kernel defined. Even though it may be argued that kernel methods can be applied directly for spike trains data given a kernel, the true meaning of using such a kernel cannot be well determined. This is one of the strengths of the explicit construction followed. In this way, the general structure of the RKHS space induced is well understood allowing for methods to be derived from their basic ideas. Additionally, we were able to establish a close mathematical relationship to several congruent spaces where the derived methods can be thoroughly comprehended. Still, it must be remarked that the mCI kernel presented here will likely not be the most appropriate for a number of problems. This was not the goal of this chapter. Instead one of our aims was to show how other kernels that operate with spike trains may be easily formulated. Depending on a specific application other kernels may be defined which lead to simpler solutions and/or are computationally simpler.

It is noteworthy that the mCI kernel is not restricted to applications with spike trains but rather can be applied to processing with any Poisson point processes. In fact, the mCI kernel can be applied for even more general point processes. Naturally, it might not be the optimum inner product for point processes other than Poisson processes since the intensity function does not fully characterizes the process but, in a sense, this is similar to the use of cross-correlation in continuous random processes, which is only sensitive to second-order statistics.

Appendix: Proofs

This section presents the proofs for Properties 1.2, 1.5, and 1.6 in Section 1.5.1.

Proof (Property 1.2). The symmetry of the matrix results immediately from Property 1.1.

By definition, a matrix is nonnegative definite if and only if $\boldsymbol{a}^T \boldsymbol{V} \boldsymbol{a} \geq 0$, for any $\boldsymbol{a}^T = [a_1, \ldots, a_n]$ with $a_i \in \mathbb{R}$. So, we have that

$$\boldsymbol{a}^T \boldsymbol{V} \boldsymbol{a} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j I(s_i, s_j), \qquad (1.44)$$

which, making use of the mCI kernel definition (Equation (1.12)), yields

$$\boldsymbol{a}^T \boldsymbol{V} \boldsymbol{a} = \int_{\mathcal{T}} \left(\sum_{i=1}^{n} a_i \lambda_{s_i}(t) \right) \left(\sum_{j=1}^{n} a_j \lambda_{s_j}(t) \right) dt$$

$$= \left\langle \sum_{i=1}^{n} a_i \lambda_{s_i}, \sum_{j=1}^{n} a_j \lambda_{s_j} \right\rangle_{L_2(\mathcal{T})}$$

$$= \left\| \sum_{i=1}^{n} a_i \lambda_{s_i} \right\|_{L_2(\mathcal{T})} \geq 0, \tag{1.45}$$

since the norm is nonnegative.

Proof (Property 1.5). Consider the 2×2 CI kernel matrix,

$$\boldsymbol{V} = \begin{bmatrix} I(s_i, s_i) & I(s_i, s_j) \\ I(s_j, s_i) & I(s_j, s_j) \end{bmatrix}.$$

From Property 1.2, this matrix is symmetric and nonnegative definite. Hence, its determinant is nonnegative [7, p. 245]. Mathematically,

$$\det(\boldsymbol{V}) = I(s_i, s_i) I(s_j, s_j) - I^2(s_i, s_j) \geq 0,$$

which proves the result of Equation (1.16).

Proof (Property 1.6). Consider two spike trains, $s_i, s_j \in \mathcal{S}(\mathcal{T})$. The norm of the sum of two spike trains is

$$\left\| \lambda_{s_i} + \lambda_{s_j} \right\|^2 = \left\langle \lambda_{s_i} + \lambda_{s_j}, \lambda_{s_i} + \lambda_{s_j} \right\rangle \tag{1.46a}$$

$$= \left\langle \lambda_{s_i}, \lambda_{s_i} \right\rangle + 2 \left\langle \lambda_{s_i}, \lambda_{s_j} \right\rangle + \left\langle \lambda_{s_j}, \lambda_{s_j} \right\rangle \tag{1.46b}$$

$$\leq \left\| \lambda_{s_i} \right\|^2 + 2 \left\| \lambda_{s_i} \right\| \left\| \lambda_{s_j} \right\| + \left\| \lambda_{s_j} \right\|^2 \tag{1.46c}$$

$$= \left(\left\| \lambda_{s_i} \right\| + \left\| \lambda_{s_j} \right\| \right)^2, \tag{1.46d}$$

with the upper bound in step 1.46c established by the Cauchy–Schwarz inequality (Property 1.5).

References

1. Aronszajn, N. Theory of reproducing kernels. Trans Am Math Soc **68**(3), 337–404 (1950)
2. Berg, C. Christensen, J.P.R., Ressel, P. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions. Springer-Verlag, New York (1984)
3. Bohte, S.M., Kok, J.N., Poutré, H.L.: Error-backpropagation in temporally encoded networks of spiking neurons. Neurocomputing **48**(1–4), 17–37 (2002). DOI 10.1016/S0925-2312(01)00658-0
4. Carnell, A., Richardson, D.: Linear algebra for time series of spikes. In: Proceedings European Symposium on Artificial Neural Networks, pp. 363–368. Bruges, Belgium (2005)
5. Dayan, P., Abbott, L.F. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. MIT Press, Cambridge, MA (2001)
6. Diggle, P., Marron, J.S. Equivalence of smoothing parameter selectors in density and intensity estimation. J Acoust Soc Am **83**(403), 793–800 (1988)
7. Harville, D.A. Matrix Algebra from a Statistician's Perspective. Springer, New York (1997)

8. Haykin, S. Adaptive Filter Processing, 4th edn. Prentice-Hall, Upper Saddle River, NJ (2002)
9. Kailath, T. RKHS approach to detection and estimation problems–part I: Deterministic signals in gaussian noise. IEEE Trans Inform Theory **17**(5), 530–549 (1971)
10. Kailath, T., Duttweiler, D.L. An RKHS approach to detection and estimation problems–part III: Generalized innovations representations and a likelihood-ratio formula. IEEE Trans Inform Theory **18**(6), 730–745 (1972)
11. Kailath, T., Weinert, H.L. An RKHS approach to detection and estimation problems–part II: Gaussian signal detection. IEEE Trans Inform Theory **21**(1), 15–23 (1975)
12. Maass, W., Bishop, C.M. (eds.) Pulsed Neural Networks. MIT Press, Cambridge, MA (1998)
13. Maass, W., Natschläger, T., Markram, H. Real-time computing without stable states: A new framework for neural computation based on perturbations. Neural Comp **14**(11), 2531–2560 (2002). DOI 10.1162/089976602760407955
14. Mardia, K.V., Jupp, P.E. Directional Statistics. John Wiley & Sons, West Sussex, England (2000)
15. McClurkin, J.W., Gawne, T.J., Optican, L.M., Richmond, B.J. Lateral geniculate neurons in behaving primates. II. Encoding of visual information in the temporal shape of the response. J Neurophysiol **66**(3), 794–808 (1991)
16. Mercer, J. Functions of positive and negative type, and their connection with the theory of integral equations. Phil Trans R Soc Lond – A **209**, 415–446 (1909)
17. Moore, E.H. On properly positive Hermitian matrices. Bull Am Math Soc **23**, 59 (1916)
18. Paiva, A.R.C., Park, I., Príncipe, J.C. Reproducing kernel Hilbert spaces for spike train analysis. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-2008, Las Vegas, NV, USA (2008)
19. Paiva, A.R.C., Xu, J.W., Príncipe, J.C. Kernel principal components are maximum entropy projections. In: Proceedings of International Conference on Independent Component Analysis and Blind Source Separation, ICA-2006, pp. 846–853. Charleston, SC (2006). DOI 10.1007/11679363_105
20. Parzen, E. Statistical inference on time series by Hilbert space methods. Tech. Rep. 23, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, CA (1959)
21. Parzen, E. On the estimation of a probability density function and the mode. Ann Math Stat **33**(2), 1065–1076 (1962)
22. Parzen, E. Time Series Analysis Papers. Holden-Day, San Francisco, CA (1967)
23. Parzen, E. Statistical inference on time series by RKHS methods. In: Pyke, R. (ed.) Proceedings of 12th Biennal International Seminar of the Canadian Mathematical Congress, pp. 1–37 (1970)
24. Príncipe, J.C., Xu, D., Fisher, J.W. Information theoretic learning. In: Haykin, S. (ed.) Unsupervised Adaptive Filtering, vol. 2, pp. 265–319. John Wiley & Sons, New York (2000)
25. Ramsay, J.O., Silverman, B.W. Functional Data Analysis. Springer-Verlag, New York (1997)
26. Reiss, R.D. A Course on Point Processes. Springer-Verlag, New York (1993)
27. Richmond, B.J., Optican, L.M. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. J Neurophysiol **51**(1), 147–161 (1987)
28. Rieke, F., Warland, D., de Ruyter van Steveninck, R., Bialek, W. Spikes: Exploring the Neural Code. MIT Press, Cambridge, MA (1999)
29. Schölkopf, B., Burges, C.J.C., Smola, A.J. (eds.) Advances in Kernel Methods: Support Vector Learning. MIT Press, Cambridge, MA (1999)
30. Schölkopf, B., Smola, A., Müller, K.R. Nonlinear component analysis as a kernel eigenvalue problem. Neural Comp **10**(5), 1299–1319 (1998)
31. Schrauwen, B., Campenhout, J.V. Linking non-binned spike train kernels to several existing spike train distances. Neurocomputing **70**(7–8), 1247–1253 (2007). DOI 10.1016/j.neucom.2006.11.017
32. Schreiber, S., Fellous, J.M., Whitmer, D., Tiesinga, P., Sejnowski, T.J. A new correlation-based measure of spike timing reliability. Neurocomputing **52–54**, 925–931 (2003). DOI 10.1016/S0925-2312(02)00838-X

33. Snyder, D.L. Random Point Process in Time and Space. John Viley & Sons, New York (1975)
34. van Rossum, M.C.W. A novel spike distance. Neural Comp **13**(4), 751–764 (2001)
35. Vapnik, V.N. The Nature of Statistical Learning Theory. Springer, New York (1995)
36. Victor, J.D., Purpura, K.P. Nature and precision of temporal coding in visual cortex: A metric-space analysis. J Neurophysiol **76**(2), 1310–1326 (1996)
37. Victor, J.D., Purpura, K.P. Metric-space analysis of spike trains: theory, algorithms, and application. Netw Comp Neural Sys **8**, 127–164 (1997)
38. Wahba, G. Spline Models for Observational Data, *CBMS-NSF Regional Conference Series in Applied Mathematics*, Vol. 59. SIAM (1990)

Chapter 2
Investigating Functional Cooperation in the Human Brain Using Simple Graph-Theoretic Methods

Michael L. Anderson, Joan Brumbaugh, and Aysu Şuben

Abstract This chapter introduces a very simple analytic method for mining large numbers of brain imaging experiments to discover functional cooperation between regions. We then report some preliminary results of its application, illustrate some of the many future projects in which we expect the technique will be of considerable use (including a way to relate fMRI to EEG), and describe a research resource for investigating functional cooperation in the cortex that will be made publicly available through the lab web site. One significant finding is that differences between cognitive domains appear to be attributable more to differences in patterns of cooperation between brain regions, rather than to differences in which brain regions are used in each domain. This is not a result that is predicted by prevailing localization-based and modular accounts of the organization of the cortex.

2.1 Introduction and Background

Hardly an issue of science or nature goes by without creating a stir over the discovery of "the" gene for some disease, trait, or predisposition, or "the" brain area responsible for some behavior or cognitive capacity. Of course, we know better; the isolable parts of complex systems like the brain or the human genome do what they do only in virtue of the cooperation of very many other parts, and often only by operating within and taking advantage of specific environmental and developmental

Michael L. Anderson
Department of Psychology, Franklin and Marshall College, Lancaster, PA 17604, USA: Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA,
e-mail: `michael.anderson@fandm.edu`

Joan Brumbaugh
Department of Psychology, Franklin and Marshall College, Lancaster, PA 17604, USA

Aysu Şuben
Department of Psychology, Franklin and Marshall College, Lancaster, PA 17604, USA

contexts. But while it is true that we have gotten better about acknowledging the limitations of our instinctive reductionism – a bit of humility that the media would do well to absorb into its reporting – actual scientific practice has yet to be much affected by awareness of those limits. A recent case in point is John Anderson's project to map ACT-R components to brain regions [3]. The motivations for the project are of course entirely sound: if ACT-R is to be a realistic model of human cognition, then that model ought to have some significant, testable relationship to the neural bases of cognition. In this particular set of experiments, the authors identify eight ACT-R modules and match each one to a different region of interest. They then look for, and find, significant fit between the predictions for the BOLD signal in those regions, based on the activity of the ACT-R modules while solving a particular arithmetic task, and the measured BOLD signal in human participants performing the same task. On its face, this is an intriguing result and seems to offer compelling support for the ACT-R model. But the methodological assumption of the project – that there is a 1:1 mapping of ACT-R modules and brain areas – is highly suspect. Nor are the authors unaware of this difficulty, and in fact they specifically caution against making any inference from their approach to the functional organization of the brain:

> Some qualifications need to be made to make it clear that we are not proposing a one-to-one mapping between these eight regions and the eight functions. First, other regions also serve these functions. Many areas are involved in vision and the fusiform gyrus has just proven to be the most useful to monitor. Similarly, many regions have been shown to be involved in retrieval, particularly the hippocampus. The prefrontal region is just the easiest to identify and seems to afford the best signal-to-noise ratio. Equally, we are not claiming these regions only serve one function. This paper has found some evidence for multiple functions. For instance, the motor regions are involved in rehearsal as well as external action (213–4).

Although we should appreciate the authors' candor here, the caveat seriously undermines the ability to interpret their results. If from the discovery that activity in an ACT-R module predicts the BOLD signal in specific brain region, we can neither infer that the region serves that specific function (because it is also activated in other tasks), nor that the function is served by that region (because other regions are activated by the target task), then we are not left with much. And yet despite the authors' awareness of these problems, they stick by the methodology that causes them.

Why might this be so? Naturally, all scientists are faced with the necessity of making simplifying abstractions to increase the tractability of their work; but as the authors found themselves, the assumption of a 1:1 mapping of modules to brain areas is not an approximation to reality, but appears to be fundamentally misleading. So what would account for the fact that they persist in applying methodological assumptions that they know to be inadequate? Given the scientific stature of the

authors, the question prompts reflection on the range and adequacy of the methodological tools actually available for work in this area. One sticks with improper tools only when the other options appear even worse. And while there are indeed more sophisticated tools for cooperation-sensitive investigations of neuroscientific data, those techniques are typically highly complex, hard to master, and – most importantly – produce results that can be difficult to interpret.

To help address these related problems, this chapter will describe a very simple analytical technique that we have been using in our lab to make cooperation-sensitive investigations tractable. In this chapter, we will outline that method, report some preliminary results of its application, and illustrate some of the many future projects in which we expect this technique (and the underlying database of brain imaging studies) will be of considerable use.

2.2 Graph Theory and Neuroscience

A graph is a set of objects called points, vertices, or nodes connected by links called lines or edges. Graphs have proven to be a convenient format to represent relationships in very many different areas, including computer networks, telephone calls, airline route maps, and social interactions [18, 19]. In neuroscience, graphs have been used for such purposes as investigating neural connectivity patterns [27], correcting brain images [17], and analyzing the patterns of neural activations in epilepsy [32]. Nevertheless graphs and graph theory – the branch of mathematics concerned with exploring the topological properties of graphs [15] – remain at this time underutilized tools with enormous potential to advance our understanding of the operations of the brain.

Our approach to investigating functional cooperation in the cortex involves building co-activation graphs, based on applying some simple data analysis techniques to large numbers of brain imaging studies. The method consists of two steps: first, choosing a spatial segmentation of the cortex to represent as nodes (current work uses Brodmann areas, but alternate segmentation schemes could easily be used; see below); and second, performing some simple analyses to discover which regions – which nodes – are statistically likely to be co-active. These relationships are represented as edges in our graphs.

For this second step we proceed in the following way. Given a database of brain imaging studies containing information about brain activations in various contexts (we describe the particular database we have been using in the next section), we first determine the chance likelihood of activation for each region by dividing the number of experiments in which it is reported to be active by the total number of experiments in the database. Then, for each pair of regions, we use a χ^2 measure to determine if the regions are more (or less) likely to be co-active than would be predicted by chance. We also perform a binomial analysis, since a binomial measure can provide directional information. (It is sometimes the case that, while area A and area B are co-active more (or less) often than would be predicted by chance, the

effect is asymmetric, such that area B is more active when area A is active, but not the reverse.)

Figure 2.1 shows the results of one such analysis, for a set of action and attention tasks. The graphs represent Brodmann areas that are significantly more likely than chance to be co-active ($\chi^2 > 3.84$); it is hypothesized that the network of co-activated areas revealed by such analysis represents those areas of the cortex that cooperate to perform the cognitive tasks in the given domain. The co-activation graphs are superimposed on an adjacency graph (where edges indicate that the Brodmann areas share a physical border in the brain) for ease of visual comparison.

<div align="center">Action   Attention</div>
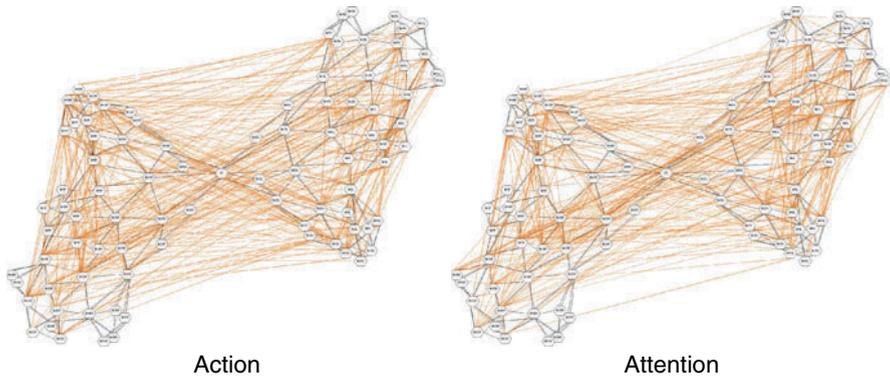
Fig. 2.1: Cortex represented as adjacency + co-activation graphs. Here the Brodmann areas are nodes, with *black lines* between adjacent areas and *orange lines* between areas showing significant co-activation. The graph on the *left* shows co-activations from 56 action tasks, and the graph on the *right* shows co-activations from 77 attention tasks. Edges determined using the threshold $\chi^2 > 3.84$. Graphs rendered with aiSee v. 2.2.

Note that co-activation analysis is similar to, but distinct from, the approach adopted by [31] in discovering "functional connectivity." The main difference is that edges in functional connectivity graphs indicate temporal co-variation between brain regions. Moreover, the results they report generally represent the dynamics of simulated neural networks (based on the structure of biological brain networks), rather than the static analysis of data-mining imaging experiments. Hence we adopt the term "functional cooperation" to distinguish our results from theirs. Nevertheless, there is presumably much to be gained by leveraging both sorts of analysis; in a later section we describe one such future project for bringing co-activation and co-variation graphs together.

The results of such analysis are not just visually striking, but afford the application of some well-understood mathematical techniques to better understand features of brain organization and functional cooperation. Of course, exactly what sorts of

techniques are appropriate, and how the end results should be interpreted, depend a great deal on the nature of the underlying data. Thus, in the next section we describe the database that we have been working with and how other researchers can get access to it for their own use. Then, in the final section, we will describe some of the projects to which we have applied this resource and some of the future possibilities.

2.3 A Database of Imaging Experiments

Over the last year or so we have compiled a database containing 665 experiments in 18 cognitive domains. The database currently consists of every qualifying imaging study in the *Journal of Cognitive Neuroscience* from 1996 to 2006, as well as the 135 experiments from [11] that were used in previous studies [4, 6]. To qualify for inclusion in the database, the study had to be conducted on healthy adults and to use a subtraction-based methodology for analyzing results. The database contains only post-subtraction activations. The data recorded for each experiment include the publication citation, the domain and sub-domain, the imaging method, the Talairach coordinates of each reported activation, the Brodmann area of each reported activation, the relative placement of the activation in the Brodmann area (e.g., frontal, caudal, ventral, dorsal), and the comparison used to generate the results. The domain labels are consistent with those used by the BrainMap database [22]. For experiments where coordinates were reported in MNI coordinates, a software package called GingerALE was used to translate these into Talairach coordinates [21]. When the authors of the study reported the Brodmann areas of their activations, these were recorded as reported. Where the authors did not report Brodmann areas, a software package called the Talairach demon [24] was used to provide Brodmann area labels for the coordinates. This program reports a distance in millimeters from the coordinate to the reported Brodmann area; this is the range, and it is recorded in cases where the BA label was generated using the software. The range is useful for excluding from analysis Brodmann area labels for coordinates that are further than desired from the reported area. Our plans are to continue to add to the database and analysis, and to publish versions at 1 year intervals beginning in the fall of 2008. The published versions of the database will contain the base data detailed above, as well as co-activation graphs, and will be prepared according to the following procedure: first, we will only include in the co-activation analysis sample domains containing some minimum number of experiments (e.g., 50 or 100, to be determined by what is feasible given the state of the database at that time). Having identified these domains, we will generate a concordance of authors to be sure that no individual labs are overrepresented in any given domain. The samples will be balanced by lab by randomly excluding experiments from overrepresented authors. At this point we will choose a target n based on the number of experiments in the domain containing the fewest number of experiments. An equal number of experiments will be

randomly selected from the remaining domains. This set of experiments, equally balanced between the domains, will be the sample for that year's co-activation analysis.

On this balanced sample we will run at least the following kinds of analysis. (1) For each domain, and for the entire set, we will generate a co-activation graph, constructed using the method outlined above, using Brodmann areas as nodes, and including only activations with a range (see above) of less than 5 mm. The calculated chance of activation and co-activation, as well as the binomial probability and χ^2 value will be reported for each pair of Brodmann areas, allowing researchers to set their own probability thresholds. (2) For each of the co-activation graphs, we will do a clique analysis (see below). Lancaster et al. [23] review some methods for generating cliques from brain activation data, and there are many other well-established methods for extracting cliques of various descriptions from graphs [1, 8, 9, 16]. Finally, (3) for all of the co-activation graphs and cliques, we will project them onto the adjacency graph (shown above) and calculate the average minimum graph distance (the "scatter" in the cortex) of the included nodes. All of this data will be made available for download from the lab web site, at

```
http://www.agcognition.org/brain_network
```

Before moving on to the next section, where we describe some of the uses to which these data have been put, and how it can be applied in the future, it is worth saying a word about our reliance on Brodmann areas as the basis for the analyses. It is of course legitimate to wonder whether the sub-division of the cortex into Brodmann areas will be a feature of our final functional map of the human brain; one rather suspects it will be fully superseded by some yet-to-be developed topographical scheme. Yet Brodmann areas remain the lingua franca in Cognitive Neuroscience for reporting findings, and sticking to this tradition will make results using these analyses easier to relate to past findings. Moreover, for the purposes we have described here – investigating the functional cooperation between brain areas involved in supporting different functions – virtually any consistent spatial division of the brain will do, and regions the size of Brodmann areas offer adequate spatial resolution for the required analysis. For, while the spatial resolution of a single fMRI image is on the order of 3 mm or better, there are questions both about the accuracy and precision of repeated fMRI, both within and between participants, effectively reducing its functional resolution [28]. It is arguable, then, that the use of Brodmann-sized regions of the cortex for representing the contribution of individual brain areas to cognitive tasks is consistent with the realistic (conservatively estimated) spatial resolution of current imaging technologies [10, 34]. In any case, it should be noted that the coordinates of each activation are also recorded in the database; if a Brodmann-based spatial scheme does not appear to produce useful or legitimate results, other spatial divisions of the cortex can certainly be substituted, and the very same sort of analysis performed. For instance, one can use the ALE (activation likelihood estimates) paradigm [33] to extract probable activations for arbitrarily defined neural volumes and build graphs from these data [23].

2.4 The Usefulness of Co-activation Graphs

With brain imaging data in this format, it becomes possible to formulate some very simple questions and use some well-understood methods to answer them. For instance, a long-standing project in our lab has been adjudicating between functional topographies of the brain based on the principle of localization and those based on the principle of redeployment. Localization-based approaches to functional topography, insofar as they typically expect brain regions to be dedicated to a small and domain-restricted set of cognitive functions, would be committed to the notion that differences in cognitive domains would be reflected primarily in differences in which brain regions support tasks in the domain. In contrast, redeployment-based approaches, being based on the idea that most brain regions are used in many different tasks across cognitive domains, would expect very little difference in which brain regions were used in each domain. However, because redeployment nevertheless expects brain regions to have fixed low-level functions [3–5], it is committed to the notion that differences in functions and domains must instead be the result of differences in the ways in which the areas cooperate in supporting different tasks. To put this in more concrete visual terms, imagine a simplified brain with six regions that together support two different cognitive domains. If one supports a localization-based (or a classical modular) organization for the brain, one would expect the regional cooperation patterns to look like those in the diagram on the left. In contrast, redeployment predicts an organization that looks something more like that shown in the diagram on the right (Fig. 2.2).

Localization-based organization   **Redeployment-based organization**
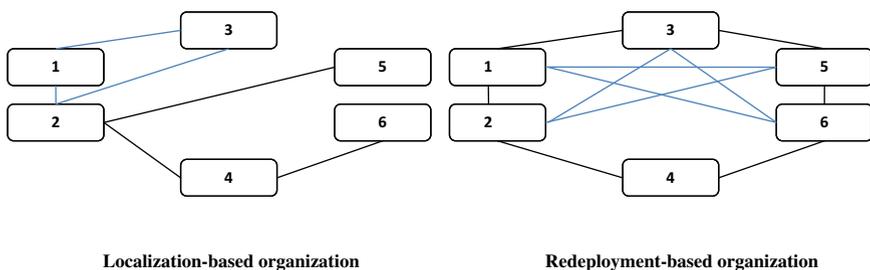
Fig. 2.2: Two different possibilities for the functional organization or the cortex. Figure shows an imagined brain with six regions supporting two cognitive domains. Localization predicts that domain 1 (*blue*) and domain 2 (*black*) will utilize different brain areas, while redeployment predicts that the domains will utilize many of the same brain areas, cooperating in different patterns.

There is an obvious analog for these features in our co-activation graphs: comparing the graphs from different domains, node overlaps indicate Brodmann areas that support tasks in both domains, whereas edge overlaps would indicate a similar pattern of cooperation between Brodmann areas. Thus, localization predicts little node

overlap between co-activation graphs (and therefore also low edge overlap), while redeployment predicts a great deal of node overlap, but little edge overlap. Using our database of imaging data, we did a co-activation analysis for the eight cognitive domains having more than 30 experiments: action; attention; emotion; language; memory; mental imagery; reasoning; and visual perception. The number of experiments (472 total) was not balanced between domains and authors, but otherwise followed the procedures outlined above. Using Dice's coefficient as our measure $(d = 2(o1,2)/(n1+n2))$, where o is the number of overlapping elements and n is the total number of elements in each set), we compared the amount of node and edge overlap between each of the eight domains. As predicted by redeployment, we found a high degree of node overlap ($d = 0.81$, SD $= 0.04$) but very little edge overlap ($d = 0.15$, SD $= 0.04$). The difference is significant (two-sample Student's-t test, double-sided $p << 0.001$). Figure 2.3 shows a graph of the results. This is just one among a number of findings that suggest that redeployment is the better supported approach to understanding the functional topography of the cortex [4–6].
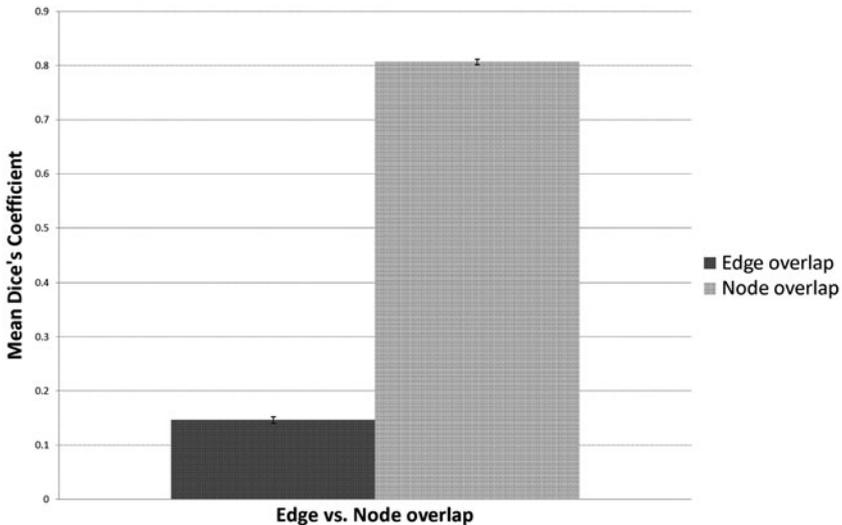
Fig. 2.3: Mean overlap of nodes vs. edges. A graph of the average Dice's coefficient for similarity between the sets of nodes and edges in a pair-wise comparison of co-activation graphs from eight cognitive domains. Difference between the means is significant ($p << 0.001$).

Looking at node and edge overlaps is just a simple example of the sorts of comparisons one might make using data in this format. Others more specific to graph-based representations also readily suggest themselves. For instance, one common form of analysis in graphs is a clique analysis, so called because of its origin in the analysis of social networks [2]. A clique is a maximal complete sub-graph – that

is, a set of nodes in a graph that are fully connected with one another, but not fully connected with any other node in the whole graph. In this context, a clique would indicate a set of Brodmann areas that are fully co-active with each other, but not with other areas of the brain; any such neural cliques would obviously be structures of interest. As in the case of social networks, however, this definition may be too strict for many purposes. Intuitively, we would be interested in sets of nodes that are cohesive and relatively isolated – that is, nodes that are highly but not necessarily fully connected, and much more connected with each other than with other nodes in the graph. These would represent sets of brain regions that are generally co-active with each other, but that operate with relative independence from the rest of the brain. Alba [2] offers the notion of a sociometric clique (an *n*-clique of diameter *n*), as well as measures of cohesiveness and isolation, that could be adopted here to discover sets of brain regions with the desired properties. Cohesive, isolated sociometric cliques seem likely to correspond to the neural components that cooperate to support a set of closely related cognitive functions or sub-functions. Whether this is so is an open scientific question, but such cliques are a far more plausible target for investigations into the neural components supporting particular cognitive functions than are individual brain areas. To return us to the issue with which this chapter began: co-activation graphs allow one to discover (among other things) neural cliques; in our view, what Anderson et al. should be doing is trying to match ACT-R modules to these sorts of structures, and not to individual brain areas.

These are far from the only research avenues that these data offer. One can also look at other features of the graphs, such as local topography, which may help make plausible inferences about underlying function. For instance, a hub-and-spoke pattern of co-activation may indicate broadcast or information consolidation functions; in contrast, long strings of connected nodes might indicate serial processing.

We could go on indefinitely, but the point is not to exhaustively list all the possible analyses one might make with graph-based co-activation data. Instead we would like to take the opportunity to call to mind the fact that, at very many points in the history of science, great progress has been made just in virtue of finding the right format for otherwise well-known data. In a field as young as Cognitive Neuroscience it is still more than possible for simple ideas to make a transformative impact; co-activation graphs may be one of those ideas.

2.5 Relating fMRI to EEG

We would like to conclude by describing one longer term application of co-activation graphs about which we are especially excited. As the reader is no doubt aware, a long-standing issue in experimental and clinical neuroscience has been the question of how to relate data from EEG/MEG to fMRI. Chief among the many obstacles standing in the way of relating the two have been (1) questions over whether each technology measures the same underlying neural activity [26] and (2) difficulty in finding the right representational format for the relation, given the vastly differ-

ent temporal scale of the two data streams [20]. However, recent research seems to indicate a mitigation of the first issue; and co-activation graphs may contribute to a novel approach to the second. We will discuss each of these in turn.

Although there have been for some time, and continue to be, questions about the neurophysiologial bases of the fMRI signal, converging evidence strongly suggests that the BOLD signal is best correlated with local field potentials [25, 7, 35]. This is good news for the project of relating EEG and fMRI, because recent work has shown that EEG signals can also be analyzed to give estimations of LFP [29, 30]. Although this is hardly to be considered the last word on the subject, it appears that differences in underlying neurophysiological basis do not necessarily pose an obstacle to relating the two sources of data.

This brings us to the vast differences in temporal resolution. Since existing fMRI data cannot be made faster, typical solutions to the mismatch in temporal resolution have involved lowering the resolution of the EEG signal, by sampling signals over much longer timescales, and applying mathematical or statistical procedures (e.g., temporal averaging) to generate a relevant structure such as a local maximum in the 3D current distribution; this can then be compared to equivalent structures from fMRI. Vitacco et al. [36] applied this method to relate EEG and fMRI in a word classification task, but while they were able to obtain agreement between local maxima for group mean data, there was much poorer correspondence for individual subjects. One reason for this problem may be that, in averaging or otherwise manipulating EEG signals, one may be generating artifacts rather than discovering real features of the data. This is not to say that such attempts at data fusion are not promising, only that there is room for the introduction and evaluation of alternate approaches.

We have already outlined our approach to mining large numbers of fMRI studies and representing the results in graph format. This is relevant to the current issue because Chaovalitwongse et al. [13] recently developed a way to represent EEG data that also emphasized cooperative activity and also involved a graph-based representation scheme. In the scheme developed by Chaovalitwongse et al., cooperation between brain areas is measured in terms of the co-variance between EEG electrodes. Although the discovery of temporal correlation in large data sets is far from a trivial problem. Chaovalitwongse et al. [14, 12] have developed different methods to make such data mining tractable.

In discussions with Prof. Chaovalitwongse, we quickly realized that combining our two approaches could help address the issue of relating fMRI and EEG, because in approaches that focus on the cooperation of brain areas the small-scale temporal features of the EEG signal are de-emphasized, and the graph-based representational formats are entirely compatible; given the same underlying spatial segmentation of the cortex, the two cooperation graphs can be directly overlaid.

Of course, while it is clear that co-activation and co-variation graphs can be easily overlaid, what is unknown is whether there is any systematic relation between EEG co-variance and fMRI co-activation. We are currently putting together a research project to help answer this question (insofar as each graph is providing genuine information about which brain areas cooperate in supporting various cognitive tasks,

it certainly seems plausible that there would be some such relation). While it is by no means certain that any such relation will be found, the potential payoff is enormous. Among other things, it suggests it would be possible to mine the vast trove of fMRI data to provide baseline expectations for normal brain function in terms of the temporal correlation between brain areas. Since this can be observed cheaply, noninvasively, and in real time with EEG, it would be of great use in clinical settings for detecting deviations from normal function, such as might be observed prior to the onset of an epileptic seizure [12].

2.6 Conclusion

This chapter introduced a very simple analytical method for mining large numbers of brain imaging experiments to discover functional cooperation between brain regions. We reported some preliminary results of its application, illustrated some of the many future projects in which we expect the technique will be of considerable use, and described a research resource for investigating functional cooperation in the cortex that will be made publicly available through the lab web site. We hope and expect the availability of this resource will help spur new and innovative discoveries in the cognitive and computational neurosciences.

References

1. Abello, J., Pardalos, P.M., Resende, M.G.C. On maximum clique problems in very large graphs in external memory algorithms. In: Abello, J., Vitter, J. (eds.) AMS-DIMACS Series on Discrete Mathematics and Theoretical Computer Science, Vol. 50 (1999)
2. Alba, R.D. A graph-theoretic definition of a sociometric clique. J Math Sociol **3**, 113–126 (1973)
3. Anderson, J.R., Qin, Y., Jung, K.J., Carter, C.S. Information processing modules and their relative domain specificity. Cogn Psychol **54**, 185–217 (2007)
4. Anderson, M.L. Evolution of cognitive function via redeployment of brain areas. Neuroscientist **13**(1), 13–21 (2007)
5. Anderson, M.L. Massive redeployment, exaptation, and the functional integration of cognitive operations. Synthese **159**(3), 329–345 (2007)
6. Anderson, M.L. The massive redeployment hypothesis and the functional topography of the brain. Philos Psychol **21**(2), 143–174 (2007)
7. Attwell, D., Iadecola, C. The neural basis of functional brain imaging signals. Trends Neurosci **25**(12), 621–25 (2002)
8. Bock, R.D., Husain, S.Z. An adaptation of Holzinger's b-coefficients for the analysis of sociometric data. Sociometry **13**, 146–53 (1950)
9. Bonacich, P. Factoring and weighting approaches to status scores and clique identification. J Math Sociol **2**, 113–20 (1972)
10. Brannen, J.H., Badie, B., Moritz, C.H., Quigley, M., Meyerand, M.E., Haughton, V.M. Reliability of functional MR imaging with word-generation tasks for mapping Broca's area. Am J Neuroradiol **22**, 1711–1718 (2001)
11. Cabeza, R., Nyberg, L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. J Cogn Neurosci **12**, 1–47 (2000)

12. Chaovalitwongse, W., Fan, Y.J., Sachdeo, R. On the k-nearest dynamic time warping neighbor for abnormal brain activity classification. IEEE Trans Syst Man Cybern A Syst Hum **37**(6), 1005–1016 (2007). To appear

13. Chaovalitwongse, W., Iasemidis, L.D., Pardalos, P.M., Carney, P.R., Shiau, D.S., Sackellares, J.C. Performance of a seizure warning algorithm based on the dynamics of intracranial EEG. Epilepsy Res **64**, 93–133 (2005)

14. Chaovalitwongse, W., Pardalos, P.M., Prokopyev, O.A. Electroencephalogram (EEG) time series classification: Applications in epilepsy. Ann Operations Res **148**, 227–250 (2006)

15. Diestel, R. Graph Theory, 3rd edn. Springer-Verlag, Heidelberg (2005)

16. Gross, J.L., Yellen, J. Graph Theory and its Applications, 2nd edn. Discrete Mathematics and Its Applications. Chapman & Hall/CRC, London (2005)

17. Han X., Xu, C., Braga-Neto, U., Prince, J.L. Topology correction in brain cortex segmentation using a multiscale, graph-based approach. IEEE Trans Med Imaging **21**, 109–121 (2002)

18. Hayes, B. Graph theory in practice: Part I. Am Sci **88**(1), 9–13 (2000)

19. Hayes, B. Graph theory in practice: Part II. Am Sci **88**(2), 104–109 (2000)

20. Horwitz, B., Poeppel, D. How can EEG/MEG and fMRI/PET data be combined? Hum. Brain Mapp. **17**, 1–3 (2002)

21. Laird, A.R., Fox, M., Prince, C.J., Glahn, D.C., Uecker, A.M., Lancaster, J.L., Turkeltaub, P.E., Kochunov, P., Fox, P.T. Ale meta-analysis: Controlling the false discovery rate and performing statistical contrasts. Hum Brain Mapp **25**, 155–164 (2005)

22. Laird, A.R., Lancaster, J.L., Fox, P.T. Brainmap: The social evolution of a functional neuroimaging database. Neuroinformatics **3**, 65–78 (2005)

23. Lancaster, J., Laird, A., Fox, M., Glahn, D., Fox, P. Automated analysis of meta-analysis networks. Hum Brain Mapp **25**, 174–184 (2005)

24. Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T. Automated talairach atlas labels for functional brain mapping. Hum Brain Mapp **10**, 120–131 (2000)

25. Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. Nature **412**, 150–157 (2001)

26. Nunez, P.L., Silberstein, R.B. On the relationship of synaptic activity to macroscopic measurements: Does co-registration of EEG with fMRI make sense? Brain Topogr **13**, 79–96 (2000)

27. Sporns, O., Ktter, R. Motifs in brain networks. PLoS Biol **2**, e369 (2004)

28. Özcan, M., Baumgärtner, U., Vucurevic G. Stoeter, P., Treede, R.D. Spatial resolution of fMRI in the human parasylvian cortex: Comparison of somatosensory and auditory activation. NeuroImage **25**(3), 877–887 (2005)

29. Grave de Peralta Menendez, R., Gonzales Andino, S., Morand, S., Michel, C., Landis, T. Imaging the electrical activity of the brain. Electra Hum Brain Mapp **9**, 1–12 (2000)

30. Grave de Peralta Menendez, R., Murray, M.M., Michel, C., Martuzzi, R., Gonzales Andino, S.L. Electrical neuroimaging based on biophysical constraints. NeuroImage **21**, 527–539 (2004)

31. Sporns, O., Tononi, G., Edelman, G.M. Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. Cereb Cortex **10**, 127–141 (2000)

32. Suharitdamrong, W., Chaovalitwongse, A., Pardalos, P.M. Graph theory-based data mining techniques to study similarity of epileptic brain network. In: Proceedings of DIMACS Workshop on Data Mining, Systems Analysis, and Optimization in Neuroscience (2006)

33. Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A. Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. Neuroimage **16**, 765–780 (2002)

34. Ugurbil, K., Toth, L., Kim, D.S. How accurate is magnetic resonance imaging of brain function? Trends Neurosci. **26**(2), 108–114 (2003)

35. Viswanathan, A., Freeman, R.D. Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. Nat Neurosci **10**(10), 1308–1312 (2007)

36. Vitacco, D., Brandeis, D., Pasual-Marqui, R., Martin, E. Correspondence of event-related potential tomography and functional magnetic resonance imaging during language processing. Hum Brain Mapp **17**, 4–12 (2002)

Chapter 3
Methodological Framework for EEG Feature Selection Based on Spectral and Temporal Profiles

Vangelis Sakkalis and Michalis Zervakis

Abstract Among the various frameworks in which EEG signal analysis has been traditionally formulated, the most widely studied is employing power spectrum measures as functions of certain brain pathologies or increased cerebral engagement. Such measures may form signal features capable of characterizing and differentiating the underlying neural activity. The objective of this chapter is to validate the use of wavelets in extracting such features in the time–scale domain and evaluate them in a simulated environment assuming two tasks (control and target) that resemble widely used scenarios of assessing and quantifying complex cognitive functions or pathologies. The motivation for this work stems from the ability of time–frequency features to encapsulate significant power alteration of EEG in time, thus characterizing the brain response in terms of both spectral and temporal activation. In the presented algorithmic scenario, brain areas' electrodes of significant activation during the target task are extracted using time-averaged wavelet power spectrum estimation. Then, a refinement step makes use of statistical significance-based criteria for comparing wavelet power spectra between the target task and the control condition. The results indicate the ability of the proposed methodological framework to correctly identify and select the most prominent channels in terms of "activity encapsulation," which are thought to be the most significant ones.

Vangelis Sakkalis

Department of Electronic and Computer Engineering, Technical University of Crete, Chania 73100, Greece, e-mail: `sakkalis@ics.forth.gr`

Michalis Zervakis

Department of Electronic and Computer Engineering, Technical University of Crete, Chania 73100, Greece, e-mail: `michalis@display.tuc.gr`

3.1 Introduction

Electroencephalographic (EEG) measures have been successfully used in the past as indices of cerebral engagement in cognitive tasks or in the identification of certain brain pathologies. Higher brain functions typically require the integrated, coordinated activity of multiple specialized neural systems that generate EEG signals at various brain regions. Linear [7,18] and nonlinear signal analysis methods have been applied in order to derive information regarding patterns of local and coordinated activity during performance of specific tasks [11] or in various pathologies [2, 13]. The inherent complexity and the dynamic nature of brain function make the evaluation using EEG a rigorous job. Nevertheless, EEG signal analysis provides the advantage of high time resolution and thus it can deduce information related to both local and widespread neuronal activations in short-time periods, as well as their time evolution.

Traditional spectral analysis techniques with Fourier transform (FT) and more specifically the windowed power spectral density function, known as the periodogram [16], form the most commonly used analytical tool for spectral representation and evaluation of activity on different EEG frequency bands [7, 15] – namely *delta (δ), theta (θ), alpha (α), beta (β), and gamma (γ)*. However, this approach considers the EEG signal as a stationary process, which assumption is not satisfied in practice, thus restricting the actual confidence on results. A more promising methodology is based on the time-varying spectral analysis that takes into account the nonstationary dynamics of the neuronal processes [1]. The short-time Fourier (STFT) and the wavelet transforms are the most prevalent analysis frameworks of this class. The first approach uses a sliding time window, whereas the second one forms the projection of the signal onto several oscillatory kernel-based wavelets matching different frequency bands. Currently, such time-varying methods have been widely applied in event-related potential (ERP) data, where distinct waveforms are associated with an event related to some brain function [3]. Under certain assumptions, both time–frequency transforms are in fact mathematically equivalent, since they both use windows that under certain conditions can provide the same results [4]. The reason why these approaches are often regarded as different lies in the way they are used and implemented. Wavelet transform (WT) is typically applied with the relative bandwidth $(\Delta f/f)$ held constant, whereas the Fourier approach preserves the absolute bandwidth (Δf) constant. In other words, STFT uses an unchanged window length, which leads to the dilemma of resolution; a narrow window leads to poor frequency resolution, whereas a wide window leads to poor time resolution. Consequently, according to the Heisenberg uncertainty principle one cannot accurately discriminate frequencies in small time intervals. However, the WT can overcome the resolution problem by providing multiresolution analysis. The signal may be analyzed at different frequencies with different resolutions achieving good time resolution but poor frequency resolution at high frequencies and good frequency resolution but poor time resolution at low frequencies. Such a setting is suitable for short duration of higher frequency and longer duration of lower frequency components of the EEG bands. For the purposes of this study the wavelet approach is used.

In this work we attempt to retrieve additional information (as compared to traditional spectral analysis methods) by making use of the time profile of the EEG signal during the target task under study. The motivation for this work stems from the fact that the WT method is able to extract not only the spectral activations but also the time segments at which they occur. It constitutes the cornerstone of our feature extraction scheme and is used for analyzing task-related or control EEG signals by effectively capturing the power spectrum (PS) of each frequency band and channel. In particular, it encodes the activation differences between the mental states of interest. The subsequent feature selection steps apply test statistics on the extracted "time-averaged" PS features. In addition, our approach introduces an extra refinement step that makes further use of the time profile provided by the WT as to derive and encode the temporally activated brain regions and bands. The proposed EEG feature extraction and selection method may also be applied to other similar nonstationary biological signal analysis problems.

3.2 Methods

3.2.1 Methodology Overview

Two different cognitive tasks are assumed for simplicity: the control and the target ones that involve a modulated rather than random activity. The latter task encapsulates the crucial information for extracting both the frequency bands and the location of brain activity, in terms of channel references or groups of channels (related to specific brain areas) as an index of cerebral engagement in certain mental tasks. The testing hypothesis suggests that the target task induces activity on certain brain lobes, reflected on the associated electrodes in a way significantly different compared to a control task. The WT constitutes the cornerstone of feature extraction and is used in analyzing task-related or control EEG signals by effectively capturing the power spectrum (PS) of each band and channel, particularly encoding the activation differences between the tasks. From the technical point of view, statistics is used to extract and select salient features, testing for significance in both the time and scale domains of the signal. The feature selection steps apply test statistics on the extracted "time-averaged" PS features, but in addition our approach introduces an extra refinement step that makes further use of the time profile of the WT, as to derive and encode the temporally activated brain regions and bands. Test statistics form appropriate means for the design of feature selection criteria strictly based on statistical significance; they are simple to implement and often perform better than other heuristic selection methods. To that respect, we base our selection on statistical tests that rely on statistical properties of the feature data under consideration. Hopefully the identified channels and lobes may elucidate any neurophysiological pathways involved in brain function.

A generic overview of the proposed methodology emphasizing the various statistical approaches is illustrated in Fig. 3.1. Different statistical decisions are possible according to the profile of the data under examination. The first choice is based on whether the data are normally distributed, whereas the second is based on the number of different groups under examination – i.e., whether two or more classes (tasks) are being tested. A detailed view of feature selection and refinement blocks matching our data characteristics is presented in Fig. 3.2. The steps involved, as well as their implementation issues, are analyzed in the following sections.
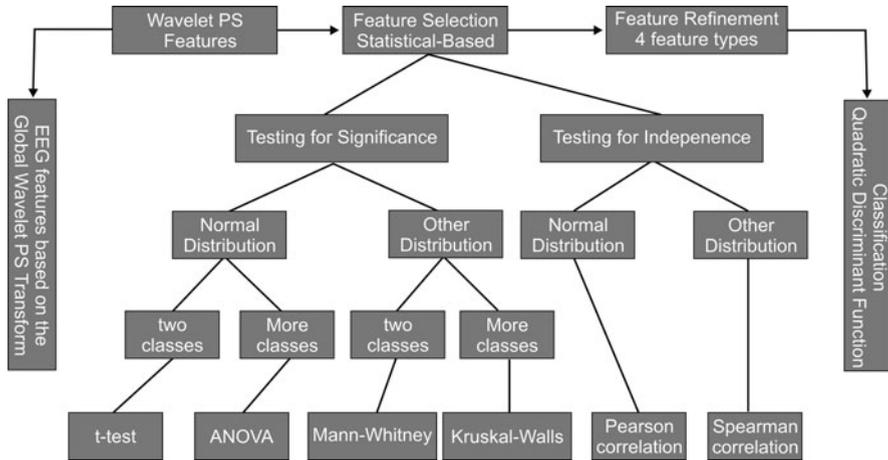
Fig. 3.1: The proposed methodology uses significance-based statistics to reduce the dimensionality of the problem and select the most salient and descriptive feature vectors. Different statistical decisions are possible according to the profile of the data under examination. If one is interested in discriminating two or more classes of normally distributed data, t-test or analysis of variance (ANOVA) tests are appropriate candidates, respectively. If the data is nonnormally distributed, Mann–Whitney and Kruskal–Walls tests are the alternatives.

3.2.2 Feature Extraction (Step 1)

Over the past decade the WT has developed into an important tool for analysis of time series that contain nonstationary power at many different frequencies (such as the EEG signal), as well as a powerful feature extraction method [9]. There are several types of wavelet transforms, namely the discrete (DWT) and the continuous (CWT), which involve the use of orthogonal bases or even nonorthogonal wavelet functions, respectively [8]. CWT is preferred in this approach, so that the time and scale parameters can be considered as continuous variables. In the WT, the notion
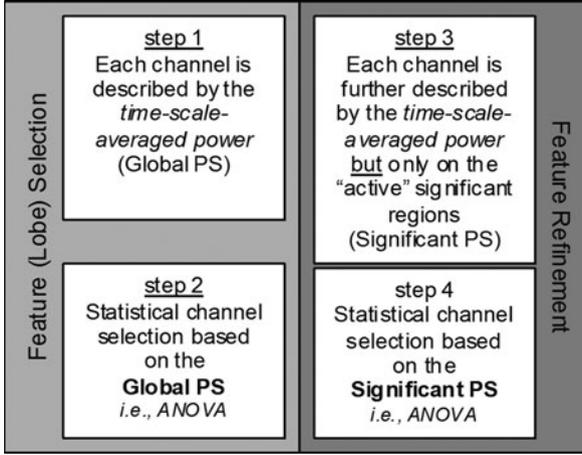
Fig. 3.2: The diagram of the proposed algorithmic transitions, heading toward derivation of significant activity channels and bands.

of scale s is introduced as an alternative to frequency, leading to the so-called time–scale representation domain.

The CWT of a discrete sequence x_n with time spacing δt and N data points ($n = 0, N-1$) is defined as the convolution of x_n with consecutive scaled and translated versions of the wavelet function $\psi_0(\eta)$:

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi^* \left[(n'-n)\delta t/s\right], \tag{3.1}$$

$$\psi_0(\eta) = \pi^{1/4} e^{i\omega_0 \eta} e^{-\eta^2/2}, \tag{3.2}$$

where η and $\omega_0 = 6$ indicate nondimensional "time" and frequency parameters, respectively and $\psi^*(\cdot)$ denotes the complex conjugate operation. In our application, $\psi_0(\eta)$ describes the most commonly used wavelet type for spectral analyses, i.e., the *normalized complex Morlet wavelet* given in (3.2). The wavelet function ψ_0 is a normalized version of ψ that has unit energy at each scale, so that each scale is directly comparable to each other. The normalization is given as

$$\psi \left[(n'-n)\delta t/s\right] = (\delta t/s)^{1/2} \psi_0 \left[(n'-n)\delta t/s\right]. \tag{3.3}$$

In principle, a complex wavelet function is better suited for capturing oscillatory behavior than a real one, because it captures both the amplitude and the phase of EEG signal. The scale set is given by

$$s_j = s_0 2^{j\delta j}, \ j = 0, \cdots, J, \tag{3.4}$$

where $s_0 = 2\delta t$ is the smallest scale chosen and δj specifies the width of the wavelet function. In our case $\delta j = 0.25$, implying that there is a scale resolution of four sub-octaves per octave [5]. The larger scale is determined by the value of J specified in (3.5), which in our case is $J = 29$:

$$J = \delta j^{-1} \log_2(N\delta t / s_0). \tag{3.5}$$

Finally, the *power spectrum* of the WT is defined by the square of coefficients in (3.1) of the wavelet series as $\|W_n(s)\|^2$. By adopting the above settings a smooth wavelet power diagram is constructed as in Fig. 3.3b for the signal in Fig. 3.3a.
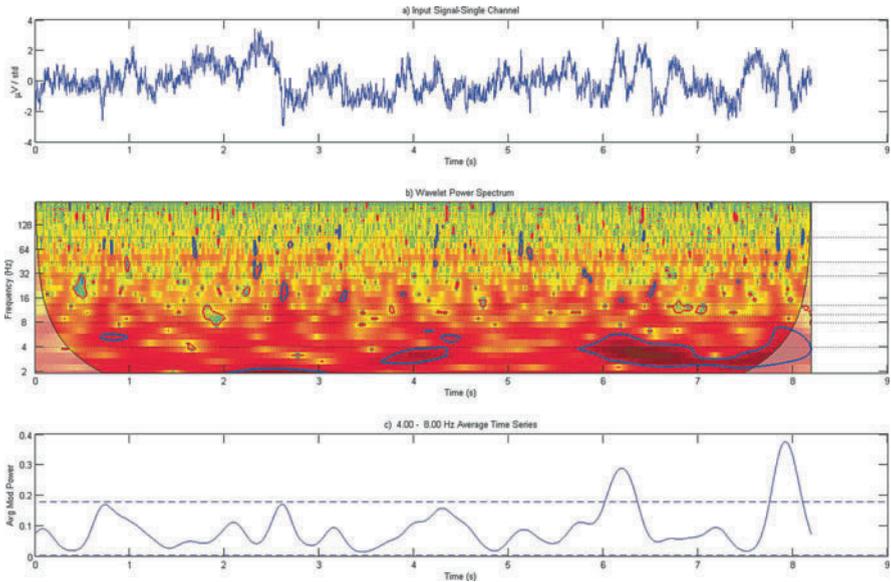
Fig. 3.3: (**a**) A typical normalized EEG signal acquired from a single electrode. (**b**) The wavelet power spectrum presented as a color-coded picture. Mapped scales to frequencies are calibrated on the y-axis, with the *horizontal dashed lines* indicating the different frequency bands. The significant regions over the time–scale transform are indicated by *closed contours*. Power increase and decrease is bounded by *blue* and *red* contours, respectively. The outer elliptical region at the edges of this second graph indicates the cone of influence in which errors (edge effects) may be apparent due to the transformation of a finite-length series EEG signal. (**c**) The scalogram of a selected averaged band (Theta 4–8 Hz) reflecting characteristic EEG activity while the participant is performing a complex mathematical calculation [14]. The significance levels are indicated by the *horizontal dashed lines*. PS values greater above the *upper dashed line* indicate significant increase, whereas PS values below the *lower dashed line* indicate significant decrease over the expected control power levels.

As noted before, there exists a concrete relationship between each scale and an equivalent set of Fourier frequencies, often known as *pseudofrequencies* [10]. For the Morlet wavelet used this relationship is $f = \frac{\omega_0 + \sqrt{2+\omega_0^2}}{4\pi s}$, which in our case ($\omega_0 = 6$); this gives a value of $f = 1/(1.03s)$. In this study the power spectra is classified in six sequential frequency bands that are coarsely mapped to the scales tabulated in Table 3.1.

Table 3.1: Frequency bands – scale set mapping

Band	Frequency	Scale
Theta (θ)	4–8	21, 22, 23, 24
Alpha1 (α_1)	8–10	20
Alpha2 (α_2)	10–13	18, 19
Beta (β)	13–30	14, 15, 16, 17
Gamma1 (γ_1)	30–45	11, 12, 13
Gamma2 (γ_2)	45–90	7, 8, 9, 10

The first stage of our feature extraction method is based on capturing the *time-averaged power spectrum* $\overline{W_n}^2$ for each electrode and scale, which is computed by averaging the power spectrum $\|\overline{W_n}\|^2$ over time:

$$\overline{W_n}^2(s) = (1/N) \sum_{n=0}^{N-1} \|W_n(s)\|^2. \tag{3.6}$$

Further averaging in scale is performed, in order to map a single feature per frequency band of interest. Thus, the *scale-averaged power spectrum* $\overline{W_n}^2$ is defined as the weighted sum of the wavelet power spectrum $\|W_n(s)\|$ over scales s_{j_1} to s_{j_2} within each frequency band, with scale correspondences defined in Table 3.1. Based on these definitions, the average power over time and frequency band is obtained as

$$\overline{W_{s,n}} = (\delta j/\delta t/C_\delta) \sum_{j=j_1}^{j_2} \left(\|W_n(s_j)\|^2/s_j \right), \tag{3.7}$$

where C_δ is a constant scale-independent factor used for the exact reconstruction of a $\delta(\cdot)$ function from its wavelet transform (for the Morlet wavelet it equals to 0.776) [17]. Once the average PS for each of the studied EEG bands is calculated for each EEG channel and task, we have a high number feature vectors (bands x channels) per task (class), representing each participant (subject), which is actually the time–scale-averaged PS (Global PS – Fig. 3.2 – Step 1) over the band of interest.

3.2.3 Feature Selection (Step 2)

In data mining and classification applications, feature selection and reduction of dimensionality in the feature space play a crucial role in the effective design by regularizing and restricting the solution space. It is of practical concern that a large number of features may actually degrade the performance of the classifier if the number of training samples is limited in comparison to the number of features [12]. This study proposes a statistical method for mining the most significant channels, resembling the way many clinical neurophysiological studies evaluate the brain activation patterns.

Hence, the second step (Fig. 3.2 – Step 2) of our design involves the statistical test selection of features, which depends upon the feature-vector properties and the experimental design. The distribution of features plays the most important role, since it is the one to judge which statistical test is the most appropriate (Fig. 3.1). Normality of the feature set may be tested using the D'Agostino–Pearson test [19]. Once normality is met and supposing that two classes are being discriminated, t-test or analysis of variance (ANOVA) is the ideal test to use in our application. The ANOVA test is superior for complex analyses for two reasons, the first being its ability to combine complex data into one statistical procedure (more than two groups may be compared). The second benefit over a simple t-test is the ANOVA's ability to determine interaction effects. One of the common assumptions underlying ANOVA is that the groups being compared are independent of each other. In the case of a related studies design (the same subjects perform each task), either matched pairs or repeated measures are more appropriate, e.g., a repeated measures ANOVA [19] with common measures factors being the two tasks and the number of channels, testing for significance at the level of 0.05. For those bands where the significance criterion is fulfilled, follow-up post hoc tests for each channel are performed to accentuate the best candidate channels to preserve as features, which resemble the most significant brain areas in terms of activity.

3.2.4 Feature Refinement (Steps 3 and 4)

The aforementioned steps derive a significant channels' subset, based only on task differentiation confidence intervals using Global PS measures. To further refine the features and optimize the whole process, we propose to isolate only those time segments of the EEG signal where notable activity differences occur from the control to the arithmetic task. The aim is to further map the EEG signal into a feature vector that best characterizes the EEG pattern of activity for the target task in terms of significant temporal and spectral content. As we are interested in ongoing EEG activity within various tasks, the temporal activity of EEG events is of interest. Notice that we focus on significant (bursty and/or sequential) activations and not on the evolution of brain operation during the task. Thus, we are mostly focused on the time-localized EEG activity itself, without particular interest to the temporal

relation of these events. We may describe the next step as an attempt to crop up the most significantly different regions from control to target activity out of the bulk initial signal (may be either significant power increase or decrease while performing the requested task compared to the control condition). In fact, this study proposes a way to derive the so-called *significant PS activity* on significantly activated EEG time segments, by testing for significance in the wavelet–time domain the "active" task over the control task (Significant PS – Fig. 3.2 – Step 3). The control task spectra de-fine the *mean time-averaged wavelet power spectrum* over all subjects performing the control task, as

$$\overline{W}(s) = (1/P) \sum_{p=1}^{P} \|W_n^p(s)\|^2,\tag{3.8}$$

where p is the subject index and $W_n^p(s)$ is computed as in (3.1) for each subject. P is the total number of participants. It should be noticed that all EEG signals are normalized to zero mean and identity variance. Further rescaling and comparisons may be performed using each subject's actual signal variance in order to include subject-specific information. Significant power increase on the "active" task is calculated using the 95% confidence level at each scale by multiplying the control task spectrum in Equation (3.8) by the 95th percentile value for a chi-squared distributed variable χ^2 with two degrees of freedom χ_2^2. This is justified because the wavelet power spectrum is derived from the Morlet wavelet in a complex product with the signal, so that both the squares of the real and the imaginary parts of the result are being χ^2 distributed with one degree of freedom each [17, 6]. In a similar manner, significant power decrease is measured using the lower power limit of 5% confidence level at each scale, by multiplying the control task spectrum in Equation (3.8) by the 5th percentile value for the chi-squared distributed variable χ_2^2 . Figure 3.3 depicts one subject's initial normalized EEG signal (Fig. 3.3a) together with its WT (Fig. 3.3b). The significant regions over the time–scale-transformed domain that differentiate the two tasks are indicated by the closed contours; red for significantly increased and blue for decreased activity. Figure 3.3c illustrates another view of the scalogram focusing on a selected averaged band, i.e., (Theta 4–8 Hz). The significance levels in this case are indicated by horizontal dashed lines.

Having derived this significant information, we are now able to form the so-called *significant power spectral* (significant PS) features, which are obtained from the signal energy over those time- and band-localized regions where apparent significant differentiation is indicated (contours in Fig. 3.3b). For the computation of these features, Equation (3.6) is adapted as

$$\overline{W_s t}^2 = (1/m) \sum_{m=m_i}^{m_{i+1}} \|W_m(s)\|^2, i = 1, \cdots, I,\tag{3.9}$$

where m is the total number of time points delimited between the boundaries m_i and m_{i+1} of all significant regions I denoted by each contour in Fig. 3.3b and i is the index of each significant region. Finally, the last step (Fig. 3.2 – Step 4) is actually a repetition of the statistical testing in the second step on the new feature set. ANOVA

or any other better suited statistical method (as described previously) may be used to further sort out and select the best candidate features (significant energy per time, band and electrode), in terms of their task discriminating power.

3.3 Results

The proposed methodology is tested on simulated data, where there exist well-defined spatiotemporal differences in frequency content between the target and the control tasks, as discussed in the following section. In addition, the performance of the proposed approach, as well as its results on actual experimental dataset, is discussed in [14].

3.3.1 Simulation Test

Two different tasks are simulated by two different groups of signals. The first group (control task) consists of 10 simulated spatiotemporal signals, each one comprised of five channels. The idea is to reflect 10 participants virtually registered with a 5-channel-EEG system each. All the channels of the control task are randomly generated quasi-white noise signals, approximately 9-s-long (500 Hz sampling rate – 4,608 samples). The second group (target task) comprises of three channels (channels 1, 3, 4) reflecting white noise and two channels (channels 2, 5) encoding frequency-modulated signals mixed again with quasi-white noise. Channel 2 consists of a time-varying theta EEG signal occurring at a fixed latency, linearly modulated (5–7 Hz) and varying in length randomly between 512 and 1,024 samples among subjects, and a gamma EEG signal, linearly modulated (30–90 Hz) and varying in length randomly between 1,024 and 2,048 samples among subjects, all mixed with quasi-white noise. In a similar manner, channel 5 consists of an alpha band linearly modulated signal (9–12 Hz) varying in length randomly (768–1,536 samples) and a gamma linearly modulated signal (30–90 Hz) varying in length randomly between 512 and 1,024 samples, mixed with quasi-white noise. Quasi-white noise covers the interval between the modulated signals. Such a generated signal (channel 2) together with the wavelet time–frequency representation is depicted in Fig. 3.4. Theta and gamma bands are apparent at different latencies. The tabulated channels in Table 3.2 are the significant ones extracted with the proposed approach from the six (most widely studied) frequency bands (delta, theta, alpha, beta, gamma1, and gamma2). The channels listed in the first column are the selected ones after the first statistical test (Step 2), whereas the channels listed in the second column are the refined ones after the second statistical selection (Step 4). Although the first stage can identify both channels (2 and 5) with the pre-specified frequency content, it is not able to discriminate correctly the activated frequency bands because of leakage effects between bands, as illustrated in Fig. 3.4. In contrast, the second stage focuses
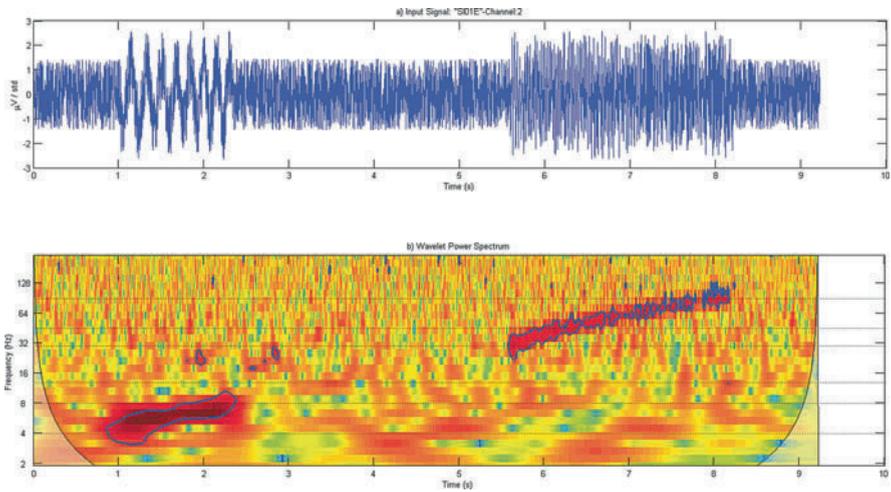
Fig. 3.4: (**a**) The simulated channel 2 consists of a time-varying (among different participants) theta linearly modulated signal (length   2 s) occurring at a fixed latency and a gamma linearly modulated signal (length   3 s) mixed with quasi-white noise. Quasi-white noise also covers the interval between the modulated signals. (**b**) The wavelet PS time–frequency representation picture. The significant regions over the time–frequency transform are indicated by the contours. The significant signal segments (contours) are successfully discriminated from the white noise background.

on the significant regions and is able to detect and correctly account for the energy content of the selected regions.

3.4 Discussion

Using the wavelet transform method on EEG signals, cortical activation evaluation is normally performed by means of comparing a target task (while participant is engaged with a difficult cognitive task or reflects certain pathology) and compares it with a rest condition. This method, in contrast to traditional spectral ones, can estimate changes between EEG signals without being bounded to the stationarity assumption and can provide information for the entire time evolution of the signal.

The simulation test and the results presented justify the suggestion that relevant characteristics are temporally localized in the most significant regions (contours in the WT scalogram), rather than in the entire segment length of the EEGs. The *Global* PS only partially encapsulates the significant information, since there is significant frequency leakage between the bands due to the transient response of the time–frequency filter in different frequencies. Using such features, both channels 2 and 5 in the simulated case induce activity in almost every band. However, the proposed

Table 3.2: Statistical feature – channel selection results

Band	Channel (step 2)	Channel (step 4)	Target
Delta (δ)	2	–	–
Theta (θ)	2, 5	2	2
Alpha (α)	2, 5	5	5
Beta (β)	–	–	–
Gamma1 ($gamma_1$)	2, 5	2, 5	2, 5
Gamma2 ($gamma_2$)	2, 5	2, 5	2, 5

methodology with its second statistical feature selection scheme can efficiently isolate the channels and the correct band activations. Traditional FT spectral analysis methods pose intrinsic limitations on encapsulating the time variation of the signal. Beyond traditional spectral analysis, the WT enables the consideration of time specific significant regions as in Step 3 of the proposed methodology. WT is proved to be a useful measure to detect time-varying spectral power and performs better than traditional time–frequency methods in identifying activity, especially on a shorter temporal scale in high frequencies, which could indicate neuronal synchronous activation in some cortical regions. This is an advantage to previous methodologies, since high-frequency bands are weak and difficult to evaluate using spectral methods.

A qualitative reasoning arising from the application of this methodology to actual data is discussed in [14], where the certain methodology is applied to a complex mathematical reasoning task. Finally, the presented method reveals additional signal characteristics, since it captures not only its average power but also the time-localized activation of the signal.

3.5 Conclusion

The proposed algorithmic approach emphasizes the idea of selecting EEG features based on their statistical significance and further supports the use of time–scale WT domain in order to select significant EEG segments capable of describing the most prominent task-related changes.

Results suggest that the proposed methodology is capable of identifying regions of increased activity during the specified target task. The entire process is automated in the sense that different feature types can be adaptively (according to the data profile) extracted and further refined in a way "transparent" to the user. Such processes may be transferred to a clinical environment if the methods prove to be valuable for the diagnosis of certain pathologies by comparing any routine EEG against a database of pathological ones.

Furthermore, the added value of this approach over other classical Fourier-based methods lies in its ability to further utilize time-domain characteristics of the WT in a way comparable to the evoked potential applications, without making any compromise in the statistical validity of the results.

References

1. Bianchi, A., Mainardi, L., Cerutti, S. Time-frequency analysis of biomedical signals. Trans Inst Measur. and Contr **22**, 321–336 (2000)
2. Breakspear, M., Terry, J., Friston, K., Harris, A., Williams, L., Brown, K., Brennan, J., Gordon, E. A disturbance of nonlinear interdependence in scalp EEG of subjects with first episode schizophrenia. NeuroImage **20**, 466–478 (2003)
3. Bressler, S. The Handbook of Brain Theory and Neural Networks. MIT Press, Cambridge MA 412-415, (2002)
4. Bruns, A. Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? J Neurosci Methods **137**(2), 321–332 (2004)
5. Burrus, S., Gopinoth, R., Haitao, G. Introduction to Wavelets and Wavelet Transforms. Prentice Hall Inc., Upper Saddle River, NJ (1998)
6. Chatfield, C. The Analysis of Time Series: An Introduction, 4th edn, p. 241. Chapman and Hall, London (1989)
7. Dumermuth, G., Molinari, L. Spectral analysis of the EEG. Some fundamentals revisited and some open problems. Neuropsychobiology **17**, 85–99 (1987)
8. Farge, M.: Wavelet transforms and their applications to turbulence. Annu Rev Fluid Mech **24**, 395–457 (1992)
9. Kalayci, T., Ozdamar, O. Wavelet preprocessing for automated neural network detection of EEG spikes. IEEE Eng Med Biol Mag **14**, 160–166 (1995)
10. Meyers, S., Kelly, B., O'Brien, J. An introduction to wavelet analysis in oceanography and meteorology: With application to the dispersion of Yanai waves. Mon Wea Rev **121**, 2858–2866 (1993)
11. Micheloyannis, S., Sakkalis, V., Vourkas, M., Stam, C., Simos, P. Neural networks involved in mathematical thinking: Evidence from linear and non-linear analysis of electroencephalographic activity. Neurosci Lett **373**, 212–217 (2005)
12. Raudys, S., Pikelis, V. On dimensionality, sample size, classification error and complexity of classification algorithm in pattern recognition. IEEE Trans Pattern Anal Machine Intell, **PAMI-2**(3), 242–252 (1980)
13. Sakkalis, V., Giurcaneanu, C., Xanthopoulos, P., Zervakis, M., Tsiaras, V., Yang, Y., Micheloyannis, S. Assessment of linear and nonlinear synchronization measures for analyzing EEG in a mild epileptic paradigm. Accepted to be published in, 2008. IEEE Trans Inf Tech (2008). DOI: 10.1109/TITB.2008.923141.
14. Sakkalis, V., Zervakis, M., Micheloyannis, S. Significant EEG features involved in mathematical reasoning: Evidence from wavelet analysis. Brain Topogr **19**(1–2), 53–60 (2006)
15. Simos, P., Papanikolaou, E., Sakkalis, E., Micheloyannis, S. Modulation of gamma-band spectral power by cognitive task complexity. Brain Topogr **14**, 191–196 (2002)
16. Stoica, P., Moses, R. Introduction to Spectral Analysis. Prentice-Hall, Upper Saddle River, NJ (1997). 24-26

17. Torrence, C., Compo, G. A practical guide to wavelet analysis. Bull Am Meteorol Soc **79**, 61–78 (1998)
18. Weiss, S., Mueller, H. The contribution of EEG coherence to the investigation of language. Brain Lang **85**, 325–343 (2003)
19. Zar, J.: Biostatistical Analysis, 4th edn. Prentice Hall, Upper Saddle River, NJ (1999)

Chapter 4

Blind Source Separation of Concurrent Disease-Related Patterns from EEG in Creutzfeldt–Jakob Disease for Assisting Early Diagnosis

Chih-I Hung, Po-Shan Wang, Bing-Wen Soong, Shin Teng, Jen-Chuen Hsieh, and Yu-Te Wu

Abstract Creutzfeldt–Jakob disease (CJD) is a rare, transmissible, and fatal prion disorder of brain. Typical electroencephalography (EEG) patterns, such as the periodic sharp wave complexes (PSWCs), do not clearly emerge until the middle stage of CJD. To reduce transmission risks and avoid unnecessary treatments, the recognition of the hidden PSWCs' forerunners from the contaminated EEG signals in the early stage is imperative. In this study, independent component analysis (ICA) was employed on the raw EEG signals recorded at the first admissions of five patients to segregate the co-occurrence of multiple disease-related features, which

Chih-I Hung
Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University, Taipei, Taiwan, ROC; Integrated Brain Research Laboratory, Department of Medical Research and Education, Veterans General Hospital, Taipei, Taiwan, ROC

Po-Shan Wang
The Neurological Institute, Taipei Municipal Gan-Dau Hospital; Department of Neurology, National Yang-Ming University School of Medicine, Taipei, Taiwan, ROC; The Neurological Institute, Taipei Veterans General Hospital, Taipei, Taiwan, ROC

Bing-Wen Soong
The Neurological Institute, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; Department of Neurology, National Yang-Ming University School of Medicine, Taipei, Taiwan, ROC

Shin Teng
Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University, Taipei, Taiwan, ROC; Integrated Brain Research Laboratory, Department of Medical Research and Education, Veterans General Hospital, Taipei, Taiwan, ROC

Jen-Chuen Hsieh
Integrated Brain Research Laboratory, Department of Medical Research and Education, Veterans General Hospital, Taipei, Taiwan, ROC; Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan, ROC, e-mail: jchsieh@vghtpe.gov.tw

Yu-Te Wu
Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University, Taipei, Taiwan, ROC; Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan, ROC, e-mail: yute.wu@msa.hinet.net

were difficult to be detected from the smeared EEG. Clear CJD-related waveforms, i.e., frontal intermittent rhythmical delta activity (FIRDA), fore PSWCs (triphasic waves), and periodic lateralized epileptiform discharges (PLEDs), have been successfully and simultaneously resolved from all patients. The ICA results elucidate the concurrent appearance of FIRDA and PLEDs or triphasic waves within the same EEG epoch, which has not been reported in the previous literature. Results show that ICA is an objective and effective means to extract the disease-related patterns for facilitating the early diagnosis of CJD.

4.1 Introduction

Creutzfeldt–Jakob disease (CJD) is a rare prion disorder of brain, with an approximated incidence of 0.5–1 case per million persons per year. The subtypes of human prion diseases can be familial, sporadic, or acquired, which are characterized by combination of clinical findings such as duration of disease, EEG changes, age at onset, and predominant neurological signs. Sporadic CJD (sCJD) is the most common subtype of CJD that usually develops in the 5th to 7th decade of life, with a mean age of onset of 62 years old (median 65). Survival times ranging from 1 to 58 months have been reported [24]. The clinical presentations such as memory loss, visual disturbances, involuntary movements, myoclonus, dementia, and coma can be observed subsequently from early to the terminal stage of the disease. Since CJD is a rapidly progressive, uniformly fatal, and transmissible spongiform encephalopathy, detection of the CJD symptom in the early stage is crucial to avoid the fatal transmission.

Electroencephalography (EEG), cerebral magnetic resonance imaging (MRI), and cerebrospinal fluid analysis (CSF analysis) are currently the most common diagnostic means of CJD. To evaluate these techniques, Collins et al. investigated the influence of several clinical parameters, such as prion protein gene codon 129 polymorphism, molecular subtype, age at disease onset, and illness duration, on the diagnostic sensitivity to EEG, cerebral MRI, and the CSF analysis. They reported that the CSF analysis had the highest sensitivity for early diagnosis since the 14-3-3 protein could be detected from the CSF after the disease had onset [14]. However, Geschwind et al. concluded that the sensitivity of CSF analysis in their study was only 53% and advised that it was risky to exclude the diagnosis of CJD in the case of negative CSF results [7]. Besides, the use of CSF 14-3-3 analysis, regardless of methods, is problematic since universally accepted standards are not available for performing such tests. Magnetic resonance brain imaging is another developing tool for detecting CJD. The study conducted by the Schröter et al. revealed T2-weighted MRI alternations in 109 (67%) out of 162 sporadic CJD patients [20], whereas the sensitivity of abnormal T2-weighted or diffusion-weighted MRI reported by Collins et al. was 43% [3]. Accordingly, efforts to develop more effective techniques for the aid to early diagnosis are of potentially great importance.

EEG is one of the major techniques used to diagnose CJD and has been included in the World Heath Organization diagnostic classification criteria [24]. In general, EEG patterns of sCJD exhibit longitudinal changes along with the course of the disease, ranging from frontal intermittent rhythmical delta activity (FIRDA), i.e., slow waves with 1–3 Hz, in the early stage to periodic lateralized epileptiform discharges (PLEDs) or prototypical periodic sharp wave complexes (PSWCs) in the middle and late stages [1, 3, 6, 25]. The temporal waveforms and the spatial dominances of FIRDA, PLEDs, and PSWCs are presented in the Fig. 4.1a, b, respectively. The morphology of PLEDs shows complexes which consist of a bi- or multiphasic spike or sharp wave and may include a slow wave [5]. The PSWCs mainly comprise simple sharp waves, i.e., monophasic, biphasic, and triphasic waves, with a typical duration of 200–600 ms, although complexes with mixed spikes, polyspikes, and slower waves may appear from times to times [5, 25, 24]. The peak-to-peak intervals of PSWCs are usually between 0.5 and 2 s. The major difference between the PLEDs and the PSWCs is their topographical dominances. The former is more hemispherically lateralized while the latter is more focal in the early stage and becomes diffusive after the middle stage. Since the PSWCs are not evident until the middle or late stage, detection of the PSWCs predecessors, such as FIRDA, PLEDs, and focal tirphasic waves, hidden in the smeared EEG signals is critical for the early diagnosis.
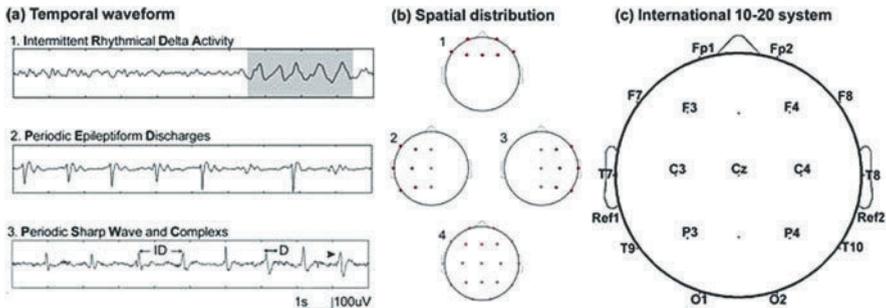
Fig. 4.1: (**a**) Temporal waveforms of FIRDA, PLEDs, and PSWCs. The PLEDs mainly consist of a bi- or multiphasic spike or sharp wave. The PSWCs mainly comprise simple sharp waves, i.e., monophasic, biphasic, and triphasic waves, with a typical duration of 200–600 ms and the peak-to-peak intervals are usually between 0.5 and 2 s. (**b**) Spatial dominances of FIRDA, PLEDs, and PSWCs. The FIRDA is usually observed in the frontal areas, the PLEDs are hemispherically lateralized, and the PSWCs are usually more focal in early stage and become diffusive after middle stage. (**c**) The whole scalp of each subject was covered with 19 EEG electrodes placed onto anatomical locations according to the international 10–20 system, where Fp, F, C, P, O, and T represent the abbreviations of frontal polar, frontal, central, parietal, occipital, and temporal, respectively. b1: frontal dominant, b2: left-side lateralization, b3: right-side lateralization, b4: generalized distribution, ID: interval duration, D: duration, 0.5 s<ID<2 s, D<600 ms, typical triphasic wave.
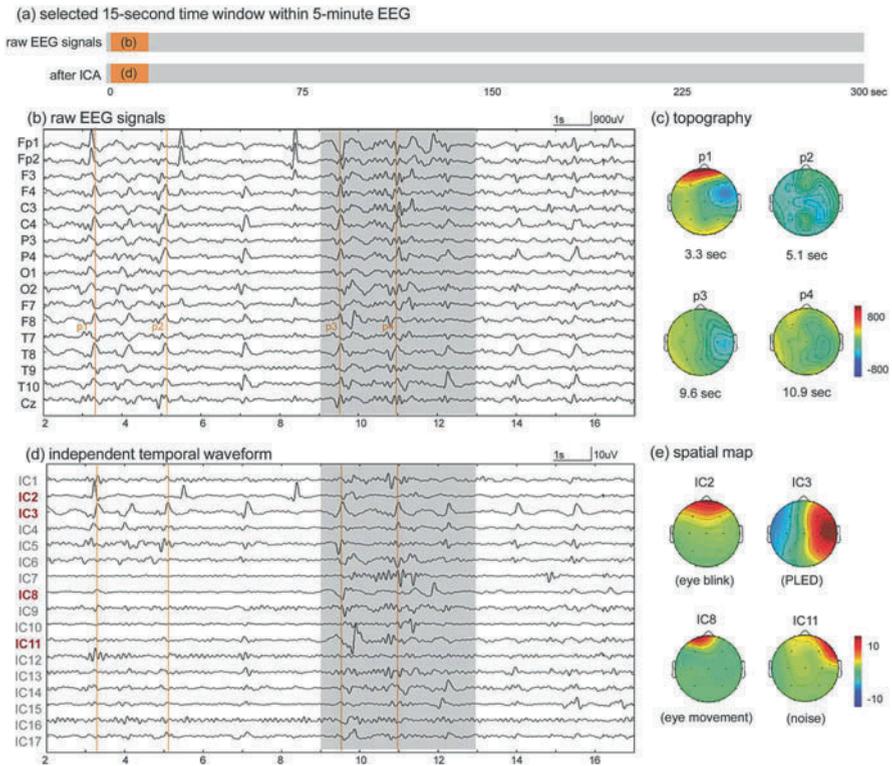
Fig. 4.2: The first selected EEG segment and ICA results from patient 1. Once W and S were resolved by ICA (Equation 4.5), rows of S representing the temporal waveforms of independent sources were displayed in (**d**), and each column of W-1 denoting the relative (spatial) weightings of each sources was depicted as a topography map in (**e**). (**a**) A 15-s time window (2–17 s) within 5-min data used to display results in (**b**) and (**d**). (**b**) The illustration of a 15-s segment where signals in the shaded areas were severely contaminated by large eye movements and environmental noises. (**c**) The topographical maps generated at four peak time points p1, p2, p3, and p4 (*vertical lines* in **b**) of four waves in IC3 at 3.3, 5.1, 9.6, and 10.9 s. (**d**) The 17 decomposed ICs show that diseased-related pattern was PLEDs (IC3) and the artifacts were eyeblinks (IC2), eye movements (IC8), and noise (IC11). (**e**) The corresponding spatial maps of IC2, IC3, IC8, and IC11.

EEG recordings are overlapping potentials contributed from individual neurons inside the brain as well as from the artifacts produced outside the brain [5]. Figures 4.2b, 4.3b, and 4.4b illustrate parts of typical segments of raw EEG signals recorded from the first admissions of patient 1 (the early stage of CJD). The shaded areas show that the brain activities are severely contaminated by significantly large eye-movement potentials and environmental noises, which makes the visual inspection of FIRDA, PLEDs, and tirphasic waves in the early stage of CJD a difficult task.

Fig. 4.3: The second selected EEG segment and ICA results from patient 1. (**a**) The 15-s time window (152–167 s) used to display results in (**b**) and (**d**). (**b**) The illustration of a 15-s segment where signals in the *shaded areas* were severely contaminated by large eye movements and environmental noises. (**c**) The topographical maps generated at four peak time points p1, p2, p3, and p4 (*vertical lines* in **b**) of four waves in IC4 at 154, 156.6, 159.9, and 162.9 s. (**d**) The 17 decomposed ICs show that diseased-related pattern was focal triphasic waves (IC4) and the artifacts were eyeblinks (IC2), eye movements (IC8), and noise (IC15). (**e**) The corresponding spatial maps of IC2, IC4, IC8, and IC15.

To recover the CJD-related patterns from EEG data, we employed the independent component analysis (ICA) [11, 23] in this study. ICA has been successfully applied to remove nonphysiological artifacts from EEG data [14, 15], to segregate Rolandic beta rhythm from magnetoencephalographic (MEG) measurements of the right index finger lifting [18], to extract the task-related features from the motor imagery EEG and the flash visual evoked EEG in the studies of the brain computer interface [10, 17], to analyze the interactions during temporal lobe seizures in stereotactic depth EEG [22], to separate generalized spike-and-wave discharges into the primary and secondary bilateral synchrony [13], and to segment spatiotemporal hemodynamics from perfusion magnetic resonance brain images [16].
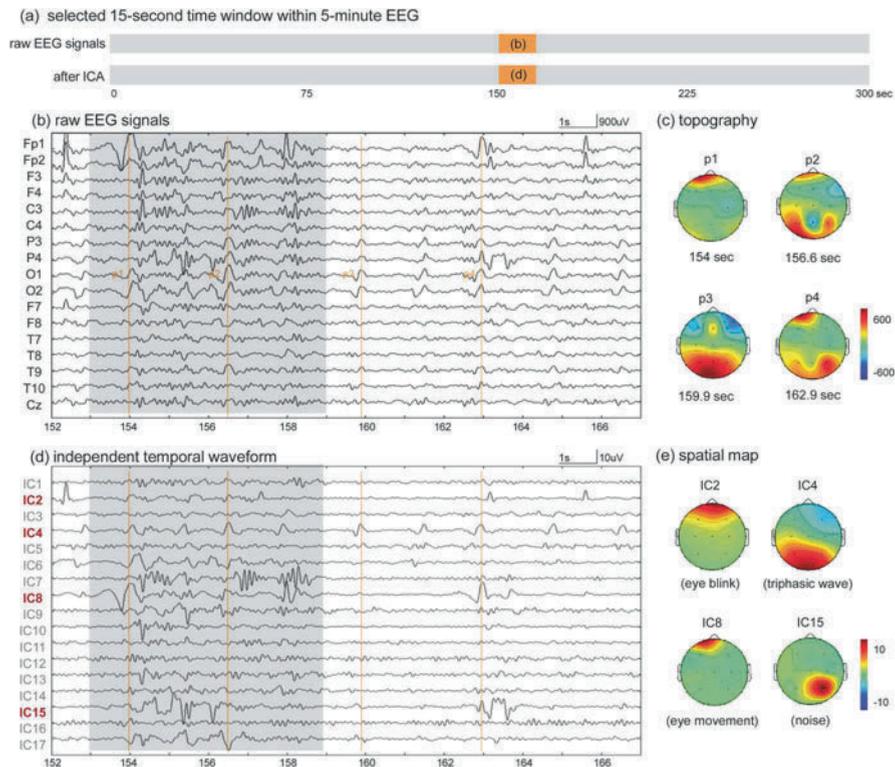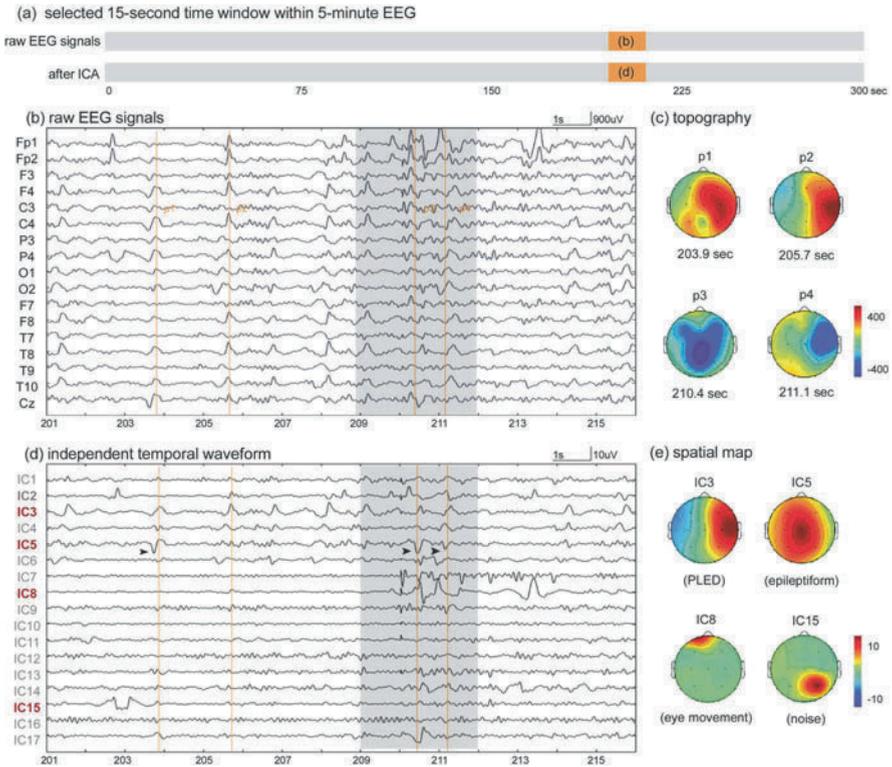
Fig. 4.4: The third selected EEG segment and ICA results from patient 1. (**a**) The 15-s time window (201–216 s) used to display results in (**b**) and (**d**). (**b**) The illustration of a 15-s segment where signals in the *shaded areas* were severely contaminated by large eye movements and environmental noises. (**c**) The topographical maps generated at four peak time points p1, p2, p3, and p4 (*vertical lines* in **b**) of four waves in IC3 at 203.9, 205.7, 210.4, and 211.1 s. (**d**) The 17 decomposed ICs show that diseased-related patterns were PLEDs (IC3) and epileptiforms (IC5) and the artifacts were eye movements (IC8) and noise (IC15). (**e**) The corresponding spatial maps of IC3, IC5, IC8, and IC15.

4.2 Patients and EEG Recordings

Five patients (all male) with sporadic CJD, aged 73, 74, 85, 52, and 80 years old were recruited in this study (for details, see Table 4.1). All of them met the criteria of probable CJD defined by WHO, were examined by board-certified neurologists, and underwent extensive diagnostic workups, including clinical, neurophysiological, neuroradiological examinations, and the CSF analysis. Disease onset was determined retrospectively based on history and clinical presentations as reported by the patients themselves and their relatives. The onset times of patient 1 to patient 5 were

Table 4.1: Clinical data of probable CJD patients

Patient	Gender	Age at onset	Disease onset	Clinical presentation	Original EEG report
1	M	73 y/o	6	Memory impairment	PLED, DBS 7 Hz
2	M	74 y/o	9	Memory impairment	FIRDA, DBS 7–8 Hz
3	M	85 y/o	4	Memory impairment	PLED, DBS 6–7 Hz
4	M	52 y/o	5	Memory impairment	DBS 4–5 Hz
5	M	80 y/o	3	Memory impairment	Periodic epileptiform

Disease onset time: weeks before the first admission. DBS: diffuse background slowing

6, 9, 4, 5, 3 weeks, respectively, before the first EEG recording. The EEGs were acquired using a 19-channel Nicolet EEG system (digitized at 250 Hz) with Ag/AgCl surface electrodes, which were placed based on the configuration of the international 10–20 system (Fig. 4.1c). We used the referential montage, rather than the bipolar or standard EEG, because the EEG signals can be expressed as $\mathbf{X} = \mathbf{AS} - \mathbf{Ref}$ so that the mixing matrix can be obtained directly from FastICA (the **Ref** term was eliminated in the zero-mean preprocessing of FastICA). The use of bipolar montage would make the recovery of the mixing matrix much more difficult since the bipolar EEG signals are formulated as $\mathbf{X} = (\mathbf{A_1} - \mathbf{A_2})\mathbf{S}$ with the additional constrain $\mathbf{A1}(i,j) = \mathbf{A2}(i,j+1)$. Five-minute EEG recording was clipped for each subject, which was bandpass filtered between 0.5 and 10 Hz prior to the ICA process. In this study, the infinite impulse response (IIR) digital filter was designed based on the Butterworth magnitude response:

$$\|P_L(\Omega)\| = \frac{1}{\sqrt{1 + \Omega^{2L}}}, \quad 1 \leq L, \tag{4.1}$$

where Ω was the analog frequency and L was the order of the normalized low-pass analog filter [21]. Furthermore, the associated s-plane poles were given by

$$s_k = \exp\left(\frac{j(2k + L - 1)\pi}{2L}\right), \quad 1 \leq k \leq 2L. \tag{4.2}$$

The bandpass filtering of the EEG was performed by the 6th-order high-pass filter followed by the 16th-order low-pass filter, which were implemented using MATLAB build-in functions.

Figure 4.2b displays a 15-s waveform of the 17-channel EEG (excluding two referential electrodes, Ref1 and Ref2) from one patient. We selected several time points at which the negative peaks or positive peaks (Figs. 4.2b, 4.3b, and 4.4b) are in conjunction with the corresponding topographic maps (Figs. 4.2c, 4.3c, and 4.4c) which may possess some physiological meanings. However, due to the mixture of source signals, such as disease-related waveforms, environmental noises, and eye-movement artifacts, the disease-related compartments can be barely discerned either from the waveforms or from the topographic maps. It should be noted that the 15-s time windows showed in the Figs. 4.2, 4.3, and 4.4 were selected merely to

demonstrate such obscure mixture in the raw EEGs (Figs. 4.2b, 4.3b, and 4.4b). In our implementation, ICA was applied to the whole 5-min recording of each patient and the selection of the interval of interest prior to ICA calculation was not needed.

4.3 Methods

4.3.1 Independent Component Analysis and Extraction of CJD-Related Components

Independent component analysis is a statistical method that has been developed to extract independent signals from a linear mixture of sources. Let $\mathbf{X}_{mxn}$ denote the measured data with m and n being the number of channels and the number of data samples, respectively. In the context of ICA, it is assumed to be linear combinations of unknown independent components and can be expressed as

$$\mathbf{X}_{mxn} = \mathbf{A}_{mxk} \cdot \mathbf{S}_{kxn}, \tag{4.3}$$

where $\mathbf{S}$ contains k independent sources with the same data length as $\mathbf{X}$, and $\mathbf{A}$ is a constant mixing matrix with the kth column representing the spatial weights corresponding to the kth component of $\mathbf{S}$. Given the measurement $\mathbf{X}$, ICA techniques attempt to recover both the mixing matrix $\mathbf{A}$ and the independent sources $\mathbf{S}$. In the present study, all calculations were performed using the FastICA algorithm [11, 23]. The FastICA technique first removes means of the row vectors in the $\mathbf{X}$ matrix and then uses a whitening procedure, implemented by principal components analysis [1], to transform the covariance matrix of the zero-mean data into an identity matrix. In the next step, FastICA searches for a rotation matrix to further separate the whitened data into a set of components which are as mutually independent as possible. In combination with previous whitening process, the matrix $\mathbf{X}$ is transformed into a matrix $\mathbf{S}$ via an unmixing matrix $\mathbf{W}$, i.e.,

$$\mathbf{S}_{kxn} = \mathbf{W}_{kxm} \mathbf{X}_{mxn}, \tag{4.4}$$

so that rows of $\mathbf{S}$ are mutually independent. The fixed-point method for solving $W = (w_i, \cdots, w_k)^T$ in the FastICA, where k is the number of independent sources, can be summarized as follows [11]:

For $i = 1 : k$,

1. Randomly choose a weighting vector w_i
2. Let $w^+ = E\{xg(w_i^T x)\} - E\{g'(w_i^T x)\}w_i$, where
   $g(u) = \tanh(cu)$, $1 \leq c \leq 2$
3. Let $w_i = w_i^+ / \|w_i^+\|$
4. Go back to step 2 if not converge.

5. Decorrelation by Gram–Schmidt-like scheme, Let

$w_i = w_i - \sum_{j=1}^{j-1} w_i^T w_j w_j$

6. Renormalize w_i, Let $w_i = w_i / \|w_i\|$

end

Since EEG can be considered as a linear combination of electric brain activities [5], we employed ICA to extract the disease-related components from the EEG of five patients. In this study, each preprocessed epoch was arranged across m channels ($m = 17$) and n sampled points ($n = 250 * 300$) into a matrix $\mathbf{X}$. The ith row contains the observed signal from the ith EEG channel, and the jth column vector contains the observed samples at the jth time point across all channels. FastICA was applied on each preprocessed epoch to resolve the $\mathbf{W}$ and $\mathbf{S}$. After estimating the unmixing matrix $\mathbf{W}$, we can recover the temporal waveforms by applying the inverse matrix of $\mathbf{W}$ on both sides of Equation (4.4) to yield

$$\underset{mxn}{\mathbf{X}} = \underset{mxk}{\mathbf{W}^{-1}} \cdot \underset{kxn}{\mathbf{S}}, \tag{4.5}$$

where $\mathbf{W}^{-1}$ is the best estimation of the mixing matrix $\mathbf{A}$ in Equation (4.3). In the cocktail-party problem, a popular example of ICA model, the kth row of $\mathbf{S}$ represents the voice from the kth speaker, and the element of mixing matrix $\mathbf{A}$ in the mth column and kth row, i.e., represents the weighting of the voice from the kth speaker recorded in the mth microphone. In other words, the kth column of $\mathbf{A}$ represents the weightings of the voice of kth speaker at each microphone. In this study, $\mathbf{S}$ represents the time sequences of activation sources, i.e., temporal waveforms of ICs in Figs. 4.2, 4.3, 4.4, and 4.5, and $\mathbf{A}$ stands for the weighting of sources recorded from electrodes. Since $\mathbf{W}$ is the estimated unmixing matrix, each column represents a spatial map describing the weightings of the corresponding temporal component at each EEG channel. These spatial maps will hereinafter be referred to as IC spatial maps. The validation of applying ICA to decompose EEG data has been addressed in the previous studies [10, 13, 14, 15, 16, 18, 17, 22, 23, 26]. In this study, we have also varied the data length, namely 1-, 2-, 3-, 4-, and 5-min epoch of data, to evaluate the performance of ICA and applied PCA on the same data sets for comparing their results on the feature extraction.

4.3.2 Bayesian Information Criterion

We have adopted the Bayesian information criterion (BIC) [2, 9, 19], which was based on the estimation of posterior probability $P(X|A,k)$ given the number of sources k and the observed data $\mathbf{X}$ to estimate the number of sources. The posterior probability was the function of $\mathbf{A}$ given by

$$P(X|A,k) = \prod_k \frac{1}{\sqrt{|2\pi\Lambda_k|}} \left(\frac{1}{|\det(A)|} \right)^T \cdot \exp\left(-\frac{1}{2} \sum_{t,t'} \hat{S}_{k,t} (\Lambda_k^{-1})_{t,t'} \hat{S}_{k,t'} \right), \tag{4.6}$$
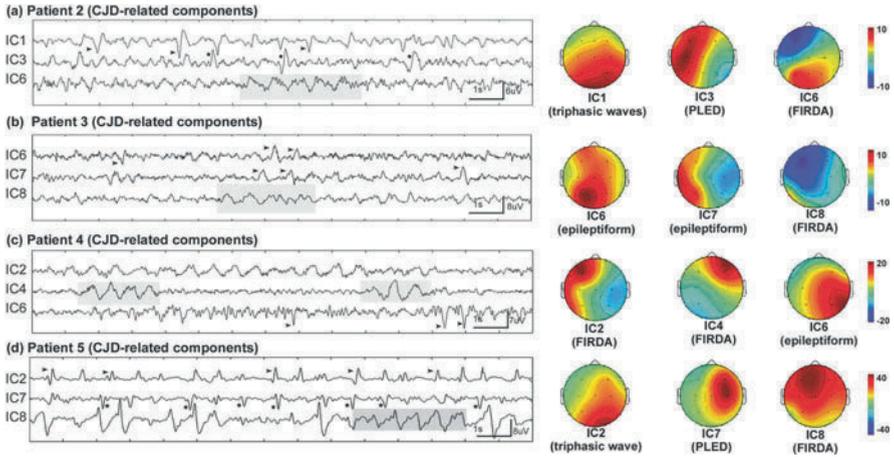
Fig. 4.5: Summarized ICA results from patient 2 to patient 5. Each panel shows the selected ICs and corresponding spatial maps for one patient. (**a**) The ICA results display generalized triphasic waves (IC1), PLEDs lateralized to the left hemisphere (IC3), and slow waves at delta frequency (*shaded area* of IC6). (**b**) The ICA results show epileptiforms (IC6, IC7) and FIRDA (*shaded area* of IC8). (**c**) The ICA results show the prominent FIRDA over left frontal-temporal area (IC2) and right frontal region (*shaded area* of IC4), and epileptiforms on the right temporal-occipital lobe (IC6). (**d**) The ICA results show periodic triphasic waves on the right occipital lobe (IC2), the PLEDs on the right frontal-central area (IC7), and the diffused delta waves (IC8).

where the notation $\hat{S}_{k,t}$ was the sources estimated from **A** and **X**, $\hat{S}_{k,t} = \sum_l (A^{-1})_{k,l} X_{l,t}$, Λ_k was the covariance matrix of sources, and t was the time point. In theory, the number of sources that produced the maximal posterior probability would be selected since the predicted model was best fit to the observed data.

4.4 Results

4.4.1 Determination of the Number of Sources

A number of sources ranging from 2 to 17 (the number of channels) were introduced to compute the posterior probabilities and the results in Fig. 4.6 demonstrated that values of posterior probabilities were comparable when N was between 12 and 17. In fact, the resultant CJD-related components were also comparable when N varied from 12 to 17. Instead of using the BIC for determining the number of sources, we simply used the number of channels as the number of sources, as suggested by the previous studies [10, 14, 15, 16, 18, 17, 26].
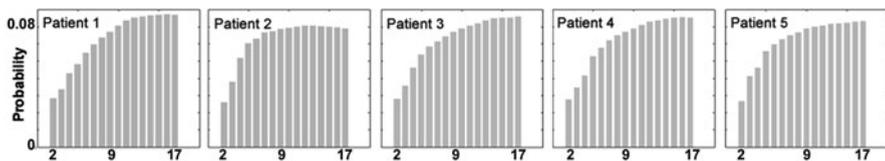
Fig. 4.6: The number of sources estimated by using the Bayesian information criterion (BIC) from patient 1 to patient 5. Each panel shows the estimated posterior probabilities (histograms) of a patient. A number of sources range from 2 to 17 (the number of channels) were given for computing the posterior probabilities (Equation 4.6). It is evident that all the estimated posterior probabilities are comparable in each plot when the source numbers are between 12 and 17.

4.4.2 CJD-Related Feature Extraction

We have observed that the distinct disease-related patterns were likely to occur in different time windows. Three 15-s windows (Figs. 4.2a, 4.3a, and 4.4a) were selected to illustrate the ICA results obtained from a 5-min EEG data. The resultant independent temporal waveforms (patient 1, 73 y/o) were presented in Figs. 4.2d, 4.3d, and 4.4d, respectively, and the corresponding spatial maps elucidating CJD-related characteristics or artifacts were depicted in Figs. 4.2e, 4.3e, and 4.4e, respectively. The CJD-related components shown in Figs. 4.2d, e, 4.3d, e, and 4.4d, e are the PLEDs lateralized to the right hemisphere (IC3), triphasic waves on the occipital lobe (IC4), and the PLEDs (IC3) as well as the epileptiforms covering the whole brain (IC5), respectively. The component IC2 was the artifact caused by eyeblinks since the spikes occurred intermittently with irregular shapes and large weights exhibited in the prefrontal area of the corresponding spatial map. Similarly, IC8 was identified as an artifact due to left eye movements. The remaining ICs may correspond to spontaneous brain activities irrelevant to CJD or artifacts and were not taken into account in the analysis.

Figure 4.5 summarizes the individual CJD-related components from the other patients. Each panel shows the selected temporal independent components and the corresponding spatial maps for one patient. The ICA results from patient 2 (74 y/o), display generalized triphasic waves (IC1), PLEDs lateralized to the left hemisphere (IC3), and slow waves at delta frequency (shaded area of IC6) (Fig. 4.5a). In Fig. 4.5b, epileptiforms (IC6, IC7) and FIRDA (shaded area of IC8) were resolved from patient 3 (85 y/o). Figure 4.5c shows the prominent FIRDA over the left frontal-temporal area (IC2) and the right frontal region (shaded area of IC4), and epileptiforms on the right temporal-occipital lobe (IC6) from patient 4 (52 y/o). Finally, Fig. 4.5d displays that the positive periodic triphasic waves appear predominantly on the right occipital lobe (IC2), the PLEDs on the right frontal-central area (IC7), and the diffused delta waves (IC8) from patient 5 (80 y/o).

Figure 4.7 shows the results when the 1-, 2-, 3-, 4-, and 5-min epochs of data were analyzed by ICA. The bars with different colors in the Fig. 4.7a–e represent the
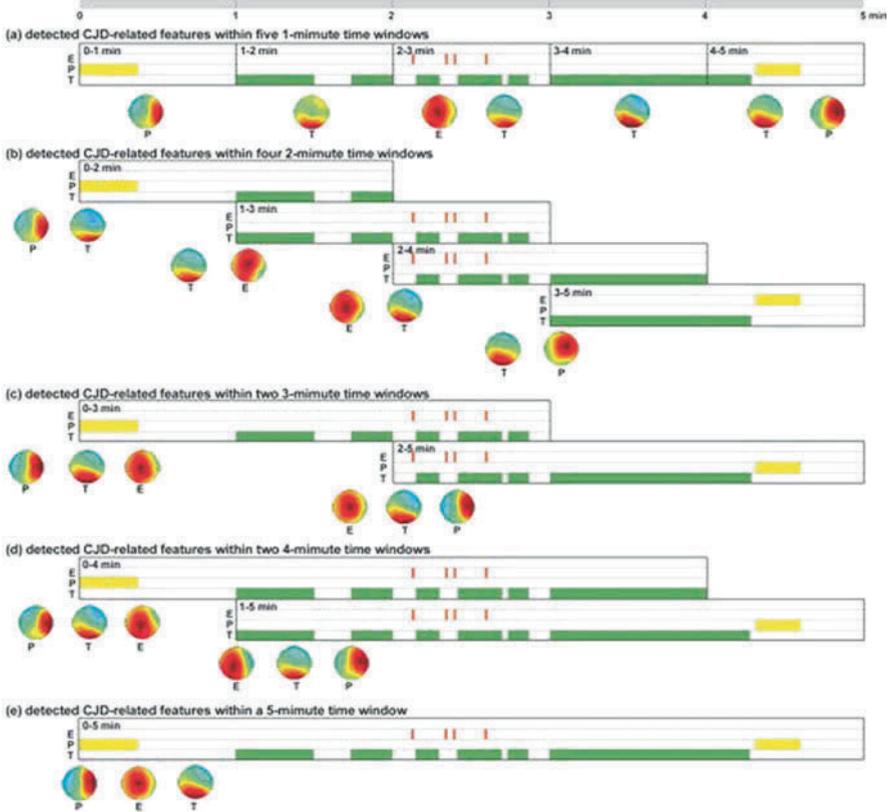
Fig. 4.7: Performance of ICA when the 1-, 2-, 3-, 4-, and 5-min epochs of data were analyzed. The bars with different colors in the panels (**a**)–(**e**) represent the time periods during which features were resolved, namely, the epileptiform, PLED, and triphasic waves. It can be seen that the ICA results remained unchanged under various data lengths where the same CJD-related patterns repeatedly appeared. Specifically, the PLED presents in the first 20 s within the first minute and in the 17th–35th s in the fifth minute of the epoch (see the *yellow bars* in the 1st and 5th windows in (**a**), 1st and 4th windows in (**b**), 1st and 2nd windows in (**c**), 1st and 2nd windows in (**d**) and in (**e**)). The epileptiform were detected within the 3rd window in (**a**), 2nd and 3rd windows in (**b**), 2nd and 3rd windows in (**c**), 1st and 2nd windows in (**d**) and in (**e**) (see orange bars). Finally, the triphasic waves can be observed across from the 2nd to the 5th windows in (**a**), which also appeared in the 1st–4th windows in (**b**), 1st and 2nd windows in (**c**), 1st and 2nd windows in (**d**) and in (**e**) (see *green bars*). It should be noted that not only the temporal features preserved the same waveforms and durations, but also the three corresponding spatial maps remained resemble. E: epileptiform, P: PLED, T: triphasic wave.

time periods during which features were resolved, namely, the epileptiform, PLED, and triphasic waves. It can be seen that the ICA results remained unchanged under various data lengths where the same CJD-related patterns repeatedly appeared. Specifically, the PLED presents in the first 20 s within the first minute and in the 17th–35th seconds in the fifth minute of the epoch (see the yellow bars in the 1st and 5th windows in (a), 1st and 4th windows in (b), 1st and 2nd windows in (c), 1st and 2nd windows in (d) and in (e)). The epileptiform were detected within the 3rd window in (a), 2nd and 3rd window in (b), 2nd and 3rd window in (c), 1st and 2nd window in (d) and in (e) (see orange bars). Finally, the triphasic waves can be observed across from the 2nd to the 5th windows in (a), which also appeared in the 1st–4th windows in (b), 1st and 2nd windows in (c), 1st and 2nd windows in (d) and in (e) (see green bars). It should be noted that not only the temporal features preserved the same waveforms and durations, but also the three corresponding spatial maps remained resemble (see Fig. 4.7). Similar results have been obtained from other patients (not shown).

4.4.3 Feature Extraction by PCA

It has been reported that the use of ICA under the assumption of source independence can separate more realistically neurophysiologic signals in comparison with the principal component analysis (PCA) [10, 12]. Since the EEG signals induced by eyeblinking or contaminated by electrical noise usually present far larger variances than physiological signals, the covariance-based PCA decomposing procedure is inferior to ICA for resolving meaningful brain activities. As shown in the Fig. 4.8b where the same time window in Fig. 4.4a was selected, the temporal waveforms of the first four principal components (eigenvectors corresponding to the first four largest eigenvalues) merely exhibit the preservation of the most power of the original signals. None of them extracted the evident eyeblinking artifacts or CJD-related features from the raw EEG as compared to the ICA results in Fig. 4.4.

4.5 Discussions

This study aims to extract the CJD-related waveforms in conjunction with the spatial dominances from the EEG recordings for the early diagnosis of CJD. Our results demonstrate that ICA is an effective tool for distinguishing FIRDA, PLEDs and PSWCs from EEG recordings in the early stage of CJD (Figs. 4.2d, e, 4.3d, e, 4.4d, e, and 4.5) with dominance in each corresponding spatial map being revealed. In comparison with the raw EEG data in the shaded areas in Figs. 4.2b, 4.3b, and 4.4b, where the CJD-related waveforms were severely smeared by the large potentials of eye movements, three PLEDs, four triphasic waves, and two epileptiforms can be evidently recovered in the shaded areas of IC3 in Fig. 4.2d, IC4 in Fig. 4.3d,
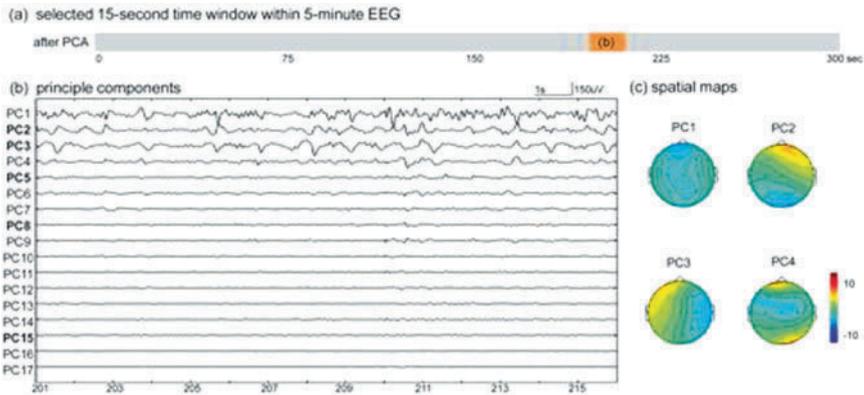
Fig. 4.8: The selected EEG segment and PCA results within the same time window as in Fig. 4.4 from patient 1. (**a**) The 15-s time window (201–216 s) is used to display results in (**b**). (**b**) The 17 decomposed PCs show that the temporal waveforms of the first four principal components (eigenvectors corresponding to the first four largest eigenvalues) merely exhibit the preservation of the most power of the original signals. (**c**) The corresponding spatial maps of PC1 to PC4. None of them extracted the evident eyeblinking artifacts or CJD-related features from raw EEG as compared to the ICA results in Fig. 4.4.

and IC5 in Fig. 4.4d, respectively. In addition, it should be noted that any 5-min IC waveform only corresponds to a single spatial map and the predominant region for IC3, IC4, and IC 5 is manifested in Figs. 4.2e, 4.3e, and 4.4e, respectively. On the contrary, the topographical maps produced from the peak times of the similar waveforms in the raw data varied from one to another. To illustrate this, we particularly chosen four peak times of the disease-related IC waveforms and displayed the topographical maps based on the raw EEG at these peak times. As shown in the vertical lines in Fig. 4.2b or d, four peak time points p1, p2, p3, and p4 of four waves in IC3 at 3.3, 5.1, 9.6, and 10.9 s were selected and the corresponding topographical maps produced from the raw data presented distinct patterns (Fig. 4.2c), which were difficult to interpret for further analysis. Similar phenomenon and difficulty can be seen in Figs. 4.3c and 4.4c.

Another salient feature of ICA is that, even a CJD-related wave hid at different time windows and obscured across multiple channels, ICA is effective to extract such waveforms from different channels into a single independent component, as illustrated by IC3 in Figs. 4.2d, e and 4.4d, e, where repeated waves of PLEDs were identified in IC3 which occurred during 2–17 and 201–216 s. Besides, muscular artifacts and environmental noise have been isolated by ICA which were in congruent with previous studies [14, 15, 23]. The intermittent high amplitude waves induced by eyeblinks with maximum over the prefrontal area were presented within IC2 in Figs. 4.2d, e and 4.3d, e, large irregular waves caused by eye movements on the left frontal region were within IC8 in Figs. 4.2 and 4.3d, e, and environmental noises

exhibiting irregularly transient waveforms in a single channel were within IC11 in Fig. 4.2d, e, IC15 in Figs. 4.3 and 4.4d, e.

Most of the previous studies have reported that only one CJD pattern appeared in each stage. The co-occurrence of FIRDA and PLEDs or triphasic waves from the same EEG data has not been explored. For example, either the FIRDA or FIRDA-like waveforms could be found in the early stage of CJD in most cases [8, 25], or the PLEDs appeared initially and were replaced by PSWCs progressively in the middle or late stage [1, 6]. The ICA results, nevertheless, illustrated that the FIRDA and PLEDs, or FIRDA and epileptiforms, or FIRDA and triphasic waves concurrently appeared in the same EEG data for each patient. As shown in Table 4.2, the PLEDs, epileptiforms, and triphasic waves from the 5-min EEG signals of patient 1 can be, respectively, recovered in IC3, IC5, and IC4, the FIRDA, PLEDs, and triphasic waves in IC6 (shaded area in Fig. 4.5a), IC3 (stars in Fig. 4.5a), and IC1 (arrows in Fig. 4.5a) from patient 2, and in IC8 (shaded area in Fig. 4.5d), IC7 (stars in Fig. 4.5d), and IC2 (arrows in Fig. 4.5d) from patient 5. In addition, FIRDA can be seen in IC8 (shaded area in Fig. 4.5b) and epileptiforms in IC6 and IC7 (arrows in Fig. 4.5b) from patient 3, and FIRDA in IC2 and IC4 (shaded area in Fig. 4.5c) and epileptiforms (arrows in Fig. 4.5c) in IC6 from patient 4. These findings suggest that the EEG in the early stage of CJD is heterogeneous and concurrent appearance of different CJD patterns should be taken into account in the diagnosis.

Table 4.2: The concurrent appearance of different CJD waveforms in the same EEG data from each patient

Patient	FIRDA	PLEDs	Epileptiform	Triphasic wave
1 (Figs. 4.2, 4.3, and 4.4)		IC3 IC5	IC4	
2 (Fig. 4.5)	IC6	IC3		IC1
3 (Fig. 4.5)	IC8		IC6, IC7	
4(Fig. 4.5)	IC2, IC4		IC6	
5 (Fig. 4.5)	IC8	IC7		IC2

It should be noted that only the PSWC had been reported with a 85% specific to the late CJD, the unaccompanied occurrence of each pattern, such as FRIDA, epileptiform, PLED, and triphasic waves, might be seen in other neurological disorders. Therefore, the hypothesis that EEGs of the CJD manifested the co-occurrence of multiple disease-related features was further tested against the Alzheimer's disease (AD) group with five patients who were all male and aged 85, 73, 45, 72, and 79 years old, i.e., age and gender matched with the CJD group. After applying ICA on the AD group, we examined the independent components to detect the disease-related features. No co-occurrence of multiple disease-related features was found in the AD group, except that two ICs were detected to consist of FIRDA in patient 1 and one IC consisted of the epileptiform in patient 4. Based on the co-occurrence of multiple disease-related features exhibited in both groups, the difference between AD and CJD groups was statistically significant (two-sample Wilcoxon test,

$p < 0.05$). Accordingly, the concurrent existence of multiple features presented in the early EEG of CJD patients can be used as an assistive tool for the early diagnosis of CJD.

The order of same CJD-related components may vary from patient to patient since both the mixing matrix A and source matrix S are unknown, which allows the change of the order of rows in S. To see this, we can substitute a permutation matrix P and its inverse into the model, $\mathbf{X} = \mathbf{AS}$, to give $\mathbf{X} = (\mathbf{AP^{-1}})(\mathbf{PS})$. The matrix $\mathbf{AP^{-1}}$ is a new unknown mixing matrix to be solved by the FastICA algorithm [11] and the rows of $\mathbf{PS}$ are original sources but in different order because each row or column in $\mathbf{P}$ consists of only one nonzero element with value 1. It is much easier to detect the CJD-related patterns from the unmixed signals rather than from the obscured mixing signals as illustrated in Figs. 4.2, 4.3, and 4.4, although the same CJD-related sources would occur at different channels among patients. In addition, we found that the ICs consisting of larger spikes, such as irregular waveforms and bursts, tended to be decomposed earlier from the mixing signals in the calculation of FastICA. All the CJD-related features, i.e., sharp waves or epileptiform, have been recognized from ICs lower than IC8.

It is noted that the matrix $\mathbf{S}$ has lower amplitude in comparison with the matrix $\mathbf{X}$. Such an amplitude difference comes from the nature of the linear mixing model and the algorithm of FastICA. Based on the vector form of the model $x_j = a_j s_1 + \cdots + a_{ij} s_i + \cdots$, it can be rewritten into the form $x_j = a_{j1} s_1 + \cdots + (a_{ji} \alpha^{-1})(\alpha s_i) + \cdots$, where α is any arbitrarily nonzero scalar. In other words, the solutions of mixing $\mathbf{A}$ and source matrix $\mathbf{S}$ are not unique since any source can be multiplied by a nonzero scalar which can always be canceled by dividing the corresponding column of $\mathbf{A}$ by the same scalar. In order to fix the magnitude of the independent components, each source is restricted to have unit variance in the FastICA calculation [11]. As a result, the resolved matrix $\mathbf{S}$ has lower amplitude than the matrix $\mathbf{X}$.

4.6 Conclusions

We have employed ICA to detect the co-occurrence of multiple CJD-related patterns from the EEG recording for aiding to the early diagnosis. Results demonstrate that ICA is an effective tool for simultaneously recovering the FIRDA, PLEDs, and triphasic waves (early PSWCs) that can be hardly discerned by visual inspection from the contaminated EEG recordings. The concurrent appearance of FIRDA and PLEDs or triphasic waves from the same EEG data suggests that the heterogeneity of EEG in the early diagnosis of CJD should be taken into account.

References

1. Au, W., Gabor, A., Vijayan, N., et al. Periodic lateralized epileptiform complexes (pleds) in creutzfeldt-jakob disease. Neurology **30**, 611–617 (1980)
2. Calamante, F., Mørup, M., Hansen, L. Defining a local arterial input function for perfusion mri using independent component analysis. Magn Reson Med **52**, 789–797 (2004)
3. Cambier, D., Kantarci, K., Worrell, G., et al. Lateralized and focal clinical, eeg, and flair mri abnormalities in creutzfeldt-jakob disease. Clin Neurophysiol **114**, 1724–1728 (2003)
4. Collins, S., Sanchez-Juan, P., Master C., et al. Determinants of diagnostic investigation sensitivities across the clinical spectrum of sporadic creutzfeldt-jakob disease. Brain **129**, 2278–2287 (2006)
5. Fisch, B. Fisch and Spehlmann's EEG Primer. Elsevier Science B.V., Amsterdam (1999)
6. Fushimi, M., Sato, K., Shimizu, T., et al. Pleds in creutzfeldt-jakob disease following a cadaveric dural graft. Clin Neurophysiol **113**, 1030–1035 (2002)
7. Geschwind, M., Martindale, J., Miller, D., et al. Challenging the clinical utility of the 14–3-3 protein for the diagnosis of sporadic creutzfeldt-jakob disease. Arch Neurol **60**, 813–816 (2003)
8. Hansen, H., Zschocke, S., Sturenburg, H., et al. Clinical changes and eeg patterns preceding the onset of periodic sharp wave complexes in creutzfeldt-jakob disease. Acta Neurol Scand **97**, 99–106 (1998)
9. Hansen, L., Larsen, J., Kolenda, T. Blind detection of independent dynamic components. Proc IEEE Int Conf Acoust Speech Signal Process ICASSP **5**, 3197–3200 (2001)
10. Hung, C., Lee, P., Wu, Y, et al. Recognition of motor imagery electroencephalography using independent component analysis and machine classifiers. Ann Biomed Eng **33**, 1053–1070 (2005)
11. Hyvärinen, A., Karhunen, J., Oja, E. Independent Component Analysis. John Wiley & Sons, Inc., New York (2001)
12. Joyce, C., Gorodnitsky, I., Kutas, M. Automatic removal of eyemovement and blink artifacts from eeg data using blind component separation. Psychophysiology **41**, 313–325 (2004)
13. Jung, K., Kim, J., Kim, D., et al. Independent component analysis of generalized spike-and-wave discharges: Primary versus secondary bilateral synchrony. Clin Neurohysiol **116**, 913–919 (2005)
14. Jung, T., Makeig, S., Humphries, C., et al. Removing electroencephalographic artifacts by blind source separation. Psychophysiology **37**, 163–178 (2000)
15. Jung, T., Makeig, S., Westerfield, M., et al. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin Neurohysiol **111**, 1745–1758 (2000)
16. Kao, Y., Guo, W., Wu, Y., et al. Hemodynamic segmentation of mr brain perfusion images using independent component, bayesian estimation and thresholding. Magn Reson Med **49**, 885–894 (2003)
17. Lee, P., Hsieh, J., Wu, C., et al. The brain computer interface using flash visual evoked potential and independent component analysis. Ann Biomed Eng **34**, 1641–1654 (2006)
18. Lee, P., Wu, Y., Chen, L., et al. Ica-based spatiotemporal approach for single-trial analysis of postmovement meg beta synchronization. NeuroImage **20**, 2010–2030 (2003)
19. MacKay, D. Bayesian model comparison and backprop nets. Adv Neural Inf Process Syst **4**, 839–846 (1992)
20. Schröter, A., Zerr, I., Henkel, K., et al. Magnetic resonance imaging in the clinical diagnosis of creutzfeldt-jakob disease. Arch Neurol **57**, 1751–1757 (2000)
21. Stearns, S., David, R. Signal Processing Algorithms in Matlab. Prentice Hall PTR, New Jersey (1996)
22. Urrestarazu, E., LeVan, P., Gotman, J. Independent component analysis identifies ictal bitem-poral activity in intracranial recordings at the time of unilateral discharges. Clin Neurophysiol **117**, 549–561 (2006)
23. Vigario, R., Särelä, J., Jousmaki, V. et al. Independent component approach to the analysis of eeg and meg recordings. IEEE Trans Biomed Eng **47**, 589–593 (2000)

24. Wieser, H., Schindler, K., Zumsteg, D. Eeg in creutzfeldt-jakob disease. Clin Neurophysiol **117**, 935–951 (2006)
25. Wieser, H., Schwarz, U., Blattler, T., et al. Serial eeg findings in sporadic and iatrogenic creutzfeldt-jakob disease. Clin Neurophysiol **115**, 2467–2478 (2004)
26. Wübbeler, G., Ziehe, A., Mackert, B., et al. Independent component analysis of noninvasively recorded corticalmagnetic dc-fields in humans. IEEE Trans Biomed Eng **47**, 594–599 (2000)

Chapter 5
Comparison of Supervised Classification Methods with Various Data Preprocessing Procedures for Activation Detection in fMRI Data

Mahdi Ramezani and Emad Fatemizadeh

Abstract In this study we compare five classification methods for detecting activation in fMRI data: Fisher linear discriminant, support vector machine, Gaussian nave Bayes, correlation analysis and k-nearest neighbor classifier. In order to enhance classifiers performance a variety of data preprocessing steps were employed. The results show that although kNN and linear SVM can classify active and nonactive voxels with less than 1.2% error, careful preprocessing of the data, including dimensionality reduction, outlier elimination, and denoising are important factors in overall classification.

5.1 Introduction

Studying the functionality of the brain with versatile noninvasive tools has boost enormously in recent years. It is widely believed that blood oxygen level, the ratio of oxygenated to deoxygenated hemoglobin in the blood at the corresponding in the brain, is influenced by local neural activity. Based on the blood oxygen level-dependent (BOLD) principle, functional magnetic resonance imaging (fMRI) has become one of the typical tools in the neurological disease diagnosis and human brain research. This imaging method can quantify hemodynamic changes induced by neuronal activity in human brain at high-spatial resolution during sensory or cognitive stimulations. fMRI technology offers the promise of revolutionary new approaches to studying human cognitive processes, provided we can develop appropriate data analysis methods to make sense of this huge volume of data. The

Mahdi Ramezani
Biomedical Image and Signal Processing Laboratory (BiSIPL), School of Electrical Engineering, Sharif University of Technology, Tehran, Iran, e-mail: `Ramezani@ee.sharif.edu`

Emad Fatemizadeh
Biomedical Image and Signal Processing Laboratory (BiSIPL), School of Electrical Engineering, Sharif University of Technology, Tehran, Iran, e-mail: `Fatemizadeh@sharif.edu`

vast majority of published researches summarizes average fMRI responses when the subject responds to repeated stimuli of some type (e.g., reading, mental imagery, remembering) [5]. Other researchers have since applied various multivariate methods in analyzing distributed response patterns in the human fMRI data sets: approaches include training machine learning classifiers to automatically decode the subjects' cognitive state at a single time instant or interval [6]. These statistical pattern recognition algorithms are powerful because they project the activity of multiple voxels to achieve a discriminative separation of the activity patterns. Before performing pattern recognition algorithms, there is a need to select a subset of voxels for further analysis. The procedure of selecting particular voxels can greatly enhance classification performance [3]. The enhancement consists of avoiding the "curse of dimensionality" by reducing the dimension of the space of patterns to be labeled and removing noise features that can only degrade performance [4]. Likewise, most classification of fMRI data depends on an effective feature selection procedure being applied beforehand [5]. In a typical fMRI study, time courses of more than several thousand voxels are simultaneously acquired. Many of these are uninformative and could severely damage the performance of the algorithm. In order to perform pattern recognition more efficiently, one should use a technique to find a reasonable subset of voxels to feed the classifiers. The aim of this work is to exploit supervised classification techniques for the voxel selection procedure. The goal of these analyses is to detect the activated voxels (those voxels with highest overall responsiveness). In the present study we applied several pattern recognition techniques and data preprocessing approaches to compare their performance in classifying active and inactive voxels. We used five classification procedures: the Fisher linear discriminant (FLD), support vector machine (SVM), Gaussian nave Bayes (GNB), correlation analysis, and k-nearest neighbor classifier (kNN). This chapter is organized as follows. In the next section, we briefly explain the acquisition of fMRI used in the application. In Section 5.3, we provide data preprocessing approaches. Section 5.4 describes the pattern recognition techniques. Results of experiments are presented in Section 5.5.

5.2 Data Set

In the studies described in this chapter, a data set from the SPM site http://www.fil.ion.ucl.ac.uk/spm/data/ was used which comprises whole brain BOLD/EPI images acquired on a modified 2T Siemens MAGNETOM vision system. This data set was the first ever collected and analyzed in the functional imaging laboratory (FIL). Each acquisition consisted of 64 contiguous slices ($64 \times 64 \times 64$ $3 \times 3 \times 3$ mm voxels). Acquisition took 6.05 s, with the repetition time (TR) set arbitrarily to 7 s. At whole 96 acquisitions were made from a single subject giving 16 42 s blocks. The condition for successive blocks alternated between rest and auditory stimulation, starting with rest. Auditory stimulation was bi-syllabic words presented binaurally at a rate of 60 per minute [2]. The images were then realigned to mitigate noise caused by

head motion, smoothed to reduce the effect of high-frequency noise on the analysis, and spatially normalized to allow for intersubject comparisons within SPM5 software. After that the activation map was obtained and was used as a gold standard in training the classifiers in this study. Figure 5.1 shows the activation map. For further analysis we use a global threshold in order to identify those voxels with highest overall responsiveness. Figure 5.2 shows the activation map after thresholding.
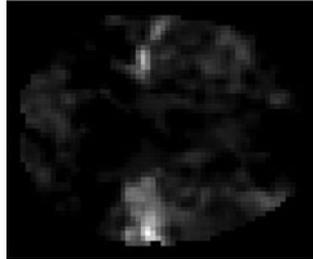
Fig. 5.1: The activation map.

Fig. 5.2: Active regions are shown as white pixels.

5.3 Data Preprocessing

To estimate the generalization ability of the classification methods, we split each data set into two nonoverlapping subsets: the training set on which each classifier was trained and the test set on which each classifier was tested. The procedure was repeated many times for different random partitions of the data, and results were averaged across the results. In order to enhance classifier performance a variety of data preprocessing steps were employed. Two of them were done for all voxels. First the data were normalized by subtracting the mean value and dividing by the overall standard deviation. Thus, each voxel had mean activity of 0 and unit standard deviation. Second the outliers were removed by setting all values that were beyond three

standard deviations from the mean to a fixed value of 3 or -3, depending on the sign of the original value. The remaining preprocessing steps were optionally performed before each classification. They were done alone or together. One of these preprocessing steps was singular vector decomposition (SVD) of the data. The SVD was done in order to denoise the data by keeping only some of principal components. For this purpose the eigenvalues of the matrix of the data were plot (Fig. 5.3) and only the important eigenvalues were kept. Another popular preprocessing approach which was used is the removal of noninformative features via subspace-based decomposition techniques. This approach proceeds by discarding the irrelevant subspace based on assumption that the sparse portion of the data space carries little, or no useful information. One of the approaches used in this study was to reduce the data dimension via principal component analysis (PCA). The principal component analysis is a representative of the unsupervised learning method which yields a linear projection for mapping the input vector of observations onto a new feature description which is more suitable for given task. It is a linear orthonormal projection which allows for the minimal mean square reconstruction error of the training data [1]. Another approach that was used to reduce the data dimension was linear discriminant analysis (LDA). The goal of the LDA is to train the linear data projection such that the class separability criterion is maximized. We further discuss the effect of these procedures in our "Results" section.
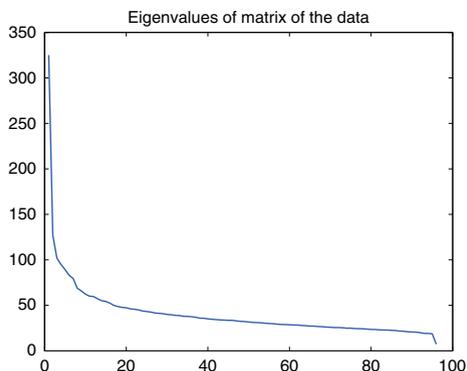
Fig. 5.3: Eigenvalues of the data matrix.

5.4 Pattern Recognition Methods

In this section, we describe our classification procedures. Five classifiers were used: Fisher linear discriminant (FLD), a linear support vector machine (SVM), Gaussian nave Bayes (GNB), correlation analysis, and k-nearest neighbor classifier (kNN). These classifiers were selected because they have been used successfully in other

applications involving high-dimensional data. We have used a Matlab implementation of the classifiers provided by statistical pattern recognition toolbox (abbreviated STPRtool) which is available online at http://cmp.felk.cvut.cz/cmp/cmp_software. html.

5.4.1 Fisher Linear Discriminant

The linear classification rule is composed of a set of discriminant functions which are linear with respect to both the input vector and their parameter vectors. In the case of the Fisher linear discriminant (FLD), the parameter vector ω of the linear discriminant function $f(x) = <\omega, x> + b$ is determined to maximize the class separability criterion. In other words, it aims to find a linear combination of voxels that discriminate between the two classes. The weights of this linear combination are given by

$$\omega = S_w^{-1}(\mu_1 - \mu_2), \tag{5.1}$$

where μ_1 and μ_2 denote the respective means of the first and second classes and S_w is the within class scatter matrix [7]. It can be shown that for Gaussian random vectors, with equal covariance matrices in both classes, this is similar to the optimal Bayesian classifier with the exception of a threshold value.

5.4.2 Support Vector Machine

SVM which is a linear classification algorithm does not assume a specific model of the data points but rather seeks to find the hyperplane (train the linear discriminant function $f(x) = <\omega, x> + b$) that separates the two classes with maximum margin. The training of the optimal parameters (ω^*, b^*) is transformed to the following programming task [7]:

$$\min \left\{ \frac{1}{2}\|\omega^2\| + C\sum_{i=1}^{N} \xi_i \right\}, \tag{5.2}$$

$$s.t \quad f(x_1)(<\omega, x_i> + b) \geq 1 - \xi_i, i = 1, 2, \dots, N. \tag{5.3}$$

$$\xi_i \geq 0, i = 1, 2, \dots, N, \tag{5.4}$$

where slack variables are used to relax the inequalities for the case of nonseparable data. The parameter C is a positive constant that controls the relative influence of the two competing terms: regularization and classification error. It is clear that minimizing the norm makes the margin maximum. This is a nonlinear optimization task subject to a set of linear inequality constraints. So the problem is a convex programming one, and the corresponding Lagrangian is given by [7]

$$\max \left\{ \frac{1}{2} \|\omega^2\| + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \mu_i \xi_i - \sum_{i=1}^{N} \lambda_i [f(x_i)(<\omega, x_i> +b) - 1 + \xi_i] \right\}, \quad (5.5)$$

$$s.t. \quad \omega = \sum_{i=1}^{N} \lambda_i f(x_i) x_i, \quad\quad\quad (5.6)$$

$$\sum_{i=1}^{N} \lambda f(x_i) = 0, \quad\quad\quad (5.7)$$

$$C - \mu_i - \lambda_i = 0, i = 1, 2, \ldots, N, \quad\quad\quad (5.8)$$

$$\lambda_i \geq 0, \mu_i \geq 0, i = 1, 2, \ldots, N. \quad\quad\quad (5.9)$$

In order to obtain the best value of the parameter C, the search space of SVM is set $C = 2$ to the power of -5 to 5. We also investigated a nonlinear variant with a radial basis function (RBF) kernel, which yielded similar results to the linear SVM.

5.4.3 Gaussian Nave Bayes

The GNB classifier uses the training data to estimate the probability distribution over fMRI observations, conditioned on the stimuli. Responses conditioned on the stimuli were modeled as Gaussians, where it was assumed that each voxel was independent of the others. The Gaussian mixture model (GMM) means and variances were estimated by maximal likelihood estimation. With the obtained model, the decision boundary for classification was the optimal boundary. The predicted class on test data was the most probable class under this model.

5.4.4 Correlation Analysis

The responses in the training set for active and inactive voxels were averaged separately to compute the mean responses for each category as templates. For prediction, the correlation coefficients between each test point (time series of a voxel) and each of the templates were obtained. Then, each test point was predicted to belong to Class 1 if the correlation coefficient for Class 1 was bigger than for Class 2, and to Class 2 otherwise.

5.4.5 k-Nearest Neighbor

The algorithm for k-nearest neighbor classifier is summarized as follows. Given an unknown feature vector (voxel), the k-nearest neighbors irrespective of class label was identified. Out of these k samples, the number of vectors that belong to class 1
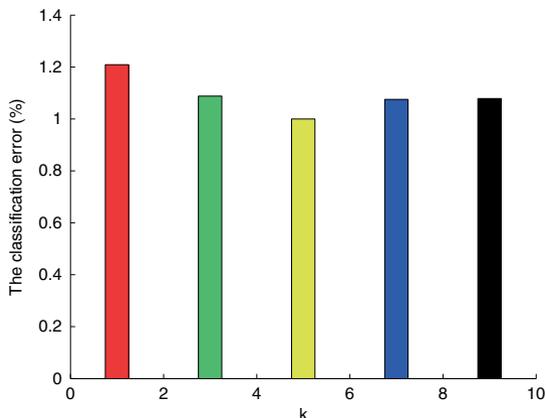
Fig. 5.4: The classification error for various values of k.

or 2 was identified. The feature vector was assigned to the class with the maximum number of samples [7]. We have used nearest neighbor with a Euclidean distance metric, considering values of 1, 3, 5, 7, and 9 for k. Figure 5.4 shows the classification error for these values of k.

As can be seen in Fig. 5.4, the 5NN classifier outperformed other classifiers. So we have chosen this kNN classifier in comparing the performance with the other pattern recognition methods, described above.

5.5 Results

In this section we present experimental results. As discussed earlier, we experimented with five classifier learning methods: FLD, SVM, GNB, correlation analysis, and kNN. We report the performance of each classifier with various preprocessing procedures mentioned in Section 5.4. By the obtained results we will be able to decide with which preprocessing procedure the classifier has the best performance. Performance for individual classifiers was measured by repeatedly splitting the data into training and test sets and averaging classification performance on each test set. Here the performance metric is a classification error. Table 5.1 shows the mean classification error of different classifiers with different preprocessing procedures. As can be seen, different preprocessing steps (denoising by SVD and dimension reduction by PCA or LDA) had a substantial impact on prediction accuracy. It is also clear that, in all experiments with different preprocessing procedures, performance is best for the linear SVM. The mean classification error was less than 1.1% for this classifier. The different preprocessing procedure used has a weak effect on the performance of the both kNN and SVM classifiers. They approximately perform equivalent (with classification error of less than 1.2%) and better than other

Table 5.1: Mean classification error of different classifiers with different preprocessing procedures

Preprocess	FLD	Linear SVM	Nonlinear SVM	GNB	Corr	kNN
None	2.261	0.9	2.25	50.824	62.522	1.168
SVD	36.228	1.05	2.03	13.766	64.775	1.055
LDA	2.858	0.9	1.692	1.572	11.288	1.198
LDA+SVD	2.826	0.82	1.078	1.768	17.066	1.108
PCA	2.38	0.9	2.098	1.842	2.17	1.06
PCA+SVD	3.128	0.6	2.02	1.57	3.022	1.106

approaches. Since correlation performs so much worse than the chance level in the first two cases of preprocessing procedures (without preprocessing and with SVD only), we will not consider it in observations of these two cases. Denoising the data by SVD in first two rows helped all algorithms besides linear SVM and FLD (actually increased FLD classification error by 34%). For instance, the accuracy of GNB was enhanced by 37% and the accuracy of kNN and nonlinear SVM were a little enhanced. Dimension reduction by PCA or LDA enhances performance on the majority of classifiers reported. It is clear that the classification error of all classifiers after reducing the data dimension via PCA is almost the same. So we can use any of the classifiers after applying the PCA without any concern about the results.

5.6 Conclusions

In this chapter we have compared different classification methods with various preprocessing procedures for detecting activation in fMRI data. The experimental results presented here demonstrate the feasibility of training classifiers to distinguish between active and inactive voxels of fMRI data. kNN and SVM perform equivalently and better than other approaches. They can classify the voxel with the classification error of less than 1.2%. Further work could include the use of other dimension reduction methods such as independent component analysis (ICA). It is also of interest to examine other classification algorithms.

References

1. Franc, V., Hlaváč, V. Statistical pattern recognition toolbox for matlab. Center for Machine Perception, Czech Technical University, Prague, Czech (2004)
2. Group, T.F.M. Spm5 manual. (2005)
3. Isabelle, G., Andre, E. An introduction to variable and feature selection. J Mach Learn Res **3**, 1157–1182 (2003)
4. Ku, S.-P., Gretton, A., Macke, J., Logothetis, N.K. Comparison of pattern recognition methods in classifying high-resolution bold signals obtained at high magnetic field in monkeys. Magn Reson Imaging **26**, 1007–1014 (2008)

5. Mitchell, T., Hutchinson, R., Just, M.A., Niculescu, R.S., Pereira, F., Wang, X. Classifying instantaneous cognitive states from fmri data. In: Americal Medical Informatics Association Annual Symposium, Washington DC 469 (2003)
6. Mitchell, T., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., Newman, S. Learning to decode cognitive states from brain images. Mach Learn **57**, 145–175 (2004)
7. Theodoridis, S., Koutroumbas, K. Pattern Recognition. Academic Press, San Diego, CA (2006)

Chapter 6
Recent Advances of Data Biclustering with Application in Computational Neuroscience

Neng Fan, Nikita Boyko, and Panos M. Pardalos

Abstract Clustering and biclustering are important techniques arising in data mining. Different from clustering, biclustering simultaneously groups the objects and features according their expression levels. In this review, the backgrounds, motivation, data input, objective tasks, and history of data biclustering are carefully studied. The bicluster types and biclustering structures of data matrix are defined mathematically. Most recent algorithms, including OREO, nsNMF, BBC, cMonkey, etc., are reviewed with formal mathematical models. Additionally, a match score between biclusters is defined to compare algorithms. The application of biclustering in computational neuroscience is also reviewed in this chapter.

6.1 Introduction

6.1.1 Motivation

With the number of database appearing in computational biology, biomedical engineering, consumers' behavior survey, and social networks, finding the useful information behind these data and grouping the data are important issues nowadays. Clustering is a method to classify the objects into different groups, so that the objects in each group share some common traits [15,31,57]. After this step, the data

Neng Fan
Department of Industrial and Systems Engineering, Center for Applied Optimization, University of Florida, Gainesville, FL, USA, e-mail: `andynfan@ufl.edu`

Nikita Boyko
Department of Industrial and Systems Engineering, Center for Applied Optimization, University of Florida, Gainesville, FL, USA, e-mail: `nikita@ufl.edu`

Panos M. Pardalos
Department of Industrial and Systems Engineering, Center for Applied Optimization, University of Florida, Gainesville, FL, USA, e-mail: `pardalos@ufl.edu`

is reduced to small subgroups and research on each subgroup will be easier and more direct. Clustering has been widely studied in past 20 years, and a general review of clustering is by Jain et al. in [31] while a survey of clustering algorithms is also available by Xu et al. in [57]. The future challenges in biological networks are available in the book edited by Chaovalitwongse et al. in [9].

However, clustering only does the work of objects without considering the features of each object may have. In other words, clustering compares two objects by the features that two share, without depicting the different features of the two. A method simultaneously groups the objects and features is called biclustering such that a specific group of objects has a special kind group of features. More precisely, a biclustering is to find a subset of objects and features satisfying these objects are related to features to some level. Such kind of subsets are called biclusters. Meantime, biclustering does not require objects in the same bicluster to behave similarly over all possible features, but to highly have specific features in this bicluster.

Besides the differences from clustering mentioned above, biclustering also has the abilities to find the hide features and specify them to some subsets of objects. We should also realize that biclustering also has relations but differences from other techniques, such as classification, feature selection, and outlier detection in data mining. Classification is a kind of supervised clustering while most algorithms used in biclustering are unsupervised, and for some supervised biclustering see [4, 40].

The biclustering problem is to find biclusters in data sets, and it may have different names such as co-clustering, two-mode clustering in some literatures.

6.1.2 Data Input

Usually, we call the objects as samples. Samples have different features and each sample may have or may not have some features. The level of a sample having some specific feature is called expression level. In real world, the samples may have quantitative features or qualitative features. The expression levels of quantitative features can be easily expressed in numerical data, while qualitative features have to use some scale measurement to be transformed into data. For some algorithms of biclustering, qualitative features are allowed.

Mainly, the biclustering algorithms are starting with matrices. There are two kinds of them usually used, and the first is more possible to be used in biclustering.

- Expression Matrix. This data matrix has rows corresponding to samples, columns to features, with entry measuring the expression level of a feature in a sample. Each row is called a feature vector of the sample. We can also call this matrix as sample-by-feature matrix.

  Sometimes, the matrix is formed from all samples' feature vectors, and the features' level in this sample will be observed directly. Generally we just scale and then put these vectors together to form a matrix if all vectors have the same length, which means they have the same set of features. However, the feature

vectors may not conform each other. In this case, we should add values (may be 0) to vectors with no corresponding features in order to form same-length vectors. In some applications, there are always large set of samples with limited features.

- Similarity Matrix. This data matrix has both rows and columns corresponding to a set of samples, with each entry measuring the similarity between two corresponding samples. It has same number of rows and columns, and it is symmetric. This matrix can be called sample-by-sample matrix.

  Note: this matrix can also be used as dissimilarity matrix with entry denoting the dissimilarity between a pair of samples. There are many similarity measurement functions to compute the (dis)similarity entries, such as Euclidean distance, Mahalanobis distance. So the similarity matrix can be computed from the expression matrix.

Since the developments of biclustering are including some time series models [38, 52], another kind of time series data is also used in biclustering. This data also can be viewed as stored in a matrix with that rows denote samples, while columns from left to right denote observed time points.

For some qualitative features in some cases, the data matrix is a kind of sign matrix. Some biclustering algorithms are still used.

Sometimes, before processing algorithms on the matrix, some steps are used, such as normalization, discretization, value mapping, and aggression, and the details of these data preparation operations are available at [16].

In the following, the data matrix usually refers to the first kind of expression matrix without explanation.

6.1.3 Objective of Task

Obviously, the objective of biclustering is to find biclusters in data. In clustering, the obtained clusters should have the propositions that the similarities among the samples within each cluster are maximized and the similarities between samples from different clusters are minimized.

For biclustering, the samples and features in each bicluster are highly related. But this does not mean the samples in this bicluster do not have other features, they just have the features in this bicluster more obvious and they still share other features. Thus, in each bicluster, the relations between the samples and the features are closer rather than relations between samples (features) from this bicluster and features (samples) from another bicluster.

Some biclustering algorithms allow that one sample or feature can belong to several biclusters (called overlapping) while some others produce exclusive biclusters. In addition, some algorithms have the property that each sample or feature must have its corresponding bicluster, while some others need not to be exhaustive and can allow only find one submatrix or several ones from data matrix to form the biclusters.

As we mentioned above, most of biclustering algorithms are unsupervised classification and it does not need to have any training sets. But supervised biclustering methods are also useful in some cases of biomedicine applications [5, 4, 40].

In this chapter, an optimization prospective of biclustering will be studied, and different objective functions will be used for different algorithms to satisfy part of objectives above. There is no such algorithm that can satisfy all objectives, and additionally, there is no such standard of justifying the algorithms. In distinct applications of biclustering, a specific or several objectives should be met so some algorithms are designed to satisfy these requirements. There are some methods trying to compare different algorithms, and we refer to [37, 44, 47, 61].

6.1.4 History

The first approach to biclustering is "direct clustering of data matrix" by Hartigan [28] in 1972. But the term "biclustering" was famous after Cheng and Church [11] using this technique to do gene expression analysis. After that, many biclustering algorithms are designed in different areas' applications, such as biological network, microarray data, word-document co-clustering, biomedical engineering, of which the most popular applications are in microarray data and gene expression data.

In 2004, Madeira and Oliveira [37] surveyed the biclustering algorithms for biological data analysis. In this survey, they identified the biclusters into four major classes: biclusters with constant values, with constant values on rows or columns, with coherent values, and with coherent evolutions. The biclustering structures of a data matrix are classified into nine groups according to algorithms: single bicluster, exclusive row and column biclusters, checkerboard structure, exclusive rows biclusters, exclusive columns biclusters, nonoverlapping biclusters with tree structure, nonoverlapping nonexclusive biclusters, overlapping biclusters with hierarchical structure, and arbitrarily positioned overlapping biclusters. In addition, the authors have also divided the algorithms into five classes: Iterative row and column clustering combination, divide and conquer, greedy iterative search, exhaustive bicluster enumeration, and distribution parameter identification. A comparison of these algorithms according to the above three classes is given in this survey.

Another review about biclustering algorithms is by Tanay et al. in [55] in 2004. In this survey, nine mostly used algorithms are reviewed and given with their pseudocodes. Mostly recent review of biclustering is by Busygin et al. in [5], and 16 algorithms are reviewed with their applications in biomedicine and text mining. In this chapter, the authors mentioned that "many of the approaches rely on not mathematically strict arguments and there is a lack of methods to justify the quality of the obtained biclusters."

In this chapter, we are trying to review and study the biclustering algorithms in mathematical and optimization prospectives. Not all of the algorithms will be covered, but most recent valuable algorithms are covered.

Since the development of biclustering algorithms, many softwares are designed to include several algorithms, including BicAT [2], BicOverlapper [48], BiVisu [10], toolbox by R(biclust) [32] etc. These software or packages allow to do data processing, bicluster analysis, and visualization of results and can be used directly to construct images.

In the toolbox named BicAT [2], it provides different facilities for data preparation, inspection, and postprocessing such as discretization, filtering of biclusters accordingly. Several algorithms of biclustering such as Bimax, CC, XMotifs, OPSM are included, and three methods of viewing data including matrix (heatmap), expression, and analysis are presented. The software BicOverlapper [48] is a tool for overlapping biclusters visualization. It can use three different kinds of data files of original data matrix and resulted biclusters to construct beautiful and colorful images such as heatmaps, parallel coordinates, TRN graph, bubble map, and overlapper. The BiVisu [10] is also a software tool for bicluster detection and visualization. Besides bicluster detection, BiVisu also provides functions for preprocessing, filtering, and bicluster analysis. Another software is a package written by R [32], biclust, which contains a collection of bicluster algorithms, such as Bimax, CC, plaid, spectral, xMotifs, preprocessing methods for two way data, and validation and visualization techniques for bicluster results. For individual biclustering software, there are also some packages available [55, 5].

6.1.5 Outline

In this chapter, we will follow the reviews of [37, 55, 5] and try to include the most recent algorithms and advancements of biclustering. The perspective of this chapter is of mathematical view, including linear algebra, optimization programming, bipartite graphs, probabilistic or statistical models, information theory, and time series. Section 6.1 has reviewed the motivation, data, objective, history, and softwares of biclustering. In Section 6.2, the bicluster type and biclustering structures are formally defined in a mathematical way. The most recent biclustering algorithms are reviewed in Section 6.3 and a comparison score is also defined. The application of biclustering in computational neuroscience will be reviewed in Section 6.4 and conclusions and future works are in Section 6.5.

6.2 Biclustering Types and Structures

6.2.1 Notations

As mentioned in Section 6.1.2, the expression matrix is mostly used in biclustering. Let $A = (a_{ij})_{n \times m}$ denote the sample-feature expression matrix, where there are n rows representing n samples, m columns representing m features, and the entry a_{ij}

denoting the expression level of feature j in sample i. Mostly, the matrix A is the required input of an algorithm, but some algorithms also use the space of samples or features.

Let $\mathcal{S} = \{S_1, S_2, \cdots, S_n\}$ be the sample set, where $S_i = (a_{i1}, a_{i2}, \cdots, a_{im})$ is also called the feature vector of sample i. Similarly, for the features, it is denoted by $\mathcal{F} = \{F_1, F_2, \cdots, F_m\}$ with each vector $F_j = (a_{1j}, a_{2j}, \cdots, a_{nj})^T$, a column vector. Thus, the matrix $A = (S_1, S_2, \cdots, S_n)^T = (F_1, F_2, \cdots, F_m)$.

A bicluster is a submatrix of data matrix. It is denoted by $B_k = (\mathcal{S}_k, \mathcal{F}_k)$ satisfying that $\mathcal{S}_k \subseteq \mathcal{S}$, $\mathcal{F}_k \subseteq \mathcal{F}$ and the entry denotes intersection entry with corresponding row (sample) and column in both A and B_k. Assume that there are K biclusters founded in data matrix A; the set of biclusters is denoted by $\mathcal{B} = \{B_k : k = 1, 2, \cdots, K\}$. Sometimes, we use $(\mathcal{S}_k, \mathcal{F})$ to denote a cluster of rows (samples) and use $(\mathcal{S}, \mathcal{F}_k)$ a cluster of columns (features). In some algorithms, the number of row clusters is not equal to that of column clusters. Let K, K' denote the number of row clusters, column clusters, respectively, the set of biclusters is $B = \{(\mathcal{S}_k, \mathcal{F}_{k'}) : k = 1, \cdots, K, k' = 1, \cdots, K'\}$. Without explanation, we assume that $K = K'$.

Additionally, $|\mathcal{S}_k|$ denotes the cardinality of itself, i.e., the number of samples in bicluster $B_k = (\mathcal{S}_k, \mathcal{F}_k)$ while for $|\mathcal{F}_k|$, similarly, the number of features. Clearly, $|\mathcal{S}| = n, |\mathcal{F}| = m$. In the following, the notation $i \in \mathcal{S}_k$ ($j \in \mathcal{F}_k$) is short for $S_i \in \mathcal{S}_k$ ($F_j \in \mathcal{F}_k$) without misleading.

Given a data matrix A, the biclustering problem is to design algorithms to find biclusters $\mathcal{B} = \{B_k : k = 1, 2, \cdots, K\}$ of it, i.e., a subset of matrices of A such that samples (rows, $\mathcal{S}_k$) of each bicluster B_k exhibit some similar behavior under the corresponding features (columns, $\mathcal{F}_k$). From this point, a bicluster problem now is transformed into a mathematical problem satisfying some requirements (which will be defined in the following under different bicluster types and structures). Usually, after finding biclusters in a data matrix, the rows and columns are rearranged so that the samples/features in a same bicluster will be together, the resulted matrix is called a proper rearrangement matrix. In the following discussions of bicluster types and biclustering structures, the requirements are all based on the rearrangement of data matrix.

6.2.2 Bicluster Types

The types of a bicluster is defined to be the relationships of entries within a bicluster. As mentioned in Section 1.4, Madeira and Oliveira [37] have identified bicluster types into following four major classes and here we follow their classification and give the mathematical representations. For first three cases, the data matrix A is required that $A \in R^2$, i.e., all entries in A are real numbers.

1. Bicluster with constant values. For a bicluster $B_k = (\mathcal{S}_k, \mathcal{F}_k)$, the following identity should be satisfied:

$$a_{ij} = \mu, \forall i \in \mathcal{S}_k, \forall j \in \mathcal{F}_k,$$

where μ is a constant number.

2. Bicluster with constant values on rows or columns. For a bicluster $B_k = (\mathcal{S}_k, \mathcal{F}_k)$ with constant values on rows, the identity for it is

$$a_{ij} = \mu + \alpha_i, \text{ or } a_{ij} = \mu \times \alpha_i, \forall i \in \mathcal{S}_k, \forall j \in \mathcal{F}_k,$$

where μ is a constant and α_i is an adjustment number for row i. The first identity is additive and the second one is multiplicative. Note in some data processing steps, the two are equivalent, for example, if doing logarithmic transformation on the second data matrix case. For the case of constant values on columns, the identity is

$$a_{ij} = \mu + \beta_j, \text{ or } a_{ij} = \mu \times \beta_j, \forall i \in \mathcal{S}_k, \forall j \in \mathcal{F}_k,$$

where μ is a constant and β_j is an adjustment number for column j.

3. Bicluster with coherent values. For a bicluster $B_k = (\mathcal{S}_k, \mathcal{F}_k)$ with coherent values, there are two transferable expressions. The first one is additive,

$$a_{ij} = \mu + \alpha_i + \beta_j, \forall i \in \mathcal{S}_k, \forall j \in \mathcal{F}_k,$$

and the second one is multiplicative,

$$a_{ij} = \mu \times \alpha_i \times \beta_j, \forall i \in \mathcal{S}_k, \forall j \in \mathcal{F}_k.$$

The method to transform the second into the first is still doing logarithmic transformation on the second data matrix.

4. Bicluster with coherent evolutions. In the above three cases, the data matrix $A \in R^2$. But for some cases, the algorithms are finding relationships of data on rows or columns without considering the real value. For example, in order-preserving submatrix (OPSM) algorithm, a bicluster is a group of rows whose values induce a linear order across a subset of columns. Thus, the value of a_{ij} is not always required in this situation since here the relationships between entries are considered. For other cases, the bicluster with coherent evolutions will be discussed in the following algorithms.

Although the biclusters are classified into these four classes, there are still other forms if the output bicluster was considered to reflect some relationships between the rows and columns within this bicluster. For example, in [7], a δ-valid pattern of bicluster is defined to satisfy $\max(a_{ij}) - \min(a_{ij}) < \delta, \forall j \in \mathcal{F}_k$ for row i.

Besides this, data initialization influences bicluster types, for example, row normalizing a bicluster with constant values on rows (type 2) will result a bicluster constant values (type 1). Similarly, column normalizing a bicluster with constant values on columns (type 2) will result a bicluster constant values (type 1).

6.2.3 Biclustering Structures

The structure of biclustering is defined to be the relationships between biclusters from $\mathcal{B} = \{B_k = (\mathcal{S}_k, \mathcal{F}_k) : k = 1, 2, \cdots, K\}$ based on the data matrix A.

For the structures of biclustering, there are some properties which should be noticed: exclusive, overlapping, and exhaustive, although some concepts or terms have been used previously. For a data matrix A, and the corresponding set of biclusters $\mathcal{B} = \{B_k = (\mathcal{S}_k, \mathcal{F}_k) : k = 1, 2, \cdots, K\}$, we have the following formal definitions.

- Exclusive (nonexclusive). A biclustering structure is said to be row exclusive if $\mathcal{S}_k \cap \mathcal{S}_{k'} = \emptyset$ for any $k, k' \in \{1, \cdots, K\}, k \neq k'$; to be column exclusive if $\mathcal{F}_k \cap \mathcal{F}_{k'} = \emptyset$ for any $k, k' \in \{1, \cdots, K\}, k \neq k'$; to be exclusive if it is both row exclusive and column exclusive.
- Overlapping (nonoverlapping). A biclustering structure is said to be overlapping if some entry a_{ij} belongs to two or more biclusters; otherwise, it is nonoverlapping.
- Exhaustive (nonexhaustive). A biclustering structure is said to be row exhaustive if any row S_i belongs to at least one bicluster; to be column exhaustive if any column F_j belongs to at least one bicluster; to be exhaustive if it is both row and column exhaustive. Otherwise, it is said to be nonexhaustive if some row or column does not belong to any bicluster.

Here, exclusive and overlapping are not opposite to each other, and it can found from structure 7. The following biclustering structures are based on these three properties.

Still following the classification of Madeira and Oliveira in [37], the biclustering structures are identified into following nine groups.

1. Single bicluster. In this single biclustering structure, only one submatrix is found, i.e., $k = 1$ and $\mathcal{B} = \{B_1 = (\mathcal{S}_1, \mathcal{F}_1)\}$, from A.
2. Exclusive row and column biclusters. Given a data matrix A, as Definition 1 in [5], the structure of exclusive row and column biclusters $\mathcal{B} = \{B_k = (\mathcal{S}_k, \mathcal{F}_k) : k = 1, 2, \cdots, K\}$ should satisfy the requirements as follows: For rows

$$\begin{cases} \mathcal{S}_k \subseteq \mathcal{S}, (k = 1, \cdots, K), \\ \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_K = \mathcal{S}, \\ \mathcal{S}_k \cap \mathcal{S}_{k'} = \emptyset, k, k' = 1, \cdots, K, k \neq k', \end{cases} \tag{6.1}$$

and for corresponding columns

$$\begin{cases} \mathcal{F}_k \subseteq \mathcal{F}, (k = 1, \cdots, K), \\ \mathcal{F}_1 \cup \mathcal{F}_2 \cup \cdots \cup \mathcal{F}_K = \mathcal{F}, \\ \mathcal{F}_k \cap \mathcal{F}_{k'} = \emptyset, k, k' = 1, \cdots, K, k \neq k'. \end{cases} \tag{6.2}$$

In proper rearrangement of rows and columns of data matrix A, the biclusters are the submatrices in a diagonal way without overlap between any two biclusters.

3. Checkerboard biclusters. The clusters $\{\mathcal{S}_k : k = 1, \cdots, K\}$ of samples $\mathcal{S}$ and the clusters of $\{\mathcal{F}_k : k = 1, \cdots, K\}$ of features $\mathcal{F}$ satisfy the same requirements (Equations (6.1) and (6.2)) as in structure 2. The set of checkerboard biclusters is

$$\mathcal{B} = \{B_{kk'} = (\mathcal{S}_k, \mathcal{F}_{k'}) : k, k' = 1, \cdots, K\},$$

i.e., any entry of A is in someone's biclusters.

Considering each bicluster as an entry, the proper rearrangement matrix of A is a $K \times K$ matrix with entry $B_{k,k'}$. In some cases, the number of samples' clusters $\mathcal{S}_k$s do not need to be the same as that of features' clusters $\mathcal{F}_k$s. This will imply a rectangle not a square matrix.

4. Exclusive rows biclusters. Given a data matrix A, the structure of exclusive rows' biclusters $\mathcal{B} = \{B_k = (\mathcal{S}_k, \mathcal{F}_k) : k = 1, 2, \cdots, K\}$ should satisfy the requirements as follows: For rows

$$\begin{cases} \mathcal{S}_k \subseteq \mathcal{S}, (k = 1, \cdots, K), \\ \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_K = \mathcal{S}, \\ \mathcal{S}_k \cap \mathcal{S}_{k'} = \emptyset, k, k' = 1, \cdots, K, k \neq k', \end{cases} \quad (6.3)$$

and for corresponding columns

$$\begin{cases} \mathcal{F}_k \subseteq \mathcal{F}, (k = 1, \cdots, K), \\ \mathcal{F}_1 \cup \mathcal{F}_2 \cup \cdots \cup \mathcal{F}_K = \mathcal{F}. \end{cases} \quad (6.4)$$

Comparing Equations (6.1) and (6.2) in structure 2, requirements for rows are same, but for columns, Equation (6.4) has no disjoint requirement between $\mathcal{F}_k$ and $\mathcal{F}_{k'}, k' \neq k$. In this structure, some features (columns) may belong to two or more biclusters (submatrices), while any sample (row) should belong to exactly one bicluster (submatrix).

5. Exclusive columns biclusters. Given a data matrix A, the structure of exclusive columns' biclusters $\mathcal{B} = \{B_k = (\mathcal{S}_k, \mathcal{F}_k) : k = 1, 2, \cdots, K\}$ should satisfy the requirements as follows: For rows

$$\begin{cases} \mathcal{S}_k \subseteq \mathcal{S}, (k = 1, \cdots, K), \\ \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_K = \mathcal{S}, \end{cases} \quad (6.5)$$

and for corresponding columns

$$\begin{cases} \mathcal{F}_k \subseteq \mathcal{F}, (k = 1, \cdots, K), \\ \mathcal{F}_1 \cup \mathcal{F}_2 \cup \cdots \cup \mathcal{F}_K = \mathcal{F}, \\ \mathcal{F}_k \cap \mathcal{F}_{k'} = \emptyset, k, k' = 1, \cdots, K, k \neq k'. \end{cases} \quad (6.6)$$

Comparing Equations (6.1) and (6.2) in structure 2, requirements for columns are same, but for rows, Equation (6.5) has no disjoint requirement between $\mathcal{S}_k$ and $\mathcal{S}_{k'}, k' \neq k$. In this structure, some samples (rows) may belong to two or more biclusters (submatrices), while any feature (column) should belong to exactly one bicluster (submatrix).

6. Nonoverlapping with tree-structured biclusters. For a data matrix A, nonoverlapping means no entry can belong to more than one bicluster. Thus some entries may not belong to any bicluster. Tree structure means in the proper rearrangement matrix, the blocks of submatrices (biclusters) are not crossing each other.
7. Nonoverlapping nonexclusive biclusters. Nonoverlapping is same as above. Nonexclusive means a sample or feature can belong to more than one biclusters, and a sample can belong to two sets of important features in two biclusters, and vice versa.
8. Nonoverlapping hierarchically structured biclusters. Nonoverlapping is same as above. Hierarchically structured means a bicluster may belong to some other "bigger" biclusters, i.e., in the set of biclusters $\mathcal{B} = \{B_k = (\mathcal{S}_k, \mathcal{F}_k) : k = 1, 2, \cdots, K\}$ of data matrix A, there exists some biclusters $B_k = (\mathcal{S}_k, \mathcal{F}_k)$ and $B_{k'} = (\mathcal{S}_{k'}, \mathcal{F}_{k'})$ such that $\mathcal{S}_k \subseteq \mathcal{S}_{k'}$ or $\mathcal{F}_k \subseteq \mathcal{F}_{k'}$.
9. Arbitrary positioned overlapping biclusters. In the set of biclusters $\mathcal{B} = \{B_k = (\mathcal{S}_k, \mathcal{F}_k) : k = 1, 2, \cdots, K\}$ of data matrix A, there exists some entry a_{ij} such that $a_{ij} \in B_k$ and $a_{ij} \in B_{k'}$ with $k \neq k'$. In the meantime, biclusters $B_k, B_{k'}$ may share some common samples or features.

To check the nine biclustering structures, and according to above definitions of exclusive and exhaustive, structures 1, 2 are exclusive; structure 3 is nonoverlapping; structure 1 is nonexhaustive; structures 2, 3, 4, and 5 are exhaustive; and the properties for some other structures can be found from its classification. Note that these structures are not always strict. For example, structures 2, 3, 4, and 5 also have nonexclusive versions (which will not satisfy above formal requirements), and for details we refer to [37].

6.3 Biclustering Techniques and Algorithms

In this section, the biclustering techniques and algorithms are divided into several class based on the methods used for different areas of mathematics, probability, or other optimization methods. Here we are concentrating on mathematical backgrounds.

6.3.1 Based on Matrix Means and Residues

For a bicluster $B_k = (\mathcal{S}_k, \mathcal{F}_k)$, several means based on the bicluster are defined. The mean of row i of B_k is

$$\mu_{ik}^{(r)} = \frac{1}{|\mathcal{F}_k|} \sum_{j \in \mathcal{F}_k} a_{ij}, \tag{6.7}$$

the mean of column j of B_k is

$$\mu_{jk}^{(c)} = \frac{1}{|\mathcal{S}_k|} \sum_{i \in \mathcal{S}_k} a_{ij}, \tag{6.8}$$

and the mean of all the entries in B_k is

$$\mu_k = \frac{\sum_{i \in \mathcal{S}_k} \sum_{j \in \mathcal{F}_k} a_{ij}}{|\mathcal{F}_k||\mathcal{S}_k|}. \tag{6.9}$$

The residue of the entry a_{ij} in bicluster B_k is

$$r_{ij} = a_{ij} - \mu_{ik}^{(r)} - \mu_{jk}^{(c)} + \mu_k, \tag{6.10}$$

the variance of bicluster B_k is

$$\mathrm{Var}(B_k) = \sum_{i \in \mathcal{S}_k} \sum_{j \in \mathcal{F}_k} (a_{ij} - \mu_k)^2, \tag{6.11}$$

and mean squared residue of the bicluster B_k is

$$H_k = \frac{\sum_{i \in \mathcal{S}_k} \sum_{j \in \mathcal{F}_k} r_{ij}^2}{|\mathcal{F}_k||\mathcal{S}_k|}. \tag{6.12}$$

The first approach of biclustering by Hartigan [28] is known as *block clustering*, with the objective function as

$$\min \mathrm{Var}(\mathcal{B}) = \sum_{k=1}^{K} \mathrm{Var}(B_k) = \sum_{k=1}^{K} \sum_{i \in \mathcal{S}_k} \sum_{j \in \mathcal{F}_k} (a_{ij} - \mu_k)^2,$$

where the number of biclusters is a given number. For each bicluster, the variance $\mathrm{Var}(B_k)$ is 0 if it is constant.

CC. Cheng and Church's Algorithm (CC) [11] defines a bicluster to be a submatrix for which the mean squared residue score is below a user-defined threshold δ, i.e., $H_k \leq \delta$, where δ represents the minimum possible value. To find the largest bicluster in A, they propose a two-phase strategy: removing rows and columns and then adding the removed rows and columns with some rules. First, the row to be removed is the one

$$\arg\max_i \frac{1}{|\mathcal{F}_k|} \sum_{j \in \mathcal{F}_k} r_{ij}^2,$$

and column is

$$\arg\max_j \frac{1}{|\mathcal{S}_k|} \sum_{j \in \mathcal{S}_k} r_{ij}^2.$$

Repeating these removing steps until the bicluster with $H_k \leq \delta$ obtained. Then some previously removed rows and columns can be added without violating the requirement of $H_k \leq \delta$. Yang et al. [58,59] proposed an improved version of this algorithm which allows missing data entry of A with a heuristic flexible overlapped clustering (FLOC) algorithm.

RWC. Angiulli et al. [1] proposed a random walk biclustering algorithm (RWC) based on a greedy technique and enriched with a local search strategy to escape poor local minima. The algorithm starts with an initial random bicluster B_k and searches for a δ-bicluster by successive transformations of B_k, until a gain function is improved. The transformations consist in the change of membership (called flip or move) of the row/column that leads to the largest increase of the gain function. If a bit is set from 0 to 1 it means that the corresponding sample or feature, which was not included in the bicluster B_k, is added to B_k. Vice versa, if a bit is set from 1 to 0 it means that the corresponding sample or feature is removed from the bicluster.

The gain function combines mean squared residue, row variance, and size of the bicluster by means of user-provided weights w_{res}, w_{var}, and $w_{vol}(w_{res} + w_{var} + w_{vol} = 1, 0 \leq w_{res}, w_{var}, w_{vol} \leq 1)$. The gain function is defined as

$$\text{gain} = w_{res}(2^{\Delta res} - 1) - w_{var}(2^{\Delta var} - 1) - w_{vol}(2^{\Delta vol} - 1),$$

where $\Delta res, \Delta var, \Delta vol$ are relative changes of mean squared residue, row variance, and size between a new bicluster and an old bicluster, respectively. This function assumes values in the interval $[-1, 1]$. Decreasing w_{res} and increasing w_{var} and w_{vol}, biclusters with higher row variance and larger size can be obtained.

6.3.2 Based on Matrix Ordering, Reordering, and Decomposition

The following several biclustering algorithms are based on matrix reordering or decomposition.

OPSM. Ben-Dor et al. [3] proposed order-preserving submatrix algorithm (OPSM) for biclustering. A bicluster is defined as a submatrix that preserves the order of the selected columns for all of the selected rows. In other words, the expression values of the samples within a bicluster induce an identical linear ordering across the selected features. Based on a stochastic model, the authors [3] developed a deterministic algorithm to find large and statistically significant biclusters. This concept has been taken up in a recent study by Liu and Wang [36] as OP-cluster.

ISA. Ihmels et al. [30] proposed the iterative signature algorithm (ISA) for biclustering. Given the data matrix A, the two matrices A^s, A^f are obtained by normalizing A such that $\sum_i a_{ij}^s = 0, \sum_i (a_{ij}^s)^2 = 1$ (mean, variance) for each feature F_j and similarly for sample S_i, $\sum_j a_{ij}^f = 0, \sum_j (a_{ij}^f)^2 = 1$.

Starting with an initial set of samples, all features are scored with respect to this sample set and those features are chosen for which the score exceeds a predefined threshold. In the same way, all samples are scored regarding the selected features and a new set of samples is selected based on another predefined threshold. The entire procedure is repeated until the set of samples and the set of features do not change anymore. Multiple biclusters can be identified by running the iterative signature algorithm on several initial sample sets.

xMotif. In the framework proposed by Murali and Kasif [39], biclusters are defined such that samples are nearly constantly expressed across the selection of features. In first step, the input matrix is preprocessed by assigning each sample a set of statistically significant states. These states define the set of valid biclusters: A bicluster is a submatrix where each sample is exactly in the same state for all selected features. To identify the largest valid biclusters, an iterative search method is proposed that is run on different random seeds, similarly to ISA.

OREO. DiMaggio Jr. et al. [19] proposed an algorithm of optimal re-ordering (OREO) of the rows and columns of the data matrix A to biclustering. The idea of OREO is to optimally rearrange the rows and columns of data matrix A to minimize the similarities between rows and columns in the rearranged matrix. The algorithm has three main iterative steps: optimally re-ordering rows (or columns) of the data matrix; computing the median for each pair of neighboring rows (or columns) in the final rearranged matrix, sorting these values from highest to lowest and classifying cluster boundaries between the rows (or columns) to obtain submatrices; and optimally re-ordering the columns (or rows) of each submatrix and computing the cluster boundaries for the re-ordered columns (or rows) analogous to the second step.

Here we use rows to reorder, and the authors [19] defined three associated cost measurement functions between row i and row i':

$$c_{ii'} = \sum_{j=1}^{m} |a_{ij} - a_{i'j}|, \sum_{j=1}^{m} (a_{ij} - a_{i'j})^2, \sqrt{\frac{\sum_j (a_{ij} - a_{i'j})^2}{m}}.$$

The authors [19] use two models to reorder rows in order to minimize the total similarities between rows of final rearranged matrix: the network flow model and TSP model, which are ideas from network optimization. In the network flow model, defining the binary variables

$$y_{ii'}^{\text{row}} = \begin{cases} 1, & \text{if row } i \text{ is adjacent and above } i' \text{ in the final ordering;} \\ 0, & \text{otherwise,} \end{cases}$$

and two additional ones for the topmost and bottommost rows

$$y\_\text{source}_i^{\text{row}} = \begin{cases} 1, & \text{if row } i \text{ is the topmost row in the final ordering;} \\ 0, & \text{otherwise,} \end{cases}$$

$$y\_\text{sink}_i^{\text{row}} = \begin{cases} 1, & \text{if row } i \text{ is the bottommost row in the final ordering;} \\ 0, & \text{otherwise,} \end{cases}$$

and choosing one of the three associated cost measurement functions, the optimization problem is to find solution to binary variables $y_{ii'}^{\text{row}}, y\_\text{source}_i^{\text{row}}, y\_\text{sink}_i^{\text{row}}$,

$$\min \quad \sum_i \sum_{i'} c_{ii'} y_{ii'}^{\text{row}}$$

$$s.t. \quad \sum_{i \neq i'} y_{ii'}^{\text{row}} + y\_source_i^{\text{row}} = 1 \quad \forall i$$

$$\sum_{i' \neq i} y_{ii'}^{\text{row}} + y\_sink_i^{\text{row}} = 1 \quad \forall i$$

$$\sum_i y\_source_i^{\text{row}} = 1$$

$$\sum_i y\_sink_i^{\text{row}} = 1$$

$$f\_source_i^{\text{row}} = n \cdot y\_source_i^{\text{row}} \quad \forall i$$

$$\sum_{i'} (f_{i'i}^{\text{row}} - f_{ii'}^{\text{row}}) + f\_source_i^{\text{row}} - f\_sink_i^{\text{row}} = 1 \quad \forall i$$

$$f_{ii'}^{\text{row}} \leq (n-1) \cdot y_{ii'}^{\text{row}} \quad \forall (i, i')$$

$$f_{ii'}^{\text{row}} \geq y_{ii'}^{\text{row}} \quad \forall (i, i')$$

$$y_{ii'}^{\text{row}}, y\_source_i^{\text{row}}, y\_sink_i^{\text{row}} \in \{0, 1\}.$$

In the TSP model, the variables are the same as network flow model except including variables $y\_source_i^{\text{row}}, y\_sink_i^{\text{row}}$, and the optimization problem is

$$\min \quad \sum_i \sum_{i'} c_{ii'} y_{ii'}^{\text{row}}$$

$$s.t. \quad \sum_{i'} y_{ii'}^{\text{row}} = 1 \quad \forall i$$

$$\sum_{i'} y_{i'i}^{\text{row}} = 1 \quad \forall i$$

$$y_{ii'}^{\text{row}} \in \{0, 1\}.$$

The two optimization problems induced by the models are mixed integer linear programming and can be solved by CPLEX [14].

After reordering the rows of data matrix, for rows i and $i + 1$ in the final rearranged matrix, the median of each pairwise term of the objective function $\phi(a_{i,j}, a_{i+1,j})$ is computed by $\text{MEDIAN}_j \phi(a_{i,j}, a_{i+1,j})$. In [19], top 10% of largest median values are suggested to be boundaries between re-ordered rows.

nsNMF. Pascual-Montano et al. [43] and Carmona-Saez et al. [8] proposed a biclustering algorithm based on nonsmooth nonnegative matrix factorization (nsNMF). The method nsNMF approximates the data matrix A as a product of two submatrices, W and H. Rows of H constitute basis samples, while columns of W are basis features. Coefficients in each pair of basis samples and features are used to sort features and samples in the original matrix, respectively. The biclusters are the submatrices of the sorted matrix.

Originally, the nonnegative matrix factorization is used to analyze facial images [35]. The nonnegative matrix factorization (NMF) is to decompose matrix $A = (a_{ij})_{n \times m}$ into two matrices, i.e.,

$$A \approx WH,$$

where $W = (w_{ia})_{n \times k}$ are the reduced k ($k \leq m$) basis vectors (factors), and $H = (h_{aj})_{k \times m}$ contains the coefficients of the linear combinations of the basis vectors (encoding vectors). All matrices A, W, H are nonnegative and the columns of W are normalized. Thus, the entry a_{ij} can be expressed as

$$a_{ij} \approx (WH)_{ij} = \sum_{a=1}^{k} w_{ia} h_{aj}.$$

Based on Poisson likelihood, the objective function of this factorization is to minimize the divergence function, i.e.,

$$\min D(A, WH) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(a_{ij} \log \frac{a_{ij}}{(WH)_{ij}} - a_{ij} + (WH)_{ij} \right).$$

The solution to this objective function of finding W, H uses an iterative algorithm with random number initialization [8].

The nsNMF method, which will [8] "produce more compact and localized feature representation of the data than standard NMF" of finding sparse structures in data matrix, is an improvement of NMF. The nsNMF method introduces a smooth distribution of the factors to get sparseness, and the decomposition of data matrix A is

$$A \approx WSH,$$

where the matrix $S = (1 - \theta)I + \theta \frac{ee^T}{k}$ is a positive smothness matrix, I is the identity matrix, e is a row vector of k 1s, and θ controls the sparseness of the model, satisfying $0 \leq \theta \leq 1$. And now the objective function for nsNMF method is

$$\min D(A, WSH) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(a_{ij} \log \frac{a_{ij}}{(WSH)_{ij}} - a_{ij} + (WSH)_{ij} \right).$$

When $\theta = 0$, the nsNMF backs to NMF; when $\theta \to 1$, the vector SX (X is a positive nonzero vector) tends to the constant with all elements almost equal to the average of the elements of X and all entries are equal to the same nonzero value, which is the smoothest possible vector, in the sense of "nonsparseness." The algorithm to solve this objective function can be done as the same way of previous function with small changes [8].

Bimax. Prelic et al. [44] presented a fast-and-conquer approach, binary inclusion-maximal biclustering algorithm (Bimax). This algorithm assumes that the data matrix A is binary with $a_{ij} \in \{0, 1\}$ where an entry 1 means feature j is important in sample i.

In this algorithm, a named inclusion-maximal bicluster is defined to be $B_k = (\mathcal{S}_k, \mathcal{F}_k)$ such that $a_{ij} = 1$ for any $i \in \mathcal{S}_k, j \in \mathcal{F}_k$, and there does not exist another

bicluster $B_{k'} = (\mathcal{S}_{k'}, \mathcal{F}_{k'})$ of A with $a_{ij} = 1$ for any entry in $B_{k'}$ and $\mathcal{S}_k \subseteq \mathcal{S}_{k'}, \mathcal{F}_k \subseteq \mathcal{F}_{k'}, (\mathcal{S}_k, \mathcal{F}_k) \neq (\mathcal{S}_{k'}, \mathcal{F}_{k'})$.

The Bimax algorithm is to find such inclusion-maximal bicluster of A, which is different from the SAMBA, where 0 entry can be contained in a bicluster. More specifically, the idea behind the Bimax algorithm is to partition A into three submatrices, one of which contains only 0-cells. Therefore, it can be disregarded in the following. The algorithm is then recursively applied to the remaining two submatrices U and V; the recursion ends if the current matrix represents a bicluster, i.e., contains only 1s. If U and V do not share any rows and columns of A, the two matrices can be processed independently from each other. If U and V have a set of rows in common as shown, special care is necessary to only generate those biclusters in V that share at least one common column.

6.3.3 Based on Bipartite Graphs

The following two algorithms are based on bipartite graphs since there is a close relationship between expression matrix of samples and features and weighted bipartite graph.

A bipartite graph is defined as a graph $G = (U, V, E)$, where U, V are two disjoint sets of vertices, and E is the set of edges between vertices from U and V, while no edge appears between any two vertices from U or V.

In order to do biclustering problem, the data matrix A can be transformed into a bipartite graph where each vertex in one set U denotes a sample while vertex from another set V denotes a feature. The expression level a_{ij} between samples and features is denoted by the weighted edges $(u_i, v_j) \in E$ between vertices $u_i \in U$ and $v_j \in V$ with weight $w_{ij} = a_{ij}$. A bicluster corresponds to a subgraph $H_k = (U_k, V_k, E_k)$ of $G = (U, V, E)$ where $U_k \subseteq U, V_k \subseteq V$ and $E_k \subseteq E$ and edges in E_k induced by vertices from U_k, V_k. Thus, the set $(\mathcal{S}, \mathcal{F}, A)$ is corresponding to bipartite graph $G = (U, V, E)$ and the bicluster $B_k = (\mathcal{S}_k, \mathcal{F}_k)$ is to subgraph $H_k = (U_k, V_k, E_k)$. Sometimes, we may only consider one subgraph of G and denote it as $H = (U', V', E')$. Clearly, here $|U| = n, |V| = m$.

Spectral biclustering. The first algorithm of biclustering based on bipartite graph is called spectral biclustering, proposed by Dhillon [17]. Since this biclustering algorithm has some close relationships, which will be shown later, with spectral graph theory [13], it got its name spectral biclustering. Before presenting this algorithm, several matrices are based on A and bipartite graph $G = (U, V, E)$ with edges' weight $w_{ij} = a_{ij}$.

The adjacency weighted matrix of the bipartite graph $G = (U, V, E)$ is expressed in the form of data matrix A as

$$W = (w_{ij})_{(n+m) \times (n+m)} = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix},$$

and the weighted degree d_i of vertex u_i is defined as $d_i = \sum_{j:(i,j) \in E} w_{ij}$, and the degree matrix $D_u = (d_{ij})_{n \times n}$ of the graph is a diagonal matrix as

$$d_{ij} = \begin{cases} d_i, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, we can get the degree matrix D_v. The degree matrix of the bipartite graph $G = (U, V, E)$ is

$$D = \begin{pmatrix} D_u & 0 \\ 0 & D_v \end{pmatrix},$$

where the diagonal elements of D_u and D_v are weighted degree of vertices belonging to U and V, and all other elements are 0. The Laplacian matrix of the bipartite graph $G = (U, V, E)$ for data set A is defined as

$$L = D - W = \begin{pmatrix} D_u & -A \\ -A^T & D_v \end{pmatrix}.$$

The production of spectral clustering is exclusive row and column biclusters. Therefore, the corresponding subgraphs H_k of G are disjoint with each other. The weight of edges between such subgraphs is defined as cut. Without loss of generality, assume there are two subgraphs $H_1 = (U_1, V_1, E_1)$ and $H_2 = (U_2, V_2, E_2)$ such that $U_1 \cup U_2 = U, U_1 \cap U_1 = \emptyset, V_1 \cup V_2 = V, V_1 \cap V_2 = \emptyset$, and $E_i \subseteq E$ induced all edges between U_i and V_i. Subgraphs $H_1 = (U_1, V_1, E_1)$ and $H_2 = (U_2, V_2, E_2)$ are called a partition of G. The cut of such partition of bipartite graph is the sum of weights of edges between U_1, V_2 and U_2, V_1, i.e.,

$$\text{cut}(H_1, H_2) = \sum_{\substack{i \in U_1, j \in V_2, (i,j) \in E \\ \text{and } i \in U_2, j \in V_1, (i,j) \in E}} w_{ij}.$$

Obviously, the objective of biclustering is to minimize such intersimilarities between biclusters (subgraphs). At the same time, the similarities within each bicluster should be maximized. The intrasimilarity of bicluster(subgraph) is defined as $\sum_k$. In order to balance the intersimilarities and intrasimilarities of biclusters, several different cuts are defined, such as ratio cut [27, 17, 33], normalized cut [51, 33], minimax cut [60], ICA cut [45]. The most popularly used are ratio cut and normalized cut.

For a partition $H_1 = (U_1, V_1, E_1), H_2 = (U_2, V_2, E_2)$ of the bipartite graph $G = (U, V, E)$, the ratio cut is defined as

$$\frac{\text{cut}(H_1, H_2)}{|U_1 \cup V_1|} + \frac{\text{cut}(H_2, H_1)}{|U_2 \cup V_2|},$$

and the normalized cut is defined as

$$\frac{\text{cut}(H_1, H_2)}{d_{p_1}} + \frac{\text{cut}(H_2, H_1)}{d_{p_2}},$$

where $d_{p_1} = \sum_{i \in (U_1 \cup V_1)} d_i, d_{p_2} = \sum_{j \in (U_2 \cup V_2)} d_j$.

Define the indicator vector as

$$y_i = \begin{cases} \sqrt{(n_2 + m_2)/((n_1 + m_1)(n + m))}, & i \in U_1 \cup V_1, \\ -\sqrt{(n_1 + m_1)/((n_2 + m_2)(n + m))}, & i \in U_2 \cup V_2, \end{cases}$$

where $|U_1| = n_1, |U_2| = n_2, |V_1| = m_1, |V_2| = m_2$, the objective of minimizing the ratio cut of partition $H_1 = (U_1, V_1, E_1), H_2 = (U_2, V_2, E_2)$ can be expressed as

$$\min \quad y^T L y,$$
$$s.t. \quad y^T y = 1, y^T e = 0.$$

Relax y to any real number, the solution is the eigenvector corresponding to the second smallest eigenvalue of L [13, 17]. Thus, after obtaining the indicator for each vertex of U, V, the corresponding subgraphs can be easily transformed back into biclusters. Similarly, for normalized cut, define the indicator vector as

$$y_i = \begin{cases} \sqrt{d_{U_2 \cup V_2}/(d_{U_1 \cup V_1} d)}, & i \in U_1 \cup V_1, \\ -\sqrt{d_{U_1 \cup V_1}/(d_{U_2 \cup V_2} d)}, & i \in U_2 \cup V_2, \end{cases}$$

where $d_{U_1 \cup V_1} = \sum_{i \in U_1 \cup V_1} d_i, d_{U_2 \cup V_2} = \sum_{j \in U_2 \cup V_2} d_j$, the objective of minimizing the normalized cut of partition $H_1 = (U_1, V_1, E_1), H_2 = (U_2, V_2, E_2)$ can be expressed as

$$\min \quad y^T L y, \qquad \qquad (6.13)$$
$$s.t. \quad y^T D y = 1, y^T D e = 0. \qquad \qquad (6.14)$$

Now the solution of this programming is the eigenvector corresponding to the generalized eigenvalue problem $Ly = \lambda Dy$ [51]. The above programming problems can be also modeled to mixed integer programming.

For large data matrix A, the solution of its eigenvector problem is very difficult and a method proposed by [17]. For more details of spectral biclustering, see [22]. In above, only two biclusters are obtained instead of K ones. For K biclusters, Dhillon [17] used k-means algorithm [31, 57] after obtaining the indicator vector y, and another direct approach is from [23] by defining an indicator matrix.

SAMBA. Tanay et al. [54] presented a statistical algorithmic method for bicluster analysis (SAMBA) based on bipartite graph and probabilistic modeling. Under a bipartite graph model, the weight of each edge is assigned according to a probabilistic model, thus, to find biclusters of A become to find heavy subgraphs of G with high likelihood. This method is motivated by finding the complete bipartite subgraph(biclique) of G. The idea of SAMBA has three steps: forming the bipartite graph and calculating weights of edges and nonedges (two models introduced in this step: a simple model and a refined model); applying a hashing technique to find

heaviest bicliques(biclusters) in the graph; and performing a local improvement procedure on the biclusters in each heap.

Given a data matrix A, the corresponding bipartite graph is $G = (U, V, E)$. A bicluster corresponds to a subgraph $H = (U', V', E')$ as introduced above. The weight of a subgraph is the sum of the assigned weights of edges $(u, v) \in E'$ and nonedges $(u, v) \in \bar{E}' = (U' \times V') \setminus E'$. The subgraph with assigned weights has its statistical significance and finding a bicluster is to search heavy subgraph with respect to the weight of subgraph. There are two models introduced in [54]: a simple model and a refined model.

In the simple model, let $|E| = k$, $p = k/mn$ and assume that edges occur independently and equiprobability with density p. Let $BT(k, p, n)$, binomial distribution, be the probability of observing k or more success occurs independently with p, the probability of observing a graph at least as dense as H is $p(H) = BT(k', p, n'm')$, where k', n', m' are corresponding notations in $H = (U', V', E')$. Finding a maximum weight subgraph of G is equivalent of finding a subgraph H with lowest $p(H)$. In the refined model, each edge (u, v) is an independent Bernoulli variable $p_{u,v}$, which is fraction of bipartite graphs with degree sequence identical to G that contains edge (u, v). The probability of observing H is

$$p(H) = \left(\prod_{(u,v) \in E'} p_{u,v} \right) \left(\prod_{(u,v) \in \bar{E}'} (1 - p_{u,v}) \right).$$

In practice, a likelihood ratio is chosen, i.e.,

$$\log L(H) = \sum_{(u,v) \in E'} \log \frac{p_c}{p_{u,v}} + \sum_{(u,v) \in \bar{E}'} \log \frac{1 - p_c}{1 - p_{u,v}},$$

where $p_c \geq \max_{(u,v) \in U \times V} p_{u,v}$, which corresponds to the weight of subgraph H with weight $\log \frac{p_c}{p_{u,v}} > 0$ of each edge (u, v) and $\log \frac{1-p_c}{1-p_{u,v}} < 0$ for each nonedge (u, v). Then a hash technique is applied to solve the maximum biclique problem in order to find the heavy subgraphs (biclusters). The final step of local improvement iteratively applies the best modification to the bicluster.

In a recent study of Tanay et al. [53], this SAMBA has been extended to integrate multiple types of experimental data.

6.3.4 Based on Information Theory

In [18], Dhillon et al. proposed a biclustering algorithm based on information theory. This information theoretic biclustering algorithm that simultaneously clusters both the rows and the columns is called co-clustering by Dhillon et al.

By proper transformation, the data matrix A is to be a joint probability distribution matrix $p(\mathcal{S}, \mathcal{F})$ between two discrete random variables $\mathcal{S}, \mathcal{F}$. Let K be the number of disjoint clusters of samples and K' the number of disjoint features. The set of biclusters is $\mathcal{B} = (\mathcal{S}', \mathcal{F}') = (\{\mathcal{S}_k, : k = 1, \cdots, K\}, \{\mathcal{F}_{k'} : k' = 1, \cdots, K'\})$. The mappings of C_S, C_F are objectives to find in this biclustering algorithm such that

$$C_S : \{S_1, S_2, \cdots, S_n\} \rightarrow \{\mathcal{S}_1, \cdots, \mathcal{S}_K\},$$

$$C_F : \{F_1, F_2, \cdots, F_m\} \rightarrow \{\mathcal{F}_1, \cdots, \mathcal{F}_{K'}\}.$$

The mutual information $I(\mathcal{S}, \mathcal{F})$ of two random variables $\mathcal{S}, \mathcal{F}$ is the amount of information shared between these two variables and is defined as in information theory

$$I(\mathcal{S}, \mathcal{F}) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(S_i, F_j) \log \frac{p(S_i, F_j)}{p(S_i) p(F_j)} = D(p(S, F) \| p(S) p(F)),$$

where $p(S_i, F_j), p(S_i), p(F_j)$ are probabilities from distribution matrix $p(\mathcal{S}, \mathcal{F})$, and $D(p_1 \| p_2) = \sum_x p_1(x) \log \frac{p_1(x)}{p_2(x)}$ is the relative entropy between two probability distributions $p_1(x)$ and $p_2(x)$.

The objective of this biclustering is to find optimal biclusters of A such that the loss in mutual information is minimized, i.e.,

$$\min I(\mathcal{S}, \mathcal{F}) - I(\mathcal{S}', \mathcal{F}').$$

In order to solve this objective function, $q(x, y) = p(x', y') p(x', y') p(x|x') p(y|y')$ is defined so that the objective function can be written as

$$\min I(\mathcal{S}, \mathcal{F}) - I(\mathcal{S}', \mathcal{F}') = D(p(\mathcal{S}, \mathcal{F}) \| q(\mathcal{S}, \mathcal{F})).$$

For proof of this result, we refer to [18]. Then an iterative way is used to solve by transformed the objective function [18].

6.3.5 Based on Probability

The following two biclustering algorithms (named as BBC and cMonkey) use the theory of probability.

BBC. Gu and Liu [26] proposed a Bayesian biclustering model (BBC) and implemented a Gibbs sampling [34] procedure for its statistical inference. This model can also consider an implementation of plain model [50] of biclustering.

Given data matrix A, assume the entry

$$a_{ij} = \sum_{k=1}^{K} ((\mu_k + \alpha_{ik} + \beta_{jk} + \varepsilon_{ijk}) \delta_{ik} \kappa_{jk}) + e_{ij} \left(1 - \sum_{k=1}^{K} \delta_{ik} \kappa_{jk}\right),$$

where μ_k is the main effect of bicluster k, and α_{ik} and β_{jk} are the effects of sample i and feature j, respectively, in bicluster k, ε_{ijk} is the noise term for bicluster k, and e_{ij} models the data points that do not belong to any bicluster. Here δ_{ik}, κ_{jk} are binary variables: $\delta_{ik} = 1$ indicates that row i belongs to bicluster k, and $\delta_{ik} = 0$ otherwise; similarly, $\kappa_{jk} = 1$ indicates that column j is in cluster k, and $\kappa_{jk} = 0$ otherwise. In plain model [50], the entry a_{ij} has similar assumption with less factors to be considered.

In nonoverlapping feature biclustering, $\sum_{k=1}^{K} \kappa_{jk} \leq 1$, and in nonoverlapping sample biclustering, $\sum_{k=1}^{K} \delta_{jk} \leq 1$. Here, nonoverlapping sample is discussed. The priors of the indicators κ and δ are set so that a feature can be in multiple biclusters while sample is at more than one.

In this model, an observation a_{ij} can belong to either one or none of the biclusters, and the probability distribution of a_{ij} conditional on the bicluster indicators can be rewritten as

$$a_{ij}|\delta_{ik} = 1, \kappa_{jk} = 1 \sim N(\mu_k + \alpha_{ik} + \beta_{jk}, \sigma_{\varepsilon k}^2)$$

if a_{ij} belongs to bicluster k; otherwise,

$$a_{ij}|\delta_{ik}\kappa_{jk} = 0 \text{ for all } k \sim N(0, \sigma_e^2).$$

With Gaussian zero-mean priors on the effect parameters, the marginal distribution of the a_{ij} conditional on the indicators is

$$\mathcal{B}|\delta, \kappa \sim N(0, \Sigma),$$

where Σ is the covariance of matrix of $\mathcal{B}$ and $\mathcal{B} = \{B_0, B_1, B_2, \cdots, B_K\}^T$ with $B_k = \{a_{ij} : \delta_{ik}\kappa_{jk} = 1\}, k \geq 1$ and B_0 being the vector of data points belonging to no bicluster. More specifically, Σ is a sparse matrix of the form

$$\Sigma = \begin{pmatrix} \sigma_e^2 I & 0 & \cdots & 0 \\ 0 & \Sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_K \end{pmatrix},$$

where $\Sigma_k = \text{Cov}(B_k, B_k)$ is the covariance matrix of all data points belonging to cluster k.

To make inference form above BBC model, the implemented Gibbs sampling method is used. Initializing from a set of randomly assigned values of δ's and κ's, the column indicators κ are sampled by calculating the log-probability ratio

$$\log \frac{P(V_2|\kappa_{jk} = 1, \sigma_{\mu k}^2, \sigma_{\alpha k}^2, \sigma_{\beta k}^2, \sigma_{\varepsilon k}^2, \sigma_e^2)P(\kappa_{jk} = 1)}{P(V_2|\kappa_{jk} = 0, \sigma_{\mu k}^2, \sigma_{\alpha k}^2, \sigma_{\beta k}^2, \sigma_{\varepsilon k}^2, \sigma_e^2)P(\kappa_{jk} = 0)},$$

where $V_1 = \{a_{il} : \delta_{ik} = 0 \text{ or } \kappa_{lk} = 0, l \neq j\}$, the set contains data points not in cluster k, and $V_2 = \{a_{il} : \delta_{ik} = 1, \kappa_{lk} = 1, l \neq j\} \cup \{a_{ij} : \delta_{ik} = 1\}$, the set contains data points that are or can in bicluster k. This notation follows that in [26].

In order to calculate the likelihood term in the above ratio, we need to take the inverse and determinant of the covariance matrices for the vector V_2 in both cases. For details of rest of BBC algorithm, we refer to [26].

cMonkey. Reiss et al. [46] proposed an integrated biclustering algorithm (named cMonkey used in heterogeneous genome-wide data sets for the inference of global regulatory networks. In this model, each bicluster is modeled via a Markov chain process, in which the bicluster is iteratively optimized, and its state is updated based upon conditional probability distributions computed using the cluster's previous state. Three major distinct data types are used (gene expression, upstream sequences, and association networks), and accordingly p-values for three such model components are computed: the expression component, the sequence component, and the network component. Here we only reviewed the expression component.

Given the expression data matrix A, the variance in the measured levels of feature j is $\sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n}(a_{ij} - \bar{a}_j)^2$, where $\bar{a}_j = \sum_{i=1}^{n} a_{ij}/n$. The mean expression level of feature j over the bicluster's samples $\mathcal{S}_k$ is $\bar{a_{jk}} = \mu_{ik}^{(r)}$ as defined previously. As defined in [46] the likelihood of an arbitrary measurement a_{ij} relative to this mean expression level is

$$p(a_{ij}) = \frac{1}{\sqrt{2\pi(\sigma_j^2 + \varepsilon^2)}} \exp\left[-\frac{(a_{ij} - \bar{a_{jk}})^2 + \varepsilon^2}{2(\sigma_j^2 + \varepsilon^2)}\right],$$

where ε for an unknown systematic error in condition j, here assumed to be the same for all j. The likelihood of the measurements of an arbitrary sample i among the conditions in bicluster k is $p(S_i) = \prod_{j \in \mathcal{F}_k} p(a_{ij})$, and similarly the likelihood of a feature j's measurements is $p(F_j) = \prod_{i \in \mathcal{S}_k} p(a_{ij})$.

Before the following iterative steps, the Markov chain process by which a bicluster is optimized requires "seeding" of the bicluster to start the procedure. The iterative steps include searching for motifs in bicluster, computing conditional probability that each sample/feature is a member of the bicluster, and performing moves sampled from the conditional probability.

6.3.6 Comparison of Biclustering Algorithms

Since the biclustering algorithms are designed based on different bases and used in different data, and the requirements are different for different applications, there is no standard rule to judge which biclusters produced are better. In [44], Prelic et al. defined match score of two clusters $\mathcal{S}_i, \mathcal{S}'_i$ of samples as

$$S(B_1, B_2) = \frac{|\mathcal{S}_i \cap \mathcal{S}'_i|}{|\mathcal{S}_i \cup \mathcal{S}'_i|},$$

and match score between two sets $\mathcal{B}, \mathcal{B}'$ of biclusters for matrix A as

$$S^*(\mathcal{B}, \mathcal{B}') = \frac{1}{|\mathcal{B}|} \sum_{(\mathcal{S}_i, \mathcal{F}_i) \in \mathcal{B}} \max_{(\mathcal{S}'_i, \mathcal{F}'_i) \in \mathcal{B}'} \frac{|\mathcal{S}_i \cap \mathcal{S}'_i|}{|\mathcal{S}_i \cup \mathcal{S}'_i|},$$

which reflects the average of the maximum match scores for all biclusters in $\mathcal{B}$ with respect to the biclusters in $\mathcal{B}'$.

In [44], Prelic et al. used this score to comparing the algorithms of Bimax, CC, OPSM, SAMBA, xMotifs, and ISA with respect to the data set of a metabolic pathway map. And in [12], Cho and Dhillon also use this score to compare several biclustering algorithms on human cancer microarrays data sets.

6.4 Application of Biclustering in Computational Neuroscience

Epilepsy is one of the most common nervous system disorders. It affects about 1% of the world's population with the highest incidence among infants and the elderly [20, 21]. For many years there have been attempts to control epileptic seizures by electrically stimulating the brain [25]. This alternate method of treatment is the subject of much study since the approval of the chronic vagus nerve stimulation (VNS) implant for treatment of intractable seizures [56, 24, 49]. The device consists of an electric stimulator implanted subcutaneously in the chest and connected, via subcutaneous electrical wires, to the left cervical vagus nerve. The VNS is programmed to deliver electrical stimulation at a set intensity, duration, pulse width, and frequency. Optimal parameters are determined on a case-by-case basis, depending on clinical efficacy (seizure frequency) and tolerability.

Busygin et al. used supervised consistent biclustering [6] to develop a physiologic marker for optimal VNS parameters (e.g., output current, signal frequency) using measures of scalp EEG signals.

The raw EEG data was obtained from two patients A and B at 512 Hz sampling rate from 26 scalp EEG channels arranged in the standard international 10–20 system (see Fig. 6.1). Then the EEG was transformed into a sequence of short-term largest Lyapunov exponents (STL_{max}) values. A famous practical application of STL_{max} measure of EEG signal time series is to predict epileptic seizures, see [29, 41, 42]. Thus, Lyapunov exponents are considered to be a perfect descriptor of such extremely complex dynamic system as human brain.

STL_{max} values were computed for each scalp EEG channel recorded from two epileptic patients using the algorithm developed by Iasemidis et al. [29]. Then the STL_{max} values were used as features of the two data sets. The averaged samples from stimulation periods were then separated from averaged samples from nonstimulation periods by feature selection performed within the consistent biclustering routine.
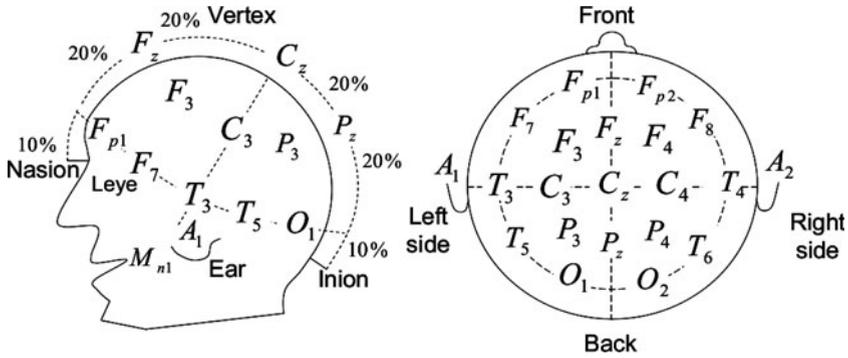
Fig. 6.1: Montage for scalp electrode placement.

As each stimulation lasted for 30 s and a 4-s time window was used to compute one element of the Lyapunov exponent time series, each stimulation provided seven data points. Since the EEG patterns of a patient may have been changing throughout the observed period due to changes in his/her conditions not relevant to the investigated phenomenon, each of the seven samples across all stimulation cycles were averaged. Thus, seven Lyapunov exponent samples have been created to represent the positive class. To create the negative class, 10 Lyapunov exponent data points were considered 250 s after each stimulation. In the similar way, these 10 samples were averaged across all stimulation cycles. So, the created negative class contains 10 averaged Lyapunov exponent data samples from nonstimulation time intervals.

Then, the biclustering experiment was done on two 26×17 matrices representing patients A and B. The patient A data were conditionally biclustering admitting with respect to given stimulation and nonstimulation classes without excluding any features. All but one feature were classified into the nonstimulation class, which indicates that for almost all EEG channels the Lyapunov exponent was consistently decreasing during the stimulation with one channel being the only exception.

Cross-validation was performed for the obtained biclustering by leave-one-out method examining for each sample whether it would be classified in the appropriate class if the feature selection was performed without it. It turned out that all classes of all 17 samples are confirmed by this method.

To make the patient B data set conditionally biclustering admitting with respect to given stimulation and nonstimulation classes only five features were selected. The one-leave-out experiment classified correctly all but four samples. The biclustering heatmaps are presented in Fig. 6.2.

The obtained biclustering results allow to assume that signals from certain parts of the brain consistently change their characteristics when VNS is switched on and could provide a basis for desirable VNS stimulation parameters. A physiologic marker of optimal VNS effect could greatly reduce the cost, time, and risk of calibrating VNS stimulation parameters in newly implanted patients compared to the current method of clinical response.
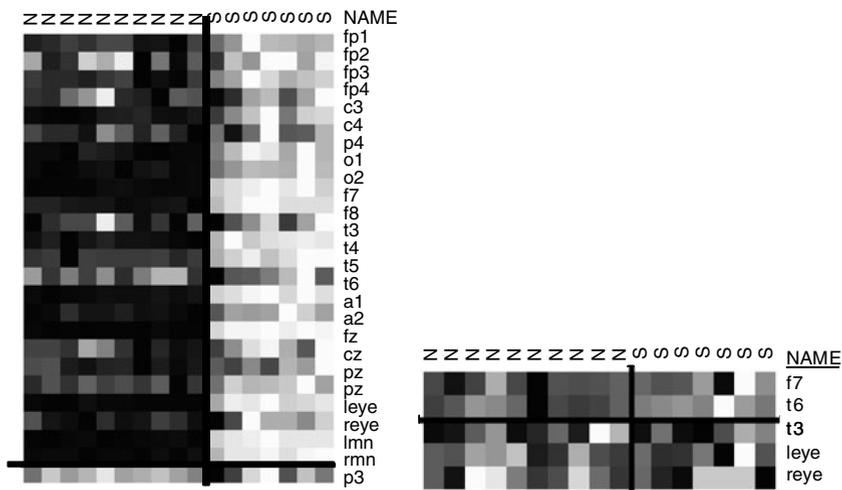
Fig. 6.2: Heatmaps for patients A and B.

6.5 Conclusions

In this review, the formal definitions of biclustering with its different types and structures are given and the algorithms are reviewed in mathematical prospective.

Biclustering is recently a hot research area with its applications in bioinformatics. Other application areas are text mining, marketing analysis, etc. In practical applications, some problems, such as the data missing, the noise of data, and data processing, influence a lot to the results of biclustering. Besides, the comparisons of biclustering algorithms are still another direction to be studied.

References

1. Angiulli, F., Cesario, E., Pizzuti, C. Random walk biclustering for microarray data. Inf Sci: Int J **178**(6), 1479–1497 (2008)
2. Barkow, S., et al. BicAT: A biclustering analysis toolbox. Bioinformatics **22**, 1282–1283 (2006)
3. Ben-Dor, A., Chor, B., Karp, R., Yakhini, Z. Discovering local structure in gene expression data: The order-preserving submatrix problem. J Comput Biol **10**, 373–384 (2003)
4. Busygin, S., Prokopyev, O.A., Pardalos, P.M. Feature selection for consistent biclustering via fractional 0–1 programming. J Comb Optim **10**/1, 7–21 (2005)
5. Busygin, S., Prokopyev, O.A., Pardalos, P.M. Biclustering in datamining. Comput Oper Res **35,** 2964–2987 (2008)
6. Busygin, S., Boyko, N., Pardalos, P., Bewernitz, M., Ghacibehc, G. Biclustering EEG data from epileptic patients treated with vagus nerve stimulation. AIP Conference Proceedings of the Data Mining, Systems Analysis and Optimization in Biomedicine, 220–231 (2007)
7. Califano, A., Stolovitzky, G., Tu, Y. Analysis of gene expression microarays for phenotype classification. Proceedings of International Conference on Computational Molecular Biology, 75–85 (2000)

8. Carmona-Saez, P., Pascual-Marqui, R.D., Tirado, F., Carazo, J.M., Pascual-Montano, A. Biclustering of gene expression data by non-smooth non-negative matrix factorization. BMC Bioinformatics **7**, 78 (2006)
9. Chaovalitwongse, W.A., Butenko, S., Pardalos, P.M. Clustering Challenges in Biological Networks, World Scientific Publishing, Singapore (2008)
10. Cheng, K.O., et al. Bivisu: Software tool for bicluster detection and visualization. Bioinformatics **23**, 2342–2344 (2007)
11. Cheng, Y., Church, G.M. Biclustering of expression data. Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology, 93–103 (2000)
12. Cho, H., Dhillon, I.S. Coclustering of human cancer microarrays using minimum sum-squared residue coclustering. IEEE/ACM Trans Comput Biol Bioinform **5**(3), 385–400 (2008)
13. Chung, F.R.K. Spectral graph theory. Conference Board of the Mathematical Sciences, Number 92, American Mathematical Society (1997)
14. CPLEX: ILOG CPLEX 9.0 Users Manual (2005)
15. Data Clustering. http://en.wikipedia.org/wiki/Data clustering, access at Dec. 8 (2008)
16. Data Transformation Steps. http://www.dmg.org/v2–0/Transformations.html, access at Dec. 8 (2008)
17. Dhillon, I.S. Co-clustering documents and words using bipartite spectral graph partitioning. Proceedings of the 7th ACM SIGKDD International Conference on Knowledging Discovery and Data Mining (KDD), 26–29 (2001)
18. Dhillon, I.S., Mallela, S., Modha, D.S. Information theoretic co-clustering. Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 89–98 (2003)
19. DiMaggio, P.A., McAllister, S.R., Floudas, C.A., Feng, X.J., Rabinowitz, J.D., Rabitz, H.A. Biclustering via optimal re-ordering of data matrices in systems biology: Rigorous methods and comparative studies. BMC Bioinformatics **9**, 458 (2008)
20. Engel, J. Jr. Seizures and Epilepsy. F. A. Davis Co., Philadelphia, PA (1989)
21. Engel, J. Jr., Pedley, T.A. Epilepsy: A Comprehensive Textbook. Lippincott-Raven, Philadelphia, PA (1997)
22. Fan, N., Chinchuluun, A., Pardalos, P.M. Integer programming of biclustering based on graph models, In: Chinchuluun, A., Pardalos, P.M., Enkhbat, R. and Tseveendorj, I. (eds.) Optimization and Optimal Control: Theory and Applications, Springer (2009)
23. Fan, N., Pardalos, P.M. Linear and quadratic programming approaches for the general graph partitioning problem, J Global Optim, DOI 10.1007/s10898-009-9520-1, (2010)
24. Fisher, R.S., Krauss, G.L., Ramsay, E., Laxer, K., Gates, J. Assessment of vagus nerve stimulation for epilepsy: Report of the therapeutics and technology assessment subcommittee of the American academy of neurology. Neurology 49, 293–297 (1997)
25. Fisher, R.S., Theodore W.H. Brain stimulation for epilepsy. Lancet Neurol 3(2), 111–118 (2004)
26. Gu, J., Liu, J.S. Bayesian biclustering of gene expression data. BMC Genom **9**(Suppl 1), S4 (2008)
27. Hagen, L., Kahng, A.B. New spectral methods for ratio cut partitioning and clustering. IEEE Trans Computer-Aided Design **11**(9), 1074–1085 (1992)
28. Hartigan, J.A. Direct clustering of a data matrix. J Am Stat Assoc **67**, 123–129 (1972)
29. Iasemidis, L.D., Principe, J.C., Sackellares, J.C. Measurement and quantification of spatiotemporal dynamics of human epilepic seizures. In: Akay, M. (ed.) Nonlinear Signal Processing in Medicine, IEEE Press (1999)
30. Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y., Barkai, N. Revealing modular organization in the yeast transcriptional network. Nat Genet **31**(4), 370–377 (2002)
31. Jain, A.K., Murty, M.N., Flynn, P.J. Data clustering: A review. ACM Comput Survey **31**(3), 264–323 (1999)
32. Kaiser, S., Leisch, F. A toolbox for bicluster analysis in r. Tech. Rep. 028, Ludwing-Maximilians-Universitat Mnchen (2008)
33. Kluger, Y., Basri, R., Chang, J.T., Gerstein, M. Spectral biclustering of microarray cancer data: Co-clustering genes and conditions. Genome Res **13**, 703–716 (2003)

34. Lazzeroni, L., Owen, A. Plaid models for gene expression data. Stat Sinica **12**, 61C86 (2002)
35. Lee, D.D., Seung, H.S. Learning the parts of objects by non-negative matrix factorization. Nature **401,** 788–791 (1999)
36. Liu, J., Wang, W. OP-cluster: Clustering by tendency in high dimensional space. Proceedings of the Third IEEE International Conference on Data Mining, 187–194 (2003)
37. Madeira, S.C., Oliveira, A.L. Biclustering algorithms for biological data analysis: A survey. IEEE Trans Comput Biol Bioinform **1**(1), 24–45 (2004)
38. Madeira, S.C., Oliveira, A.L. A linear time biclustering algorithm for time series gene expression data. Lect Notes Comput Sci **3692**, 39–52, (2005)
39. Murali, T.M., Kasif, S. Extracting conserved gene expression motifs from gene expression data. Pacific Symp Biocomput **8**, 77–88 (2003)
40. Pardalos, P.M., Busygin, S., Prokopyev, O.A. On biclustering with feature selection for microarray data sets. In: Mondaini, R. (ed.) BIOMAT 2005Ïinternational Symposium on Mathematical and Computational Biology, pp. 367–378. World Scientific, Singapore (2006)
41. Pardalos, P.M., Chaovalitwongse, W., Iasemidis, L.D., Sackellares, J.C., Shiau, D.-S., Carney, P.R., Prokopyev, O.A., Yatsenko, V.A. Seizure warning algorithm based on optimization and nonlinear dynamics. Math Prog **101**(2), 365–385 (2004)
42. Pardalos, P.M., Chaovalitwongse, W., Prokopyev, O. Electroencephalogram (EEG) time series classification: Application in epilepsy. Ann Oper Res (2006)
43. Pascual-Montano, A., Carazo, J.M., Kochi, K., Lehmann, D., Pascual-Marqui, R.D. Non-smooth Non-negative matrix factorization (nsNMF). IEEE Trans Pattern Anal Mach Intell **28**, 403–415 (2006)
44. Prelic, A., Bleuler, S., Zimmermann, P., Wille, A., Buhlmann, P., Gruissem, W., Hennig, L., Thiele, L., Zitzler, E. A systematic comparison and evaluation of biclusteringmethods for gene expression data. Bioinformatics **22**(9), 1122–1129, (2006)
45. Rege, M., Dong, M., Fotouhi, F. Bipartite isoperimetric graph partitioning for data co-clustering. Data Min Know Disc **16**, 276–312 (2008)
46. Reiss, D.J., Baliga, N.S., Bonneau, R. Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. BMC Bioinformatics **7**, 280 (2006)
47. Richards, A.L., Holmans, P.A., O'Donovan, M.C., Owen, M.J., Jones, L. A comparison of four clustering methods for brain expression microarray data. BMC Bioinformatics **9**, 490 (2008)
48. Santamaria, R., Theron, R., Quintales, L. BicOverlapper: A tool for bicluster visualization Rodrigo. Bioinformatics **24**, 1212–1213 (2008)
49. Schachter, S.C., Wheless, J.W. (eds.) Vagus nerve stimulation therapy 5 years after approval: A comprehensive update. Neurology **S4**, 59 (2002)
50. Sheng, Q., Moreau, Y., De Moor, B. Biclustering microarray data by Gibbs sampling. Bioinformatics **19**, 196–205 (2003)
51. Shi, J., Malik, J. Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell, **22**(8), 888–905 (2000)
52. Supper, J., Strauch, M., Wanke, D., Harter, K., Zell, A. EDISA: Extracting biclusters from multiple time-series of gene expression profiles. BMC Bioinformatics **8**, 334 (2007)
53. Tanay, A., Sharan, R., Kupiec, M., Shamir, R. Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. Proc Natl Acad Sci USA **101**, 2981–2986 (2004)
54. Tanay, A., Sharan, R., Shamir, R. Discovering statistically significant bilcusters in gene expression data. Bioinformatics **18**, S136–S144 (2002)
55. Tanay, A., Sharan, R., Shamir, R. Biclustering algorithms: A survey. In: Aluru, S. (ed.) Handbook of Computational Molecular Biology. Chapman Hall, London (2005)
56. Uthman, B.M., Wilder, B.J., Penry, J.K., Dean, C., Ramsay, R.E., Reid, S.A., Hammond, E.J., Tarver, W.B., Wernicke, J.F. Treatment of epilepsy by stimulation of the vagus nerve. Neurology **43**, 1338–1345 (1993)
57. Xu, R., Wunsch, D. II. Survey of clustering algorithms. IEEE Trans Neural Netw **16**(3), 645–678 (2005)

58. Yang, J., Wang, W., Wang, H., Yu, P. δ -Clusters: Capturing subspace correlation in a large data set. Proceedings of the 18th IEEE International Conference on Data Engineering, 517–528 (2002)
59. Yang, J., Wang, W., Wang, H., Yu, P. Enhanced biclustering on expression data. Proceedings of the Third IEEE Conference on Bioinformatics and Bioengineering, 321–327 (2003)
60. Zha, H., He, X., Ding, C., Simon, H., Gu, M. Bipartite graph partitioning and data clustering. Proceedings of the Tenth International Conference on Information and Knowledge Management, 25–32 (2001)
61. Zhao, H., Liew, A.W.-C., Xie, X., Yan, H. A new geometric biclustering based on the Hough transform for analysis of large-scale microarray data. J Theor Biol **251**, 264–274 (2008)

Chapter 7
A Genetic Classifier Account for the Regulation of Expression

Tsvi Achler and Eyal Amir

Abstract This work is motivated by our model of neuroscience processing which incorporates large numbers of reentrant top-down feedback regulation connections. Such regulation is fundamental and can be found throughout biology. The purpose of this chapter is to broaden this model's application.

Genes perform important life functions, responsible for virtually every organic molecule that organisms produce. The genes must closely regulate the amount of their products, because too little or too much production may be deleterious for the organism. Furthermore, they must respond efficiently and in unison to the environments that the organism faces. Networks that are closely regulated can behave as robust classifiers which can recognize and respond to their environment. Using simple examples we demonstrate that such networks perform dynamic classification, determining the most efficient set of genes needed to replace consumed products.

7.1 Introduction

7.1.1 Motivation

Genes working together are involved in the production and regulation of proteins and precursors necessary to maintain life. These genetic networks must self-regulate their expression in order to produce the correct products in practical amounts. The

Tsvi Achler

Department of Computer Science, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA, e-mail: achler@uiuc.edu

Eyal Amir

Department of Computer Science, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA, e-mail: eyal@cs.uiuc.edu

focus of this chapter is on network-wide coordination through a simple control mechanism at every gene.

A basic assumption of the genetic model of expression is that each gene is closely regulated by its products. If the rate of consumption of a gene's product exceeds the rate of production, more is produced. If little is consumed, less is produced. However, production pathways are complex. Products can share pathways of consumption or production. For example, different products may require similar enzymes. Some molecules can be converted to common precursors. A product can contribute to or be produced by separate pathways and consumed for different purposes. For regulation to be effective the expression of the genes whose products intermix must be coordinated. The hypothesis is that genes interact and regulate each other through their common products. Yet each gene is regulated by a simple control mechanism. Such networks show complex coordination, perform recognition, and have been previously described in the context of neuroscience [3, 4].

It is demonstrated that if a protocol of regulation is preserved, regulation can form a genetic classifier. The classifier monitors product consumption and finds the most efficient configuration of genes to replace the products. This configuration minimizes the amount of unused products and responds to environmental demands. A fundamental understanding of these regulatory mechanisms can guide experiment design, reveal methods to control gene expression, and advance genetic therapy approaches.

7.1.2 Background

Complex interactions occur between the genes and the cellular environment they control. Genes not only autoregulate their expression but interact with each other via numerous mechanisms within the process of converting DNA to final proteins. Gene regulatory networks integrate multiple signals to determine protein production. Expression is ultimately regulated by concentrations of products and intermediaries of metabolic pathways.

Understanding genetic-protein structure and dynamics relationships in networks is a major goal of complex systems research [8]. Although numerous relationships between specific structural and dynamical aspects of network components have been investigated [5, 6, 10], general principles behind such relationships are still unknown [9]. Thus a high degree of regulation occurs throughout genetic-protein production pathways, but many aspects are unclear.

Instead of direct gene-to-gene interactions (i.e., *gene1* promotes or inhibits *gene2*), our model focuses on a gene–product axis. Suppose *gene1* and *gene2* share the same product or pathway. The genes also share regulation. A gene's regulation of its product will affect the other gene that regulates that product. All genes that regulate the same product reach a communal equilibrium. Any change in the communal equilibrium changes the expression of multiple genes. Gene–product regulation establishes indirect and nonlinear gene-to-gene interactions. With these interactions a

system emerges that is sufficient to implement a recognition system [3, 4, 2]. The properties of such genetic regulatory networks are investigated.

7.2 Model and Methods

This section describes the structure, function, and equations of the genetic regulation model.

7.2.1 Basic Assumptions

The role of a gene promoter is to measure the amount of product available and determine gene expression levels. The most important assumption, on which this model builds upon, is that each promoter aims to produce a fixed amount of product. If too much product is consumed, the promoter signals more product must be expressed. If too little is consumed, the promoters signal less to be expressed. A gene that affects multiple products is regulated by those products. Thus, every input–output relation is regulated by feedback.

A classifier based on feedback can be surprisingly powerful [3, 4]. This structure maintains its simplicity in large networks but can still make complex recognition decisions based on distributed processing.

7.2.2 Model Structure

The proposed tight association between genes and products and promoters is depicted in Fig. 7.1.
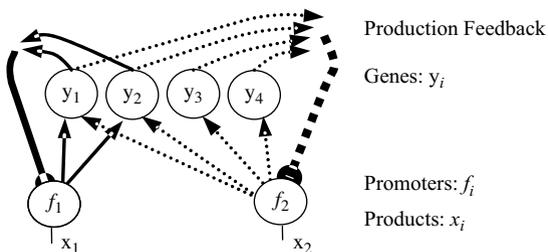

Fig. 7.1: Self-regulation. If y_1 and y_2 affect x_1 then f_1 monitors x_1 and regulates y_1 and y_2. Similarly if y_1, y_2, y_3, and y_4 affect x_2 then f_2 monitors x_2 and regulates y_1, y_2, y_3, and y_4.

Every product x has a corresponding promoter "f," which samples the concentration of x and accordingly regulates the genes associated with product x. The promoter modulates the gene's expression y based on function x/f.

The tight association creates a situation where the only way a gene can be fully promoted by a promoter, say f_1, is if it is the only active gene that affects product x_1 (reducing f_1). Each promoter maintains equilibrium of its product regardless of the state of other promoters. However, multiple promoters can affect a gene's expression.

If several genes affect the same product, no gene will be fully promoted by the amount the product is consumed. For example, if two genes affect the same product, the promoters each uses will not be available for the other. In this way they "inhibit" each other through the product's promoter, forcing these genes' promotion to be mediated through other promoters from other processes.

The more products two genes mutually interact with, the more they will blunt each other's promoters. The less products genes mutually affect, the less their promoters will mutually interfere, and the more "parallel" or independent these genes can be.

The network dynamically evaluates gene expression by

1. Determining promoter activity based on product concentration.
2. Modifying gene expression based on the promoter.
3. Redetermining promoter activity based on new product concentration.

Steps 1–3 are continuously cycled through expression, promoters, and products.

7.2.3 Model Equations

This section introduces the nonlinear equations governing this network. For any gene y denoted by index a, let N_a denote all products that gene y_a affects. Let n_a denote the number of products gene y_a affects. For any product x denoted by index b, let M_b denote all genes that affect x_b. The total amount of expression of product x_b is Y_b, which is the sum of expression from all genes that affect product x_b.

$$Y_b = \sum_{j \in M_b} y_j(t) \tag{7.1}$$

Efficacy of promoter f_b is determined by consumption of x_b and the overall production of x_b:Y_b. This is determined by

$$f_b = \frac{x_b}{Y_b} \tag{7.2}$$

The expression of y_a is dependent on its previous expression and its promoters. The equations are designed so that gene expression is proportional to the amount of product consumed and inversely proportional to their promoters based on product consumption and also depends on their previous expression levels [2, 1]. Describing

these equations with engineering control theory nomenclature, feedback regulation is "negative," stabilizing feedback.

$$y_a(t+dt) = \frac{y_a(t)}{n_a} \sum_{i \in N_a} f_i = \frac{y_a(t)}{n_a} \sum_{i \in N_a} \left(\frac{x_i}{\sum_{j \in M_i} y_j(t)} \right) \tag{7.3}$$

7.2.4 Stability

Stability of related equations has been previously analyzed [7]. If nonlinear equations are bounded and well behaved locally, they remain stable. In this model, all variables are limited to positive values. Thus the values of y cannot become negative and have a lower bound of 0. The upper values of y are bounded as well. The expression value of gene y_a will be greatest if all of its promoters f_i are maximized. The promoters will be maximized if genes coactivated by that promoter are not active. Assuming this is the case then the equation simplifies to

$$y_a(t+\Delta t) \le \frac{1}{n_a} \sum_{i \in N_a} \left(\frac{y_a(t) \cdot x_{max}}{y_a(t)} \right) = \frac{1}{n_a} \sum_{i \in N_a} x_{max} \le \frac{x_{max} \cdot n_a}{n_a} = x_{max} \tag{7.4}$$

If maximum consumption x_{max} is bounded by 1, then y_a expression is bounded by 1. The values are bounded by positive numbers between zero and the consumption level. Thus they satisfy boundary conditions and are well behaved. Furthermore as $dt \to 0$, Lyapunov functions can be written. This indicates that the networks will settle to a steady state and not display chaotic oscillations. Numerical simulations also show the equations are well behaved and several cases of gene–product interactions are demonstrated.

7.3 Results

This system attempts to replace consumed products through a minimum amount of overall gene expression. Several configurations of genes are analyzed to illustrate how the system interacts with different patterns of product consumption.

7.3.1 Composition by Overlap of Nodes

7.3.1.1 Complete Overlap

Given that two genes lead to the same product but one of them also leads to another product, how do they respond to consumption patterns to minimize expression? In a
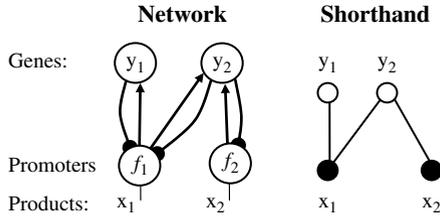
Fig. 7.2: Example 1, simple network with two genes. With feedforward and feed-back connections included (*left*). Equivalent configuration redrawn with bidirectional connections (*right*).

simple example, Example 1 (Fig. 7.2), if genes y_1 and y_2 lead to product x_1, then by definition the x_1 promoter f_1 affects genes y_1 and y_2. Similarly if gene y_2 leads to product x_2, gene y_2 is additionally regulated by product x_2. Thus gene y_1 is regulated by product x_1 and gene y_2 is regulated by products x_1 and x_2. Due to the feedback model, the activation of gene y_1 can depend on the level of product x_2, because if products x_1 and x_2 are consumed equally, then gene y_2 will be promoted at the expense of gene y_1.

The network is evaluated until it settles onto a steady state. The solutions are presented as (*products consumed*) → (*genes expressed*). Since there are two products and two genes in Example 1, the solution is written in the form $(x_1, x_2) \rightarrow (y_1, y_2)$.

The steady-state solution for Example 1 is $(x_1, x_2) \rightarrow (y_1 = x_1 - x_2, y_2 = x_2)$. The mathematical equations and their derivation follow:

$$y_1(t+dt) = \frac{y_1(t)x_1}{y_1(t)+y_2(t)}, \quad y_2(t+dt) = \frac{y_2(t)}{2}\left(\frac{x_1}{y_1(t)+y_2(t)} + \frac{x_2}{y_2(t)}\right). \quad (7.5)$$

The network solution at steady state is derived by setting $y_1(t+dt) = y_1(t)$ and $y_2(t+dt) = y_2(t)$ and solving these equations. The solutions are $y_1 = x_1 - x_2$ and $y_2 = x_2$. If $x_1 \leq x_2$ then $y_1 = 0$ and the equation for y_2 becomes $y_2 = \frac{x_1+x_2}{2}$.

This solution demonstrates efficient outcomes where minimal products are wasted. Neither x_1 nor x_2 is produced if they are not needed. For example, when products x_1 and x_2 are equally consumed $(x_1 = x_2)$ then gene y_2 is expressed and gene y_1 is silenced. This occurs because x_1 and x_2 equally usurp promoter f_1. From the perspective of the genes, gene y_1 has all of its promoters reduced when f_1 is usurped, while gene y_2 still has an independent promoter f_2. Gene y_2 expression becomes preferred and in the process inhibits gene y_1. The final result is that if product x_2 is not consumed, gene y_2 is not expressed.

If only product x_1 is consumed $(x_2 = 0)$ then only gene y_1 is expressed avoiding extraneous products. There are consumption patterns where this configuration is not efficient. For example, if only product x_2 is consumed $(x_1 = 0)$ then only y_2 is expressed but extraneous product x_1 is produced.

7.3.1.2 Incomplete Overlap

What happens if there are no efficient configurations? In these cases genes may not completely dominate. In Example 2 (Fig. 7.3), gene y_1 is replaced by gene y_3. Genes y_2 and y_3 can equally affect product x_2, but also affect independent products x_1 and x_3, respectively. If only product x_2 is consumed, or products $x_1 - x_3$ are consumed equally, either gene y_2 or gene y_3 can lead to the needed product x_2. However, in either case, there will be some extraneous products that are not consumed. Genes that lead to two products cannot express only one product. The simulations reflect this imbalance and the solution is more complicated. The mathematical solutions are

$$(x_1, x_2, x_3) \rightarrow \left(y_1 = \frac{x_1(x_1 + x_2 + x_3)}{2(x_1 + x_3)}, y_2 = \frac{x_3(x_1 + x_2 + x_3)}{2(x_1 + x_3)} \right) \qquad (7.6)$$
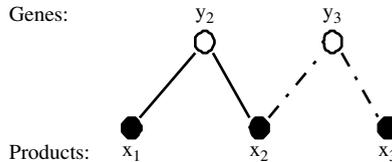
Fig. 7.3: Example 2.

When inputs are (1, 1, 1) the output cells become ($3/4$, $3/4$). Furthermore, if only the middle input is active (0, 1, 0) then the forces on both cells are symmetrical, the equation collapses to $2(y_1 + y_2) = x_2$ and the solution depends on initial conditions. Thus either gene can express x_2, and there is no preference between the genes.

7.3.2 Multiple Gene Scenarios

7.3.2.1 Three Genes

Given multiple genes with overlapping products how can they promote or inhibit each other's expression based on consumption patterns? Example 3 (Fig. 7.4) is composed of Examples 1 and 2 combined. Now three genes share products. A third gene y_3 is introduced which leads to products x_2 and x_3 (and regulated by products x_2 and x_3). This example demonstrates how genes can interact in a distributed fashion. In this configuration genes y_1 and y_3 can together turn off the expression of gene y_2.

Equation analysis: equation $y_1(t + dt)$ remains the same as Example 1. $y_2(t + dt)$ and $y_3(t + dt)$ are given by
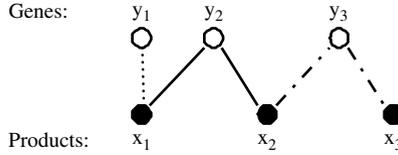
Fig. 7.4: Example 3.

$$y_2(t+dt) = \frac{y_2(t)}{2}\left(\frac{x_1}{y_1(t)+y_2(t)} + \frac{x_2}{y_2(t)+y_3(t)}\right), \qquad (7.7)$$

$$y_3(t+dt) = \frac{y_3(t)}{2}\left(\frac{x_2}{y_2(t)+y_3(t)} + \frac{x_3}{y_3(t)}\right)$$

The steady-state solution limited to positive gene expression values is

$$(x_1, x_2, x_3) \rightarrow (y_1 = x_1 - x_2 + x_3, y_2 = x_2 - x_3, y_3 = x_3). \qquad (7.8)$$

If $x_2 \leq x_3$ then $y_2 = 0$ and the equations become $\left(x_1, 0, \frac{x_2+x_3}{2}\right)$. If $x_3 = 0$ the solution becomes that of Example 1: $(x_1, x_2, 0) \rightarrow (x_1 - x_2, x_2, 0)$.

Similar to Example 1, if product x_1 is consumed, gene y_1 is expressed. If only products x_1 and x_2 are consumed equally, then gene y_2 is expressed. The underlying mechanisms remain the same as Example 1.

However, unlike Example 2, this configuration now has an efficient configuration for the case where products x_1, x_2, and x_3 are equally consumed. With equal consumption genes y_1 and y_3 are expressed equally $(1, 1, 1) \rightarrow (1, 0, 1)$ and gene y_2 expression turned off. This expression pattern most efficiently replaces this consumption pattern with the least amount of extraneous products.

This case demonstrates that information travels indirectly "through" the promoters based on gene structures. Given equal consumption of x_1 and x_2, expression of y_1 is determined by consumption of x_3 through y_3. If x_3 is not consumed (its value is 0), then gene y_1 is not expressed. If x_3 is consumed, y_1 becomes active. However, x_3 is not directly affected by y_1, and the product affected by $y_1(x_1)$ is not directly expressed by y_3. Thus genes can cooperate and function in groups, choosing the best single gene or gene combination that efficiently replaces the consumption pattern.

7.3.3 Composition by Infinite Chains

What are the limits of these interactions? The behavior of an infinitely large number of genes with overlapping products is analyzed. No matter how many genes are linked, this genetic model attempts to match the product consumption with the most efficient gene expression configuration. To demonstrate this, gene networks are composed of chained subunits linked at infinitum. The promoters and genes interact indirectly by transferring their dynamic activation through the chain.

7.3.3.1 Chain of Genes Including A 1-Product Gene

Consider the case where there are N two-product genes that are connected in a chain shown in Fig. 7.5. This configuration includes a 1-product gene similar to Examples 1 and 3. Suppose all products are consumed equally, for example, all have the value 1. The network will find the most efficient configuration of expression where no extraneous products are created. These configurations may change based on the properties of the links. For example, suppose there are N gene links. If N is an *odd* number then gene y_1 will express its single product and every second gene will express their products. The genes interspersed in between will be turned off (0). If N is *even*, y_1 will be turned off and the *even* genes expressed (1) and *odd* ones turned off. If i and j represent gene indexes the general solution becomes

$$(x_1, x_i, \cdots, x_N) \to \left(\sum_{i \le j \le N} (-1)^j x_j, \cdots, x_N \right)$$

With a 1-Product Gene (y_1)
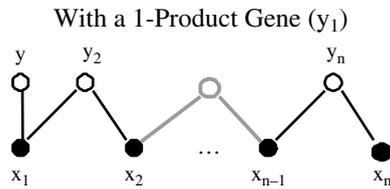
Fig. 7.5: Example 4a.

For example, with four genes chained, $N = 4 : (1,1,1,1) \to (0,\ 1,\ 0,\ 1)$. With five genes chained $N = 5 : (1,1,1,1,1) \to (1,\ 0,\ 1,\ 0,\ 1)$. If the concentrations of x are such that $y < 0$, the chain breaks at that gene and the rest of the links behave as smaller independent chains from that point (*see* Section 7.3.4).

7.3.3.2 Chain of Genes Without A 1-Product Gene

If a one product gene is not available, then the network does not have a favorable set of genes to resolve an odd number of products. Two-product genes cannot produce an odd number of products. The configuration with three products was presented in Example 2. In case of four inputs (even) distributed over three genes the solution becomes

$$\left(\frac{x_1(\Sigma X)}{2(x_1 + x_3)},\ \frac{-(\Sigma X)(x_1 x_4 - x_3 x_2)}{2(x_1 + x_3)(x_2 + x_4)},\ \frac{x_4(\Sigma X)}{2(x_2 + x_4)} \right) \quad \text{where} \quad \Sigma X = x_1 + x_2 + x_3 + x_4.$$

When all products are consumed, all x's $= 1$, the cells settle on a binary solution $(1, 0, 1)$. Thus simple solutions can be found as long as there is an even number of products consumed. Cases with $N > 4$ genes become progressively complicated to solve and beyond the scope of this chapter (Fig. 7.6).

2-Product Genes Only
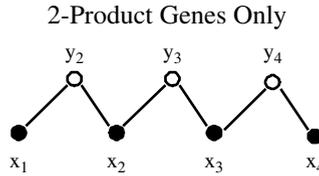
Fig. 7.6: Example 4b.

7.3.4 Subchains

If a product in the chain is not consumed, this can break the chain into independent components composed of the right and left parts of the chain from the unconsumed product. These chains can function as smaller chains. For example, if product $x_6 = 0$, the chains involving genes y_{1-6} and y_{6-N} become independent. Thus gene expression patterns are determined by distributed product-promoter dynamics involving consumption and gene structures. Further analysis remains for future research.

7.4 Discussion

This theory shows that highly regulated genes can affect one another and form a classification system. The recognition system configures the expression of multiple genes to efficiently minimize extraneous products. This chapter serves as a demonstration of this concept. Though details of the molecular mechanisms have been abstracted, this model suggests methods to control gene expression by artificially introducing products. Genetic data indicating shared promoter regions between genes may predict which genes compete.

Suppose a patient has a deleterious gene. Though still highly speculative, this model suggests that introducing artificial products which match a gene's regulation pattern may change the deleterious gene's expression. Through gene competition, artificial products may be introduced to favor other native genes which share the same production pathway and will turn off the deleterious gene.

Alternatively, if a gene has been artificially inserted but its products are not sufficiently expressed, it may be possible to inhibit native genes. This can be achieved by introducing protein products to match product patterns of the native genes.

Lastly, since promoters are distributed across genes, this system reveals how copied genes can integrate into the genome while still being closely regulated.

In summary, this chapter outlines the concept and implications of regulatory feedback systems that maintain homeostasis and explores their systematic properties. This model suggests speculative methods to control gene expression by manipulating shared molecular pathways.

References

1. Achler, T. Input shunt networks. Neurocomputing **44**, 249–255 (2002)
2. Achler, T. Object classification with recurrent feedback neural networks. Proceedings of the SPIE Evolutionary and Bio-inspired Computation: Theory and Applications, vol. 6563, Orlando (2007)
3. Achler, T., Amir, E. Input feedback networks: Classification and inference based on network structure. Artif General Intell **1**, 15–26 (2008)
4. Achler, T., Omar, C., Amir, E. Shedding weights: More with less. Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IJCNN'08), Hong Kong, pp. 3020–3027 (2008)
5. Albert, R., Jeong, H., Barabasi, A.L. Error and attack tolerance of complex networks. Nature **406**(6794), 378–382 (2000)
6. Kauffman, S., et al. Random Boolean network models and the yeast transcriptional network. Proc Natl Acad Sci USA **100**(25), 14796–14799 (2003)
7. Mcfadden, F.E. Convergence of competitive activation models based on virtual lateral inhibition. Neural Networks **8**(6), 865–875 (1995)
8. Nochomovitz, Y.D., Li, H. Highly designable phenotypes and mutational buffers emerge from a systematic mapping between network topology and dynamic output. Proc Natl Acad Sci USA **103**(11), 4180–4185 (2006)
9. Nykter, M., et al. Critical networks exhibit maximal information diversity in structure-dynamics relationships. Phys Rev Lett **100**(5), 058702 (2008)
10. Shmulevich, I., et al. The role of certain Post classes in Boolean network models of genetic networks. Proc Natl Acad Sci USA **100**(19), 10734–10739 (2003)

Part II
Modeling

Chapter 8
Neuroelectromagnetic Source Imaging of Brain Dynamics

Rey R. Ramírez, David Wipf, and Sylvain Baillet

Abstract Neuroelectromagnetic source imaging (NSI) is the scientific field devoted to modeling and estimating the spatiotemporal dynamics of the neuronal currents that generate the electric potentials and magnetic fields measured with electromagnetic (EM) recording technologies. Unlike functional magnetic resonance imaging (fMRI), which is indirectly related to neuroelectrical activity through neurovascular coupling [e.g., the blood oxygen level-dependent (BOLD) signal], EM measurements directly relate to the electrical activity of neuronal populations. In the past few decades, researchers have developed a great variety of source estimation techniques that are well informed by anatomy, neurophysiology, and the physics of volume conduction. State-of-the-art approaches can resolve many simultaneously active brain regions and their single trial dynamics and can even reveal the spatial extent of local cortical current flows.

8.1 Introduction

NSI methods model and estimate the spatiotemporal dynamics of neuronal currents throughout the brain as accessed by noninvasive and invasive surface measurements such as electroencephalography (EEG), magnetoencephalography (MEG), and electrocorticography (ECoG) [6, 22, 30, 31].

Rey R. Ramírez
MEG Program, Department of Neurology, Medical College of Wisconsin and Froedtert Hospital, Milwaukee, WI, USA, e-mail: rrramirez@mcw.edu

David Wipf
Biomagnetic Imaging Laboratory, University of California San Francisco, San Francisco, CA, USA, e-mail: david.wipf@mrsc.ucsf.edu

Sylvain Baillet
MEG Program, Department of Neurology, Medical College of Wisconsin and Froedtert Hospital, Milwaukee, WI, USA, e-mail: sbaillet@mcw.edu

Like all imaging modalities, NSI is an endeavor that encompasses a great variety of multidisciplinary knowledge: modeling the neural electrophysiology of cell assemblies, neuroanatomy, bioelectromagnetism, measurement technology, denoising, reconstruction of time-resolved brain current flows from original sensor data, and subsequent multidimensional analysis and interpretation in the spatial, temporal, spectral, and connectivity domains.

This chapter is an attempt to assemble these otherwise disparate elements in a principled manner to provide the reader an in-depth overview of this exciting evolving field – with an emphasis on estimation techniques – that let us access functional imaging at the speed of brain.

8.1.1 Neuronal Origins of Electromagnetic Signals

The measured EM signals that are generated by the brain are thought to be due primarily to ionic current flow in the apical dendrites of cortical pyramidal neurons and their associated return (a.k.a., volume) currents throughout the head tissues, i.e., the volume conductor [62, 63].

The unique architecture of each neural cell conditions the paths taken by both the synaptically driven and intrinsic tiny intracellular currents that sum up vectorially throughout the neuron to produce the dynamic net current generated by each cell. This summation results in a significant net current at the cellular level if the dendrites are organized along a single preferential direction rather than in a radial shape. Furthermore, when multiple adjacent neurons with similar morphologies are synchronously active, their cellular net currents constructively add up to produce a group current density effect at the cell assembly level. For these reasons, assemblies of pyramidal cells in neocortical layers II/III and V are considered to be the main sources of EM surface signals detected remotely. Neurons that have dendritic arbors with closed field geometries (e.g., interneurons) are thought to produce no externally measurable EM signals [37]. However, some non-pyramidal neurons such as the Purkinje cells of the cerebellar cortex have been evidenced to generate EM signals measurable at some distance [57].

Recent quantitative investigations using realistically shaped computer models of neurons suggest that EM signals generated by neocortical columns made of as few as 50,000 pyramidal cells could be detectable outside the head and on the scalp. These models also suggest that the contribution of intracellular currents due to voltage-dependent ion channels involved in fast spiking activity might well be larger than formerly expected, which supports the experimental evidence of high-frequency brain oscillations ($>100\,\mathrm{Hz}$) detected from surface signals [54].

Although still somewhat controversial, there is cumulative evidence that activity within deeper brain structures, such as the basal ganglia, amygdala, hippocampus, brain stem, and thalamus [76, 96, 88, 42, 4], may be detected remotely. However, single neurons produce weak fields, and if the current flow is spatiotemporally incoherent (e.g., a local desynchronization) the fields end up canceling. Thus, EM

recordings are particularly suited for studying spatiotemporally coherent and locally synchronized collective neural dynamics. There is a limit to how much current density a patch of cortex can support [62], thus large amplitude fields/potentials entail distributed synchronized oscillations.

From a practical standpoint, signals may be contaminated by EM artifacts originating from the heart, muscles, eyes, and the environment. These sources of perturbation to the estimation of neural currents can be attenuated and/or corrected using appropriate denoising techniques. These include noise cancelation using filters in the temporal, spatial, and frequency domains (see Section 12.3 and [6] for a review).

8.2 Measurement Modalities

All the EM recording techniques share the important benefit of high sampling rates during acquisition (up to 5 KHz on several hundreds of channels). However, they measure different, yet closely related physical quantities at different spatial scales. In principle, the inverse modeling methods described here can be applied to data acquired using MEG, EEG, ECoG, and combinations of these measurement modalities. A prerequisite is the modeling of source currents, tissue geometry and conductivity, and sensor technology. This has yield an abundant literature in the domain of forward modeling techniques, which is reviewed in Section 8.4.4.

8.2.1 Magnetoencephalography (MEG)

In MEG, an array of sensors is used to noninvasively measure components of the magnetic vector field surrounding the head [31, 97]. The magnetic fields generated by neurons are extremely weak and range by about a billion times smaller than the Earth's static magnetic field. This low signal-to-noise ratio (SNR) challenged the early development of MEG technology. The first magnetoencephalogram was recorded with a single heavily wounded coil [12]. Not long after, the superconducting quantum interference device (SQUID) was invented [103]. This extremely sensitive magnetometer (consisting of a superconducting loop with one or two Josephson junctions), coupled to a pickup coil via a flux transformer, allowed for the first low-noise MEG recordings by the early 1970s [13]. For a thorough overview of SQUID electronics and modern integrated thin-film magnetometers and gradiometers, see [31]. Importantly, to dramatically increase the SNR, MEG measurements are acquired inside a magnetically shielded room (MSR). Current state-of-the-art systems include a large number of sensors (>300), organized as a helmet-array of magnetometers and/or gradiometers (planar or axial) that can measure spatial gradients of the magnetic field. This latter arrangement has been demonstrated to be beneficial to the SNR by attenuating environmental perturbations. Distant reference

sensors can also be used to eliminate noise and to synthesize higher order virtual gradiometers [97]. Also, cost-efficient active shielding technology has been developed to further reduce the effects of sources outside the MSR and to reduce the need for heavily shielded rooms. Recent significant progress on active shielding has allowed the installation of MEG systems in a greater variety environments, with reduced MSR bulk and weight.

8.2.2 Electroencephalography (EEG)

In EEG, an array of electrodes is placed on the scalp surface to noninvasively sample the scalar field of electric potentials relative to a reference electrode [58, 61]. EEG recording technology has progressed much since the first human recordings by Hans Berger in 1924 [8] and the later work by Edgar Adrian [1]. Due to its relative low-cost and portability, EEG has become a standard technique for clinical monitoring. Modern state-of-the-art research systems use electrode caps with as many as 256 sensors. It is sometimes considered as a bridge-technique between brain imaging modalities. Indeed, some EEG systems are used for simultaneous EEG/MEG or EEG/fMRI recordings. Research is being conducted on wireless acquisition and on dry electrode technologies that do not use conductive gel, thereby reducing preparation time.

8.2.3 Electrocorticography (ECoG)

In patients undergoing ECoG, grid or strip electrode arrays are neurosurgically placed to record the electric potential more closely to the neural sources and undistorted by the skull [40, 11]. Grid arrays have typical interelectrode distances of about 1 cm or lower. Invasive measurements of the local field potential (LFP) can be recorded by depth electrodes, electrode grids, and laminar electrodes [67, 91, 77]. Although the variations of electrical potentials captured by invasive recordings are usually considered as being locally generated (that is, within the immediate vicinity of the electrode), intracranial electrodes can pick up contributions from remotely located sources when they are strongly or coherently activated.

8.3 Data Preprocessing

Data is usually preprocessed with a variety of methods prior to localization, the purpose being correction and/or rejection of cardiac, eye, muscle, respiratory, and environmental artifacts, and the extraction of features of interest. For that purpose, data channels may be processed either sequentially or all at once. In the former, so-called

univariate case, treatments include baseline subtraction (i.e., DC offset) and band-pass filtering. Signals can also be transformed to the time-frequency/scale domains with Fourier/wavelet methods, or can be bandpass filtered and Hilbert transformed to extract instantaneous amplitude and phase information [84, 83, 29]. Filtering or time-frequency analysis is of great utility for studying wavelike activity and oscillations within specific frequency bands: slow oscillations (<1 Hz), delta (1–4 Hz), theta (5–8 Hz), alpha (9–12 Hz), mu (9–12 Hz and 18–25 Hz), spindles ($\sim$14 Hz), low beta (13–20 Hz), high beta (20–30 Hz), gamma (30–80 Hz), high gamma or omega (80–200 Hz), ripples of high-frequency oscillations (HFO, $\sim$200 Hz), and sigma bursts ($\sim$600 Hz). This great variety is a reflection of the fairly large spectrum of relevant signals of electrophysiological origin accessible to NSI, using MEG or ECoG in particular, as the spatial smearing due to the skull barrier in EEG tends to obliterate its access to higher-frequency oscillations, which are supposed to originate more locally than the slower oscillations of the neural spectrum.

The continuously and simultaneously acquired time series of all MEG, EEG, and/or ECoG channels can be concatenated to form a multivariate data array $\mathbf{B} \in \Re^{d_b \times d_t}$, where d_b is the number of measurement channels and d_t is the number of time points. This data contains correlated noise generated by physiological and environmental sources. Such perturbations may be reduced using a variety of multivariate signal processing tools such as, blind source separation, subspace projection, and machine learning methods [47,92,48,85,64,104]. The general principle consists in extracting undesired features from the data using linear transformations that either aim to project noise components away from the recordings or to unmix the data into separate components before recombining those only thought to be originating from neural sources.

The signal-space projection (SSP) algorithm and principal component analysis (PCA) are two popular techniques that use the second-order statistics of the data to estimate the spatiotemporal characteristics of noisy components [92]. SSP may be applied by default to MEG data based on the statistics of an empty MSR recording to account for the perturbations that still can get into the MSR from the environment. The denoised $\mathbf{B}$ matrix can be cut into epochs time-locked to an event (e.g., stimulus onset) for single trial analysis or averaged across epochs to extract the event-related potential and/or field (ERP/F) [44]. The ERP/F can then be localized by many different inverse methods as described below.

Alternatively, blind source separation algorithms that use higher order statistics or temporal information [e.g., infomax/maximum-likelihood independent component analysis (ICA) or second-order blind identification (SOBI)] can be applied to the entire unaveraged multivariate data time series to learn a data representation basis of sensor mixing vectors (associated with maximally independent time-courses) that can be localized separately and to reject non-brain components (i.e., denoising) [7, 46, 48, 85].

For MEG, the signal-space separation (SSS) algorithm and its temporal extension (tSSS) suggest an alternative route to the rejection of external perturbations [86]. In short, SSS builds a spatial-filter which removes the EM components in the data that are generated from outside a spherical volume encompassing the brain. This is done

by projecting the data onto a magnetic signal subspace obtained from a truncated expansion of spherical harmonic basis functions of the scalar magnetic potential.

Regardless of any transformation or averaging of the original measurements, the data to be simultaneously solved can be represented as a d_b by d_v real or complex matrix $\mathbf{B}$, which contains d_v measurement column vectors. For example, if $\mathbf{B}$ is a time series matrix, its ith column vector is the d_b-dimensional measurement vector at the ith time. But $\mathbf{B}$ needs not be a time series, it can also be any given set of vectors obtained from the data that benefit from simultaneous source localization (e.g., a subspace spanned by the data). When $d_v = 1$, the single measurement problem is recovered. This case is also used for localizing individual sensor maps obtained from a decomposition of the data (e.g., ICA).

8.4 Overview of Modeling Steps

A quick overview of several aspects of modeling, which directly affect source estimation is presented in this section. Throughout this chapter, uppercase Latin and Greek letters will be used to represent matrices, and Latin lowercase letters will be used for vectors, except when in italics, which will be used for scalars. Lowercase Greek letters will be used for vectors, scalars, and functions depending on the context. The ith element of a vector will be specified by $\mathbf{a}_i$, and the notation $\mathbf{A}_{:\,i}$ and $\mathbf{A}_{i\,:}$ will be used to refer, respectively, to the ith column vector and ith row vector of $\mathbf{A}$. Also, $\mathbf{A}_{:\,\mathbf{i}}$ and $\mathbf{A}_{\mathbf{i}\,:}$ will represent matrices made out, respectively, of the column and row vectors specified by the vector of indices $\mathbf{i}$.

8.4.1 Modeling of Neural Generators

The source model refers to the mathematical model used to approximate the primary current density within a cellular assembly. A popular source model for surface and volume distributions of neural currents is the equivalent current dipole (ECD), which approximates the current density as concentrating to a single point in space $\mathbf{r_q} = (x, y, z)^T$ as expressed by $\mathbf{j}^p(\mathbf{r}) = \mathbf{q}\delta(\mathbf{r} - \mathbf{r_q})$, where $\mathbf{j}^p(\mathbf{r})$ is the 3D primary current density vector at 3D spatial coordinates $\mathbf{r}$, and δ is the Dirac delta distribution with dipole moment $\mathbf{q} = \int \mathbf{j}^p(\mathbf{r})\mathbf{dr}$ flowing along a preferred direction as derived by the average morphology of cells within the neural ensemble [31, 6]. The popularity of the ECD source model stems from its compact description of distributed current flows using a limited number of parameters: 3 for position, 2 for orientation, and 1 for amplitude. Higher dimensional parametric source models have been proposed to describe more complex geometries of the primary current through most notably, multipolar expansions [39].

EM recordings are dependent on the total flow of currents generated by neural activity. We have so far described the modeling of the primary currents generated

within cell assemblies. These currents circulate and return through secondary volume currents (see Section 8.4.4), which depend on the geometry and conductivity properties of head tissues, as discussed in the upcoming sections.

8.4.2 Anatomical Modeling of Head Tissues and Neural Sources

To conduct subject-specific anatomical modeling, the geometry of head compartments is obtained from the analysis of the T1-weighted MRI data. Head tissues, such as gray and white matter, skull bone, scalp, cerebrospinal fluid (CSF), and fat are classified from MRI data using segmentation techniques [17]. The geometry of these components is represented using surface and volume tessellation techniques for subsequent modeling of their electromagnetic properties as discussed in Section 8.4.4.

The anatomical domain of sources can then be constrained to the gray matter volume, or rather to the cortical surfaces since NSI cannot discriminate generators from different cortical layers. With this surface approach, dipole orientations can easily be constrained to point along the normal direction of the cortical surfaces (i.e., in the direction of the apical dendrites of cortical pyramidal neurons). Non-cortical structures can be modeled as volumetric source subspaces without dipole orientations constraints.

If a subject's MRI is not available, a standardized MRI may be warped to optimally fit the subject's anatomy based on the individual's digitized head shape points. The warped brain anatomy can then be used as a standardized volumetric source space and for standardized forward modeling [16].

8.4.3 Multimodal Geometric Registration

As a prerequisite to the modeling of EM signals originating from anywhere in the source space, the sensor and source positions and orientations must be expressed in the same coordinate system. This registration process is usually done by transforming (i.e., using rigid-body translation and rotation based on anatomical landmarks) the sensor positions and orientations to the coordinate system of the MRI, where the NSI generators are modeled. Errors in the definition of the fiducial anatomical landmarks in either the NSI or MRI modality can result in poor geometrical alignment and therefore, critical errors in the modeling of generators. Improved registration can be achieved by matching a larger number of fiducial points – beyond the three typical nasion and auricular locations – such as a digitized head-shape or the locations of EEG electrodes to the skin surface extracted from MRI data. Careful alignment can help minimize geometrical registration errors within the range of under 5 mm.

8.4.4 Forward Modeling

In order to obtain an estimate of the primary current density, one needs to model the EM signals produced by both the primary (i.e., impressed) and the secondary (i.e., volume, return) current density throughout the head volume conductor, which in reality has an inhomogenous and anisotropic conductivity profile. Analytic MEG forward solutions can be computed if the volume conductor is approximated by an isotropic sphere or a set of overlapping spheres [78, 36]. The same is true for EEG but using concentric spherical shells with different isotropic conductivities. Most MEG and EEG studies assume a spherically symmetric volume conduction model. Solutions and software exist to improve the level of realism of the forward volume conduction head model, as the measured signals – especially with EEG – may have significant contributions from volume currents.

Much progress has been made toward realistic EM forward modeling using numerical techniques such as the boundary element method (BEM) and the finite element method (FEM) [32, 3, 102]. The BEM assumes a homogenous and isotropic conductivity profile through the volume of each tissue shell (e.g., brain, CSF, skull, skin), but with a conductivity inhomogeneity across the boundaries of the shells. The FEM usually also assumes homogeneity and isotropy within each tissue type, but in contrast to BEM, can also be used to model the conductivity anisotropy of white matter and that of the skull's spongiform and compact layers. Although realistic modeling exploits any available subject-specific information from MRI (e.g., T1, T2, PD, DTI) or CT, standardized BEM or FEM head models can be used as a first approximation for subjects without an MR scan [19, 16]. We should note, however, that realistic modeling is ultimately limited by the uncertainty in parameters such as the in vivo individual distribution of electrical conductivity throughout head tissues, which is yet to be accessible reliably to MRI techniques [90] and electrical impedance tomography [24].

8.4.5 Inverse Modeling

The goal of inverse modeling is to estimate the location and strengths of the sources that generated the measured EM data. As in many other problems in physics, this is a so-called *ill-posed* inverse problem, which essentially means there are an infinite number of solutions that explain the measured data equally well. The main reason is that some source configurations produce no EM signals at the sensors. This means that these silent source configurations can always be added to an existing solution without affecting the fit to the data [33]. This nonuniqueness forces us to make a priori assumptions, additional to the experimental data, to further constrain the number of feasible source patterns to one unique solution [78, 31].

These additional constraints are usually handled within the general framework of regularization, which is also common to most medical imaging applications where reconstruction of source signals of measured data is required. In NSI, these con-

straints can take many forms but are generally handled by making assumptions about the nature of the sources (e.g., number of brain areas involved, constraints on their spatial extent and the relative smoothness or sparsity of the current density, priors on source parameters and hyperparameters, priors on their anatomical locations – e.g., on the cortex – and from electrophysiology – e.g., the maximum expected amplitude of currents). Thus, the accuracy and validity of the source estimates depend to some extent on the biological correctness of the assumptions and priors adopted in the models. A recent trend in the domain of NSI research consists is considering that such priors should – to some extent – be flexible and adaptive to the data under study. The rest of this chapter focuses on presenting a variety of inverse modeling approaches. We have identified three basic approaches that encompass most of the methods that have been published so far: (1) parametric source model fitting, (2) source imaging techniques explained within a general Bayesian framework, and (3) spatial scanning and filtering through beamforming.

8.5 Parametric Dipole Modeling

One of the most common assumptions adopted to handle nonuniqueness is that the measurements were generated within a small number of brain regions that can be modeled using a limited number of ECDs. The associated estimation algorithms minimize a data-fit cost function, defined typically in the least-squares sense, in the multidimensional space of nonlinear parameters. Usually, algorithms estimate five nonlinear parameters per dipole: the x, y, and z parameters that define the dipole position, and the two angles necessary to define the dipole orientation in 3D space. However, in the MEG spherically symmetric volume conductor model only one angle (on the tangent space of the sphere) is necessary because the radial dipole component is silent, thereby reducing the dimensionality to four dimensions per dipole. The dipole amplitudes are linear parameters estimated directly from data. The dimension of the space where the cost function is minimized can be reduced further to three dimensions per dipole if the dipole orientations are allowed to be obtained linearly from the data. Technically, parametric dipole modeling is performed in the sense of a least-squares fit of a model of the data, which writes differently depending on the model of noise statistics under consideration as we shall now describe.

8.5.1 Uncorrelated Noise Model

Parametric dipole fitting algorithms, minimize a data-fit cost function such as the square of the Frobenius norm of the residual,

$$\min_{\mathbf{s}} ||\mathbf{B} - \hat{\mathbf{B}}||_F^2 = ||\mathbf{B} - \mathbf{L_s}\hat{\mathbf{J}}_\mathbf{s}||_F^2 = ||(\mathbf{I} - \mathbf{L_s}\mathbf{L_s^\dagger})\mathbf{B})||_F^2 = ||\mathbf{P}_{\mathbf{L_s}}^\perp \mathbf{B}||_F^2, \qquad (8.1)$$

where **s** refers to the set of nonlinear parameters that are optimized to minimize the data-fit cost, and the Frobenius norm of a matrix is the square root of the sum of the squares of all the elements of the matrix [80, 6]. The nonlinear parameters of the ith dipole are its position vector, $\mathbf{r}_{(i)} = (x_i, y_i, z_i)^T$, and its angle vector, $\omega_{(i)} = (\phi_i, \theta_i)^T$, which specify the dipole's location and orientation. Thus, the cost is minimized in a space of dimension $5d_d$ (or $4d_d$ for the MEG single sphere head model), where d_d is the number of dipoles in the model (i.e., the order of the model).

$\hat{\mathbf{B}}$ is the part of the data explained by the ECD generative model: $\hat{\mathbf{B}} = \mathbf{L_s}\hat{\mathbf{J}_s}$, where $\mathbf{L_s}$ is the lead field or gain matrix containing d_d d_b-dimensional column vectors called gain vectors. They are computed for d_d dipoles of unit amplitude with parameters specified in **s**. The estimated d_d by d_v current matrix, $\hat{\mathbf{J}_s} = \mathbf{L_s^{\dagger}}\mathbf{B}$, contains the moments of the d_d dipoles, where $\mathbf{L_s^{\dagger}}$ is the pseudoinverse of $\mathbf{L_s}$ [23]. Thus, the ith row vector of $\hat{\mathbf{J}_s}$ contains the moments of the dipole located at position $\mathbf{r}_{(i)}$ with orientation $\omega_{(i)}$. **I** is the d_b-dimensional identity matrix, and $\mathbf{P_{L_s}^{\perp}}$ is the orthogonal projection operator onto the null space of $\mathbf{L_s}$. Note that the gain matrix needs to be recomputed at each iteration for every new **s**.

Alternatively, the orientations of the dipoles can be obtained linearly if only the positions are optimized by including the gain vectors of all three orthogonal dipole components pointing in the (x, y, z) directions, so that $\mathbf{L_s}$ is a d_b by $3d_d$ matrix and $\mathbf{J_s}$ is a $3d_d$ by d_v matrix. For this rotating dipole model, the cost function exists in a space of $3d_d$ dimensions.

This least-squares approach is equivalent to maximum likelihood estimation of the parameters that maximize the Gaussian likelihood defined by:

$$p(\mathbf{B}|\hat{\mathbf{J}_s}, \mathbf{s}, d_d, \sigma_{\Upsilon}^2) = \left(2\pi\sigma_{\Upsilon}^2\right)^{-d_b d_v/2} \exp\left(-\frac{1}{2\sigma_{\Upsilon}^2} \left\|\mathbf{B} - \mathbf{L}\hat{\mathbf{J}_s}\right\|_F^2\right), \qquad (8.2)$$

where noise is assumed to be Gaussian and uncorrelated with scalar variance σ_{Υ}^2. The parameters $\mathbf{s}^{(\mathrm{ml})}$ and $\mathbf{J_s}^{(\mathrm{ml})}$ that maximize the likelihood or equivalently minimize the negative log likelihood at convergence are the maximum likelihood estimates of the dipole positions, orientations, and amplitudes.

8.5.2 Correlated Noise Model

In the presence of correlated noise, a modified cost function can be minimized:

$$\min_{\mathbf{s}} \left\|\Sigma_{\Upsilon}^{-1/2}\left(\mathbf{B} - \mathbf{L_s}\hat{\mathbf{J}_s}\right)\right\|_F^2 = \mathrm{tr}\left(\left(\mathbf{B} - \mathbf{L}\hat{\mathbf{J}_s}\right)^T \Sigma_{\Upsilon}^{-1}\left(\mathbf{B} - \mathbf{L}\hat{\mathbf{J}_s}\right)\right), \qquad (8.3)$$

where $\Sigma_{\Upsilon}^{-1/2}$ is a whitening matrix obtained by taking the square root inverse of the noise covariance matrix, Σ_{Υ} [78]. This solution again is equivalent to a maximum likelihood estimate of the parameters using a Gaussian likelihood noise model defined by:

$$p(\mathbf{B}|\hat{\mathbf{J}}_{\mathbf{s}}, \mathbf{s}, d_d, \Sigma_{\Upsilon}) = \frac{(2\pi)^{-d_b d_v/2}}{|\Sigma_{\Upsilon}|^{1/2}} \exp\left(-\frac{1}{2} \mathrm{tr}\left(\Upsilon^T \Sigma_{\Upsilon}^{-1} \Upsilon\right)\right), \qquad (8.4)$$

where $\Upsilon = \mathbf{B} - \mathbf{L}\hat{\mathbf{J}}_{\mathbf{s}}$ is the residual data noise not explained by the model, and noise is assumed to be Gaussian and correlated.

8.5.3 Global Minimization

These cost functions are usually minimized using nonlinear optimization algorithms (e.g., Nelder–Meade downhill simplex, Levenberg–Marquardt). Unfortunately, when the number of dipoles is increased (e.g., $d_d > 1$), the profile of the cost functional has many local minima. Furthermore, it should be noted that by adding a spatial term to the data-fit cost function, dipoles can be constrained to reside as close as desired to the gray matter volume. However, such spatial penalties can introduce even more local minima problems. Robust global minimization can theoretically be achieved using computationally intensive algorithms such as simulated annealing, multistart simplex algorithms, or genetic algorithms [35, 93], but the minimization over continuous parameters makes the endeavor unpractical for source models with over a handful of ECDs.

Alternatively, instead of selecting a point estimate, one can use Markov Chain Monte Carlo (MCMC) algorithms to make Bayesian inferences about the number of sources and their spatial extents, and to compute probabilistic maps of activity anatomically constrained to gray matter [82, 9].

As a side note, it is important to distinguish the cost function from the optimization algorithm. Although the standard costs for dipole fitting have many local minima, other costs, like for example, the negative log marginal likelihood (see Section 8.6.4), have fewer local minima and can also be minimized with nonlinear optimization algorithms.

8.6 Source Space-Based Distributed and Sparse Methods

Instead of performing low-dimensional nonlinear optimization, one can assume dipoles at all possible candidate locations of interest within a grid and/or mesh called the *source space* (e.g., source points in gray matter), and then solve the underdetermined linear system of equations

$$\mathbf{B} = \mathbf{L}\mathbf{J} + \Upsilon \qquad (8.5)$$

for $\hat{\mathbf{J}}$, the d_j by d_v estimated current density matrix (d_j being the number of dipole components throughout the source space). The lead field matrix $\mathbf{L} \in \Re^{d_b \times d_j}$ linearly maps the current space onto the measurement space. Υ is the d_b by d_v noise matrix

usually assumed to be Gaussian. Since there is no unique solution to this problem, additional priors are needed to find solutions of interest. These algorithms can be best presented from the standpoint of a general Bayesian framework that makes explicit the source prior assumptions using probability density functions (pdfs). Bayes theorem

$$p(\mathbf{J}|\mathbf{B}, \mathcal{H}) = \frac{p(\mathbf{B}|\mathbf{J}, \mathcal{H})p(\mathbf{J}|\mathcal{H})}{p(\mathbf{B}|\mathcal{H})} \tag{8.6}$$

states that the posterior probability of $\mathbf{J}$ given the measurements $\mathbf{B}$ and hypothesis or Bayesian model $\mathcal{H}$ (consisting of all implicit assumptions and parameters) is equal to the likelihood of $\mathbf{J}$ multiplied by the marginal prior probability of $\mathbf{J}$, divided by the normalizing constant of the posterior called the evidence for $\mathcal{H}$ which is defined by

$$p(\mathbf{B}|\mathcal{H}) = \int p(\mathbf{B}|\mathbf{J}, \mathcal{H})p(\mathbf{J}|\mathcal{H})dJ. \tag{8.7}$$

8.6.1 Bayesian Maximum a Posteriori (MAP) Estimates

A Gaussian likelihood model is usually assumed,

$$p(\mathbf{B}|\mathbf{J}, \mathcal{H}) = \left(2\pi\sigma_\Upsilon^2\right)^{-d_b d_v/2} \exp\left(-\frac{1}{2\sigma_\Upsilon^2}\left\|\mathbf{B} - \mathbf{L}\hat{\mathbf{J}}\right\|_F^2\right), \tag{8.8}$$

together with a prior pdf, which assigns a probability density to every possible estimate before the measurement data has been taken into account. A very useful family of prior models can be obtained with the generalized Gaussian marginal pdfs

$$p(\mathbf{J}|\mathcal{H}) \propto \exp\left(-\mathrm{sgn}(p)\sum_{i=1}^{d_n}\left\|\hat{\mathbf{J}}_{\mathbf{i}:}\right\|_q^p\right), \tag{8.9}$$

where d_n is the total number of source points, p specifies the shape of the pdf or equivalently the *p-norm*-like measure to be minimized, which controls the sparsity of the estimate, and q specifies the norm of $\hat{\mathbf{J}}_{\mathbf{i}:}$ (the matrix containing the row vectors of $\mathbf{J}$ associated with the ith source point as indexed by $\mathbf{i}$), which here is assumed to be the Frobenius norm. The signum function, $\mathrm{sgn}(p)$, takes values of 1, 0, or -1 for positive, zero, or negative p, respectively. However, the special case of $p = 0$ (i.e., the so-called zero norm) rather implies minimizing the number of $\hat{\mathbf{J}}_{\mathbf{i}:}$'s with nonzero Frobenius norms. Other priors are also possible for MAP estimation.

Since the normalizing constant $p(\mathbf{B}|\mathcal{H})$ does not affect the location of the posterior mode, it can be ignored, and thus the MAP point estimate can be computed by

$$\hat{\mathbf{J}}^{(\mathrm{map})}, \hat{\sigma}_\Upsilon^{2(\mathrm{map})} = \arg\max_{\hat{\mathbf{J}}, \sigma_\Upsilon^2} \log p(\hat{\mathbf{J}}|\mathbf{B}) \propto \log p(\mathbf{B}|\hat{\mathbf{J}}, \mathcal{H}) + \log p(\hat{\mathbf{J}}|\mathcal{H}). \tag{8.10}$$

Maximizing the log posterior rather than the posterior, illustrates how these MAP approaches are equivalent to algorithms that minimize *p-norm*-like measures

$$\min_{\hat{\mathbf{J}},\sigma_\Upsilon^2} (2\sigma_\Upsilon^2)^{-1} \left\| \mathbf{B} - \mathbf{L}\hat{\mathbf{J}} \right\|_F^2 + \mathrm{sgn}(p) \sum_{i=1}^{d_n} \left\| \hat{\mathbf{J}}_{\mathbf{i}:} \right\|_q^p + \frac{d_b d_v}{2} \log(2\pi\sigma_\Upsilon^2). \quad (8.11)$$

The noise variance σ_Υ^2 is equivalent to the λ^2 parameter used in Tikhonov regularization, which can be fixed to a value, stabilized (e.g., using the empirical L-curve or generalized cross validation methods [56]), learned from the data, or adjusted to achieve a desired representation error ε using the discrepancy principle,

$$\left\| \mathbf{B} - \mathbf{L}\hat{\mathbf{J}} \right\|_F^2 = \left\| \Upsilon \right\|_F^2 = \varepsilon. \quad (8.12)$$

MAP estimates using a Gaussian prior ($p = 2$) are equivalent to noise-regularized minimum-l_2-norm solutions, often called minimum-norm estimates (MNE),

$$\hat{\mathbf{J}} = \mathbf{L}^T \left(\mathbf{L}\mathbf{L}^T + \sigma_\Upsilon^2\mathbf{I} \right)^{-1} \mathbf{B}, \quad (8.13)$$

which are widely used in the field [2,98,21]. This basic model assumes homoscedastic uncorrelated noise. Heteroscedastic uncorrelated noise can be modeled by replacing $\sigma_\Upsilon^2\mathbf{I}$ with a diagonal matrix containing the estimated variance of each channel on the diagonal. To suppress correlated noise, the matrix $\sigma_\Upsilon^2\mathbf{I}$ can be replaced with a non-diagonal noise covariance matrix Σ_Υ obtained from the measurements, which is equivalent to performing whitening.

The point estimates obtained with the Gaussian prior are spatially distributed and suffer from depth bias (i.e., deep source distributions tend to mislocalize to more superficial source points). Many different types of weighted minimum-l_2-norm algorithms can be used to partially compensate for this depth bias by assuming an a priori source covariance other than the identity matrix, which is the assumed source covariance in the standard minimum-l_2-norm approach. In its more general form, the inverse operator using a Gaussian prior is given by

$$\Omega^{(\mathrm{map}-L_2)} = \Sigma_\mathbf{J}\mathbf{L}^T \left(\mathbf{L}\Sigma_\mathbf{J}\mathbf{L}^T + \Sigma_\Upsilon \right)^{-1}, \quad (8.14)$$

where $\Sigma_\mathbf{J}$ is the source covariance matrix. Depth bias compensation is often implemented by setting $\Sigma_\mathbf{J} = \mathbf{W}\mathbf{W}^T$, where $\mathbf{W}$ is a diagonal matrix [e.g., $\mathbf{W} = \mathrm{diag}\left(\|\mathbf{L}_{:\mathbf{i}}\|_2^{-1} \right)$, $\mathbf{W} = \mathrm{diag}\left(\|\mathbf{L}_{:\mathbf{i}}\|_2^{-1/2} \right)$, 3D Gaussian function, or fMRI priors] [38, 20]. More generally, and to include the case of unconstrained dipole orientations, the source covariance matrix can be defined as $\Sigma_\mathbf{J} = \mathrm{diag}\left(\|\mathbf{L}_{:\mathbf{i}}\|_F^{-2\kappa} \right)$, where $\mathbf{L}_{:\mathbf{i}}$ is the gain matrix for the ith source point containing one column per dipole component (indexed by the vector $\mathbf{i}$), and this value is assigned to all variance diagonal elements corresponding to the ith source point. A κ value between 0.5 and 0.8 is usually adopted to avoid overcompensating with a full normalization ($\kappa = 1$). Non-diagonal $\Sigma_\mathbf{J}$ matrices can be used to incorporate source covariance and smoothness

constraints. For example, in the low resolution brain electromagnetic tomography (LORETA) method (i.e., spatial Laplacian minimization), $\Sigma_{\mathbf{J}} = \left(\mathbf{W}^T\mathbf{D}^T\mathbf{D}\mathbf{W}\right)^{-1}$, where $\mathbf{W} = \text{diag}\left(\|\mathbf{L}_{\cdot\,\mathbf{i}}\|_2\right)$, and $\mathbf{D}$ is the discrete spatial Laplacian operator [66].

To obtain more focal estimates, MAP estimation can be performed using super-Gaussian priors such as the Laplacian pdf, which is equivalent to obtaining minimum-l_1-norm solutions, often called minimum current estimates (MCE) [48, 94]. These are traditionally computed using linear programming, but can alternatively be obtained more efficiently using an expectation maximization (EM) algorithm by parameterizing the prior as a Gaussian scale mixture. This approach can be used to find MAP estimates with generalized Gaussian prior pdfs defined by $p \leq 2$ (the Laplacian being the special case $p = 1$).

These source priors can be formulated within a hierarchical Bayes framework, in which each $\mathbf{J}_{\mathbf{i}\,:}$ has a Gaussian prior, $p(\mathbf{J}_{\mathbf{i}\,:}|\alpha_i^{-1}) = \mathcal{N}(\mathbf{J}_{\mathbf{i}\,:}|0, \alpha_i^{-1}\mathbf{I})$, with zero mean, and covariance $\alpha_i^{-1}\mathbf{I}$, and each α_i^{-1} has a hyperprior $p(\alpha_i^{-1}|\gamma)$ that controls the shape of the pdf. The variances are integrated out to obtain the prior

$$p(\mathbf{J}_{\mathbf{i}\,:}|\gamma) = \int p(\mathbf{J}_{\mathbf{i}\,:}|\alpha_i^{-1})p(\alpha_i^{-1}|\gamma)d\alpha_i^{-1}. \tag{8.15}$$

Different priors can be obtained by assuming different hyperpriors. For example, the Laplacian prior is obtained with an exponential hyperprior $p(\alpha_i^{-1}|\gamma) = \frac{\gamma}{2}\exp\left(\frac{\gamma}{2}\alpha_i^{-1}\right)$, and the Jeffreys prior $p(\mathbf{J}_{\mathbf{i}\,:}) = \|\mathbf{J}_{\mathbf{i}\,:}\|_F^{-1}$ is obtained with the noninformative Jeffreys hyperprior $p(\alpha_i^{-1}) = \alpha_i$, which has the advantage of being scale invariant and parameter free.

The EM algorithm minimizes the negative log posterior by alternating between two steps. In the E-step, the conditional expectation of the inverse source variances at the kth iteration, $\mathbf{A}^{(k)} = \text{diag}(\alpha^{(k)})$, given $\mathbf{B}$, $\mathbf{J}^{(k)}$, and $\sigma_\Upsilon^{2(k)}$ is computed

$$E[\alpha_i^{(k)}|\mathbf{J}^{(k)}, \sigma_\Upsilon^{2(k)}] = \left(\frac{1}{d_v}\left\|\hat{\mathbf{J}}_{\mathbf{i}\,:}^{(k)}\right\|_F^2\right)^{\frac{p-2}{2}}. \tag{8.16}$$

In the M-step, the noise variance and the current density estimates are computed

$$\hat{\sigma}_\Upsilon^{2(k+1)} = \left(\left\|\mathbf{B} - \mathbf{L}\hat{\mathbf{J}}^{(k)}\right\|_F^2 / d_b d_v\right)^{1-\frac{p}{2}}, \tag{8.17}$$

$$\hat{\mathbf{J}}^{(k+1)} = \Sigma_{\mathbf{J}}^{(k)}\mathbf{L}^T\left(\mathbf{L}\Sigma_{\mathbf{J}}^{(k)}\mathbf{L}^T + \Sigma_\Upsilon^{(k+1)}\right)^{-1}\mathbf{B}, \tag{8.18}$$

where $\Sigma_{\mathbf{J}}^{(k)} = E[\mathbf{A}^{(k)}|\mathbf{J}^{(k)}, \sigma_\Upsilon^{2(k)}]^{-1}$ and $\Sigma_\Upsilon^{(k+1)} = \hat{\sigma}_\Upsilon^{2(k+1)}\mathbf{I}$ are the source and noise covariance matrices. Note that the noise variance update rule implements MAP estimation with a non-Gaussian prior on $\hat{\sigma}_\Upsilon^2$. In practice, the discrepancy principle is often used based on some reasonable expected representation error to avoid under-regularizing. When $\hat{\mathbf{J}}^{(k+1)} = \hat{\mathbf{J}}^{(k)}$ and $\sigma_\Upsilon^{2(k+1)} = \sigma_\Upsilon^{2(k)}$, the algorithm has converged and the MAP inverse operator for this generalized Gaussian prior (e.g., $p = 1$) can

be computed by

$$\Omega^{(\text{map}-L_p)} = \Sigma_{\mathbf{J}}^{(k)} \mathbf{L}^T \left(\mathbf{L} \Sigma_{\mathbf{J}}^{(k)} \mathbf{L}^T + \Sigma_{\Upsilon}^{(k)} \right)^{-1}. \tag{8.19}$$

In practice, the iterations are usually carried out until a threshold of change is reached (e.g., $\left\| \hat{\mathbf{J}}^{(k+1)} - \hat{\mathbf{J}}^{(k)} \right\|_2 \leq \varepsilon$). Also, to accelerate convergence one usually truncates source points from all equations for which the current is smaller than a very small threshold, but this can have a negative effect on the minimization of the cost function. If the cost is not minimized at one iteration due to this thresholding, a smaller threshold value should be used. Also, to compensate for depth bias, the lead field matrix should be weighted as explained earlier in the context of weighted minimum-l_2-norm algorithms, but in this case it should be weighted before the start of MAP optimization, and the final solution can be unweighted after convergence by multiplying with the original weight factors.

These update equations are equivalent to a generalized form of the FOCal Undetermined System Solver (FOCUSS) algorithm, which was developed as a recursive weighted minimum-norm algorithm for $p = 0$, but was later derived as a Bayesian MAP algorithm using generalized Gaussian prior pdfs [26, 25, 75, 74, 14, 69]. For the case of $p = 0$, truncation of the rows of $\mathbf{J}$ with smallest norms is usually implemented so that the minimization involves the count of nonzero rows. When $p = -2$, the magnetic field tomography (MFT) algorithm is recovered if the update rule is based on the current modulus, there is only one iteration, and the a priori weight matrix is a 3D Gaussian used for depth bias compensation [38, 76, 87]. If one is not sure whether one should use a Gaussian or Laplacian prior, one can use MCMC methods to learn which l_p-norm is optimal for that particular data set [5].

To simultaneously identify the generators of a long data time series, the matrix $\mathbf{B}\mathbf{B}^T$ can be decomposed efficiently using the SVD, and $\mathbf{B}$ in (8.17) and (8.18) can be replaced with the matrix $\mathbf{U}\mathbf{S}^{1/2}$, where $\mathbf{U}$ and $\mathbf{S}$ are the left singular vectors and singular values matrices, respectively [99].

8.6.2 Dynamic Statistical Parametric Mapping (dSPM)

Another approach directly related to the MNE is the noise normalized dynamic statistical parametric mapping (dSPM) technique, which normalizes the MNE by the noise sensitivity at each location, thereby producing statistical activity maps [15, 41]. This extra step helps compensate for depth bias. First, the linear inverse operator is computed by (8.14). This operator is equivalent to that used in Wiener filtering or in weighted minimum-l_2-norm estimation assuming correlated noise. Then the noise-normalized operator is computed, which in the case of fixed dipole orientations yields:

$$\Omega^{(\text{dspm})} = \text{diag}\,(\mathbf{v})^{-1/2}\, \Omega^{(\text{map}-L_2)}, \tag{8.20}$$

where $\mathbf{v} = \mathrm{diag}\left(\Omega^{(\mathrm{map}-L_2)}\Sigma_{\Upsilon}\Omega^{T(\mathrm{map}-L_2)}\right)$. Note that dSPM performs noise normalization after the inverse operator $\Omega^{(\mathrm{map}-L_2)}$ has been computed. Thus, the noise-normalized source activity estimates are given by

$$\hat{\mathbf{S}}^{(\mathrm{dspm})} = \Omega^{(\mathrm{dspm})}\mathbf{B} = \mathrm{diag}\left(\mathbf{v}\right)^{-1/2}\hat{\mathbf{J}}^{(\mathrm{map}-L_2)}. \tag{8.21}$$

More generally, to include the case where dipole orientation constraints are not enforced, the noise-normalized dSPM time series of source power at the ith source point is computed as

$$\hat{\mathbf{S}}_{i:}^{2(\mathrm{dspm})} = \mathrm{diag}\left(\hat{\mathbf{J}}_{\mathbf{i}:}^{T(\mathrm{map}-L_2)}\hat{\mathbf{J}}_{\mathbf{i}:}^{(\mathrm{map}-L_2)}\right)^{T}/\mathrm{tr}\left(\Omega_{\mathbf{i}:}^{(\mathrm{map}-L_2)}\Sigma_{\Upsilon}\Omega_{\mathbf{i}:}^{T(\mathrm{map}-L_2)}\right). \tag{8.22}$$

8.6.3 Standardized Low Resolution Brain Electromagnetic Tomography (sLORETA)

An alternative approach for depth-bias compensation and source standardization is the sLORETA technique [65]. In contrast to the dSPM method, the MNE is modified by the resolution matrix, $\mathbf{R} = \Omega^{(\mathrm{map}-L_2)}\mathbf{L}$, that is associated with the inverse and forward operators: $\Omega^{(\mathrm{map}-L_2)}$ and $\mathbf{L}$. For fixed dipole orientations, the pseudo-statistics of power and absolute activation at the ith source point for a time slice are respectively given by

$$\varphi_i = \frac{\hat{\mathbf{j}}_i^2}{\mathbf{R}_{ii}} \text{ and } \sqrt{\varphi_i}, \tag{8.23}$$

and the standardized sLORETA inverse operator can be written as

$$\Omega^{(\mathrm{sloreta})} = \mathrm{diag}(\mathbf{r})^{-1/2}\Omega^{(\mathrm{map}-L_2)}, \tag{8.24}$$

where $\mathbf{r} = \mathrm{diag}(\mathbf{R})$. Thus, the sLORETA activity time-series is computed by

$$\hat{\mathbf{S}}^{(\mathrm{sloreta})} = \Omega^{(\mathrm{sloreta})}\mathbf{B} = \mathrm{diag}\left(\mathbf{r}\right)^{-1/2}\hat{\mathbf{J}}^{(\mathrm{map}-L_2)}. \tag{8.25}$$

More generally, for the case of no dipole orientation constraints, the sLORETA standardized source power time series at the ith source point is computed as

$$\hat{\mathbf{S}}_{i:}^{2(\mathrm{sloreta})} = \mathrm{diag}\left(\hat{\mathbf{J}}_{\mathbf{i}:}^{T(\mathrm{map}-L_2)}\left(\mathbf{R}_{\mathbf{ii}}\right)^{-1}\hat{\mathbf{J}}_{\mathbf{i}:}^{(\mathrm{map}-L_2)}\right)^{T}. \tag{8.26}$$

Interestingly, the sLORETA algorithm is similar to the first step of the sparse Bayesian learning (SBL) algorithm explained in the next section.

8.6.4 Sparse Bayesian Learning (SBL) and Automatic Relevance Determination (ARD)

Sparse Bayesian learning (SBL) uses the same Gaussian likelihood model defined in (8.8) and also uses a hierarchical Bayes formulation similar to that explained for MAP estimation, but instead of integrating out the hyperparameters as in parameter MAP estimation, in SBL we integrate out the parameters [45, 44, 55, 89, 100, 79, 69, 99, 101, 60]. Thus, instead of finding point estimates at the posterior modes using fixed priors, it performs the evidence maximization procedure to learn adaptive hyperparameters from the data itself. SBL assumes an automatic relevance determination (ARD) prior for the current density defined as

$$p(\mathbf{J}|\alpha) = \prod_{i=1}^{d_\alpha} \mathcal{N}(0, \alpha_i^{-1}\mathbf{I}), \tag{8.27}$$

where α is a vector of hyperparameters or precisions (i.e., inverse source variances), d_α is the number of hyperparameters, and each $\mathbf{J_{i\,:}}$ has a zero-mean Gaussian prior with covariance $\alpha_i^{-1}\mathbf{I}$. The inverse source and noise variances have Gamma hyperpriors,

$$p(\alpha) = \prod_{i=1}^{d_\alpha} \text{Gamma}(\alpha_i|a,b), \tag{8.28}$$

$$p(\sigma_\Upsilon^{-2}) = \text{Gamma}(\sigma_\Upsilon^{-2}|c,d), \tag{8.29}$$

where a, b, c, and d are the degrees of freedom parameters of the Gamma distributions of α and σ_Υ^{-2} given by $\text{Gamma}(\alpha|a,b) = \Gamma(a)^{-1}b^a\alpha^{a-1}e^{-b\alpha}$ with $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$. The Gamma hyperprior results in a student-t prior for the source parameters. However, to avoid tuning the hyperprior, the Gamma distribution parameters can be set to a small number (e.g., $a = b = c = d = 10^{-4}$) to make these priors noninformative (i.e., flat in log space, as is common for scale parameters), or they can be made exactly zero, in which case we obtain the Jeffreys prior, which results in scale invariance.

SBL is an important alternative because the posterior mode may not be representative of the full posterior, and thus, a better point estimate may be obtained, the posterior mean, by tracking the posterior probability mass. In the case of the Jeffreys prior, this is achieved by finding the maximum likelihood hyperparameters $\alpha^{(\text{ml})}$ and $\sigma_\Upsilon^{2(\text{ml})}$ that maximize a tractable Gaussian approximation of the *evidence* of the hyperparameters, also known as the type-II likelihood or marginal likelihood

$$\hat{\alpha}^{(\text{ml})}, \hat{\sigma}_\Upsilon^{2(\text{ml})} = \arg\max_{\alpha,\sigma_\Upsilon^2} p(\mathbf{B}|\alpha, \sigma_\Upsilon^2) = \int p(\mathbf{B}|\mathbf{J})p(\mathbf{J}|\alpha, \sigma_\Upsilon^2)dJ = \mathcal{N}(0, \hat{\Sigma}_\mathbf{B}), \tag{8.30}$$

or equivalently by minimizing the negative log marginal likelihood

$$\hat{\alpha}^{(\text{ml})}, \hat{\sigma}_\Upsilon^{2(\text{ml})} = \arg \min_{\alpha, \sigma_\Upsilon^2} - \log p(\mathbf{B}|\alpha, \sigma_\Upsilon^2) = d_v \log \left| \hat{\Sigma}_\mathbf{B} \right| + tr \left(\mathbf{B}^T \hat{\Sigma}_\mathbf{B}^{-1} \mathbf{B} \right), \quad (8.31)$$

where $\hat{\Sigma}_\mathbf{B} = \mathbf{L}\Sigma_\mathbf{J}\mathbf{L}^T + \Sigma_\Upsilon$ is the model data covariance, and $\Sigma_\mathbf{J} = \text{diag}(\alpha)^{-1}$ is the prior source covariance matrix. The noise covariance matrix, Σ_Υ, can be assumed to be a multiple of the identity matrix (e.g., $\sigma_\Upsilon^2 \mathbf{I}$, where σ_Υ^2 is the noise variance, a hyperparameter that can also be learned from the data), or can be empirically obtained from the measurements.

In the case of Gamma hyperpriors (i.e., a, b, c, and d are nonzero), the posterior probability of the log hyperparameters given the data, that is, the product of the marginal likelihood and the hyperprior, $p(\mathbf{B}|\log \alpha, \log \sigma_\Upsilon^2)p(\log \alpha, \log \sigma_\Upsilon^2)$, is maximized, or equivalently the negative log posterior is minimized,

$$\log \hat{\alpha}^{(\text{map})}, \log \hat{\sigma}_\Upsilon^{2(\text{map})} = \arg \min_{\log \alpha, \log \sigma_\Upsilon^2} - \log p(\mathbf{B}|\log \alpha, \log \sigma_\Upsilon^2)p(\log \alpha, \log \sigma_\Upsilon^2)$$

$$= \arg \min_{\log \alpha, \log \sigma_\Upsilon^2} d_v \log \left| \hat{\Sigma}_\mathbf{B} \right| + tr \left(\mathbf{B}^T \hat{\Sigma}_\mathbf{B}^{-1} \mathbf{B} \right) +$$

$$\sum_{i=1}^{d_\alpha} (a \log \alpha_i - b\alpha_i) + c \log \sigma_\Upsilon^{-2} - d\sigma_\Upsilon^{-2}. \quad (8.32)$$

Evidence maximization is usually achieved by using Expectation–Maximization update rules

$$\alpha_i^{(k+1)} = (1+2a) \left(\frac{1}{d_v d_r} \left\| \Omega_{\mathbf{i}:}^{(k)} \mathbf{B} \right\|_F^2 + \frac{1}{d_r} tr \left(\left(\mathbf{I} - \Omega_{\mathbf{i}:}^{(k)} \mathbf{L}_{:\mathbf{i}} \right) \alpha_i^{-1(k)} \right) + 2b \right)^{-1}, \quad (8.33)$$

$$\sigma_\Upsilon^{2(k+1)} = \left(\frac{1}{d_v} \left\| \mathbf{B} - \mathbf{L}\hat{\mathbf{J}}^{(k)} \right\|_F^2 + \sigma_\Upsilon^{2(k)} tr(\mathbf{R}^{(k)}) + 2d \right) / (d_b + 2c), \quad (8.34)$$

or alternatively using the MacKay gradient update rules

$$\alpha_i^{(k+1)} = \left(\frac{1}{d_r} tr \left(\left(\mathbf{I} - \Omega_{\mathbf{i}:}^{(k)} \mathbf{L}_{:\mathbf{i}} \right) \alpha_i^{-1(k)} \right) + 2a \right) / \left(\frac{1}{d_v d_r} \left\| \Omega_{\mathbf{i}:}^{(k)} \mathbf{B} \right\|_F^2 + 2b \right), \quad (8.35)$$

$$\sigma_\Upsilon^{2(k+1)} = \left(\frac{1}{d_v} \left\| \mathbf{B} - \mathbf{L}\hat{\mathbf{J}}^{(k)} \right\|_F^2 + 2d \right) / \left(d_b - tr(\mathbf{R}^{(k)}) + 2c \right), \quad (8.36)$$

where $\mathbf{L}_{:\mathbf{i}}$ is a matrix with column vectors from $\mathbf{L}$ that are controlled by the same ith hyperparameter, d_r is the rank of $\mathbf{L}_{:\mathbf{i}}\mathbf{L}_{:\mathbf{i}}^T$, $\Omega_{\mathbf{i}:}^{(k)} = \alpha_i^{-1(k)} \mathbf{L}_{:\mathbf{i}}^T \left(\hat{\Sigma}_\mathbf{B}^{(k)} \right)^{-1}$, and $\mathbf{R}^{(k)} = \Sigma_\mathbf{J}^{(k)} \mathbf{L}^T \left(\hat{\Sigma}_\mathbf{B}^{(k)} \right)^{-1} \mathbf{L}$ is the kth resolution matrix. With fixed dipole orientations $\mathbf{L}_{:\mathbf{i}}$ is a vector, but with loose orientations $\mathbf{L}_{:\mathbf{i}}$ is a d_b by three matrix. For patch source models involving dipoles within a region, $\mathbf{L}_{:\mathbf{i}}$ is a matrix containing all gain vectors associated with the local patch of cortex. The gradient update rule is much faster than the EM rule and is similar to the update rules used in several

hybrid sLORETA/FOCUSS algorithms [81]. In practice, the discrepancy principle is often used to avoid under-regularizing. Once the optimal maximum likelihood or maximum a posteriori hyperparameters have been learned (i.e., they have stopped changing), the SBL inverse operator can be expressed as

$$\Omega^{(sbl)} = \Sigma_{\mathbf{J}}^{(sbl)} \mathbf{L}^T \left(\hat{\Sigma}_{\mathbf{B}}^{(sbl)} \right)^{-1}, \qquad (8.37)$$

and the posterior mean is given by

$$\hat{\mathbf{J}} = E\left[\mathbf{J} | \mathbf{B}; \Sigma_{\mathbf{J}}^{(sbl)} \right] = \Omega^{(sbl)} \mathbf{B}. \qquad (8.38)$$

It is important to note that many useful SBL variants can be obtained by the reparametrization of the source covariance matrix $\Sigma_{\mathbf{J}} = \sum_{i=1}^{d_\alpha} \mathbf{C}_i \alpha_i^{-1}$. In fact, if only a few hyperparameters are used, and each controls many source points, then the parametrization cannot support sparse estimates. For example, in the restricted maximum likelihood (ReML) algorithm one of the source covariance components is the identity matrix, which is controlled by a single hyperparameter [18, 68, 50, 99]. In standard SBL, $\mathbf{C}_i = \mathbf{e}_{(i)} \mathbf{e}_{(i)}^T$, where $\mathbf{e}_{(i)}$ is a vector with zeros everywhere except at the ith element, where it is one. This delta function parametrization can be extended to box car functions in which $\mathbf{e}_{(i)}$ takes a value of 1 for all three dipole components or for a patch of cortex. Alternatively, each $\mathbf{e}_{(i)}$ can be substituted by a geodesic basis function $\psi_{(i)}$ (e.g., a 2D Gaussian current density basis function) centered at the ith source point and with some spatial standard deviation [79, 70]. This approach can be extended to a multiscale algorithm, in which the source covariance matrix is composed of components across many possible spatial scales, by using multiple $\psi_{(i)}$ vectors located at the ith source point but with different spatial standard deviations [70, 73, 71, 72]. This approach can be used to estimate the spatial extent of distributed sources by using a mixture model of geodesic Gaussian distributions at different spatial scales. Such multiscale approach can also be used with parameter MAP estimation [47, 72].

The problem of finding optimal hyperpriors to handle multimodal posteriors and to eliminate the use of improper priors has been dealt with by using flat hyperpriors or by introducing MCMC strategies [59, 60]. In practice, the noninformative hyperprior works well and helps avoid the problem of determining the optimal hyperprior. Finally, as explained for parameter MAP estimation, to simultaneously localize the generators of a very long time series of any length very quickly, instead of localizing the times series matrix $\mathbf{B}$, one can use the matrix $\mathbf{U}\mathbf{S}^{1/2}$, where $\mathbf{U}$ and $\mathbf{S}$ are the left singular vectors and singular values matrices of $\mathbf{B}\mathbf{B}^T$.

8.7 Spatial Scanning and Beamforming

An alternative approach to the *ill-posed* bioelectromagnetic inverse problem is to independently scan for dipoles within a grid containing candidate locations (i.e.,

source points). Here the goal is to estimate the activity at a source point or region while avoiding the cross talk from other regions so that these affect as little as possible the estimate at the region of interest.

8.7.1 Matched Filter

The simplest spatial filter, a matched filter, is obtained by normalizing the columns of the lead field matrix and transposing this normalized dictionary. The spatial filter for location $\mathbf{r}_i$ is given by

$$\Omega_{\mathbf{i}:}^{(mf)} = \frac{\mathbf{L}_{:\,\mathbf{i}}^T}{\|\mathbf{L}_{:\,\mathbf{i}}\|_F}. \tag{8.39}$$

This approach essentially projects the data onto the column vectors of the lead-field dictionary. Although this guarantees that the absolute maximum of the map corresponds to the true maximum when only one source is active and with the correct fixed dipole orientation, this filter is not recommended since these assumptions are usually not valid, and since the spatial resolution of the filter is so low given the high correlation between dictionary columns. This approach can be extended to fast recursive algorithms, such as matching pursuit and its variants, which sequentially project the data or residual to the nonused dictionary columns to obtain fast suboptimal sparse estimates.

8.7.2 Multiple Signal Classification (MUSIC)

The MUSIC algorithm was adopted from spectral analysis and modified for spatial filtering of MEG data [53, 52]. The MUSIC cost function is given by

$$M_i = \frac{\left\|\left(\mathbf{I} - \mathbf{U}_s\mathbf{U}_s^T\right)\mathbf{L}_{:\,\mathbf{i}}\right\|_2^2}{\|\mathbf{L}_{:\,\mathbf{i}}\|_2^2} = \frac{\left\|\mathbf{P}_{\mathbf{U}_s}^{\perp}\mathbf{L}_{:\,\mathbf{i}}\right\|_2^2}{\|\mathbf{L}_{:\,\mathbf{i}}\|_2^2}, \tag{8.40}$$

where $\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ is the singular value decomposition of the data, $\mathbf{U}_s$ is a matrix with the first d_s left singular vectors that form the signal subspace, and $\mathbf{L}_{:\,\mathbf{i}}$ is the gain vector for the dipole located at $\mathbf{r}_i$ and with orientation θ_i (obtained from anatomy or using the generalized eigenvalue decomposition). $\mathbf{P}_{\mathbf{U}_s}^{\perp}$ is an orthogonal projection operator onto the data noise subspace. The MUSIC map is the reciprocal of the cost function at all locations scanned. This map can be used to guide a recursive parametric dipole fitting algorithm. The number d_s is usually carefully provided by an expert user.

8.7.3 Linearly Constrained Minimum Variance (LCMV) Beamforming

Beamformers, as used in the field of NSI, are spatial filtering algorithms that scan each source point independently to pass source signals at a location of interest while suppressing interference from other regions using only the local gain vectors and the measured covariance matrix. One of the most basic and often used linear beamformers is the linearly constrained minimum variance (LCMV) beamformer, which attempts to minimize the beamformer output power subject to a unity gain constraint:

$$\min_{\Omega_{\mathbf{i}:}} \quad \mathrm{tr}\left(\Omega_{\mathbf{i}:}\Sigma_{\mathbf{B}}\Omega_{\mathbf{i}:}^{T}\right) \text{ subject to } \Omega_{\mathbf{i}:}\mathbf{L}_{:\mathbf{i}} = \mathbf{I}, \tag{8.41}$$

where $\Sigma_{\mathbf{B}}$ is the empirical data covariance matrix, $\mathbf{L}_{:\mathbf{i}}$ is the d_b by three gain matrix of the ith source point, and $\Omega_{\mathbf{i}:}$ is the three by d_b spatial filtering matrix [95]. The solution to this problem is given by

$$\Omega_{\mathbf{i}:}^{(\mathrm{lcmv})} = \left(\mathbf{L}_{:\mathbf{i}}^{T}\Sigma_{\mathbf{B}}^{-1}\mathbf{L}_{:\mathbf{i}}\right)^{-1}\mathbf{L}_{:\mathbf{i}}^{T}\Sigma_{\mathbf{B}}^{-1}. \tag{8.42}$$

The parametric source activity at the ith source point is given by $\hat{\mathbf{S}}_{\mathbf{i}:}^{(\mathrm{lcmv})} = \Omega_{\mathbf{i}:}\mathbf{B}$. This can be performed at each source point of interest to yield a score map of activity. Note that these maps, like those obtained by sLORETA and dSPM, are not real current density estimates. This beamforming approach can be expanded to a more general Bayesian graphical model that uses event timing information to model evoked responses, while suppressing interference and noise sources [104]. This approach uses a variational Bayesian EM algorithm to compute the likelihood of a dipole at each grid location.

8.7.4 Synthetic Aperture Magnetometry (SAM)

Synthetic aperture magnetometry (SAM) is a nonlinear beamformer in which an optimization algorithm is used to the find the dipole orientation at each source point that maximizes the ratio of the total source power over noise power, the pseudo-Z deviate

$$z_i = \sqrt{\frac{\Omega_{\mathbf{i}:}\Sigma_{\mathbf{B}}\Omega_{\mathbf{i}:}^{T}}{\Omega_{\mathbf{i}:}\Sigma_{\Upsilon}\Omega_{\mathbf{i}:}^{T}}} = \sqrt{\frac{p_i}{n_i}}, \tag{8.43}$$

where Σ_{Υ} is the noise covariance, usually based on some control recording or assumed to be a multiple of the identity matrix [97]. This maximization generates a scalar beamformer with optimal dipole orientations in terms of SNR. This improves the spatial resolution of SAM relative to that of LCMV beamforming. To generate statistical parametric maps between an active task period (*a*) and a control period (*c*), the so-called pseudo-T statistic can be computed as

$$t_i = \frac{p_i(a) - p_i(c)}{n_i(a) + n_i(c)}. \tag{8.44}$$

Such maps usually have more focal activities since they contrast the differences between two states. Other scalar beamformers can be implemented. For example, an anatomically constrained beamformer (ACB) can be obtained by simply constraining the dipole orientations to be orthogonal to the cortical surfaces [34].

8.7.5 Dynamic Imaging of Coherent Sources (DICS)

Beamforming can be performed in the frequency domain using the dynamic imaging of coherent sources (DICS) algorithm, whose spatial filter matrix for frequency f is given by

$$\Omega_{\mathbf{i}:}^{(\text{dics})}(f) = \left(\mathbf{L}_{:\mathbf{i}}^T \tilde{\Sigma}_{\tilde{\mathbf{B}}}(f)^{-1} \mathbf{L}_{:\mathbf{i}} \right)^{-1} \mathbf{L}_{:\mathbf{i}}^T \tilde{\Sigma}_{\tilde{\mathbf{B}}}(f)^{-1}, \tag{8.45}$$

where $\tilde{\Sigma}_{\tilde{\mathbf{B}}}(f)$ is the cross-spectral density matrix for frequency f [29]. Note that the covariance matrix has simply been replaced in (8.42) by the cross-spectral density matrices. DICS can also be used to reveal which brain regions are coherent with external reference signals (e.g., electromyogram), and to estimate cortico-cortical coherence maps.

8.7.6 Other Spatial Filtering Methods

All the spatial filtering methods explained so far depend on the gain vectors associated only with the region of interest (i.e., they do not depend on the gain vectors associated with the rest of the source space). There are other more direct approaches to spatial filtering that incorporate the gain vectors associated with both the region of interest and the rest of the source space, and that do not necessarily use the measured covariance matrix. In the Backus–Gilbert method, a different spread matrix is computed for each candidate source location [28, 27]. The goal is to penalize the side lobes of the resolution kernels (i.e., the row vectors of the resolution matrix, defined as $\mathbf{R} = \Omega \mathbf{L}$, where $\mathbf{L}$ is the lead field matrix for the entire source space and Ω is the optimized linear operator that gives the source estimates when multiplied with the data). This usually results in a wider main lobe.

In the spatially optimal fast initial analysis (SOFIA) algorithm, virtual leadfields are constructed that are well concentrated within a region of interest compared to the rest of the source space [10]. The region of interest can be moved to every source point. A similar approach is adopted in the local basis expansion (LBEX) algorithm, which solves a generalized eigenvalue problem to maximize the concentration of linear combinations of leadfields [51].

As a final remark, it should be emphasized that all of the spatial filtering algorithms presented scan one source point or local region at a time, but can be expanded

to multisource scanning protocols that search through combinations of sources. Although multisource scanning methods can recover perfectly synchronized sources (which are usually missed by single-source methods), there is no agreed protocol to scan the immense space of possible multisource configurations.

8.8 Comparison of Methods

To compare some of the methods reviewed here, we performed a retinotopic mapping of the four visual quadrants of one subject using the Elekta MEG system, which contains 204 planar gradiometers and 102 axial magnetometers. Data were processed with SSS and bandpass filtered (2–55 Hz). One hundred epochs were averaged for each quadrant. A BEM forward model was used and sources were constrained to the cortical surface of the subject.

A schematic of the black and white checkerboard visual stimuli used is shown in Fig. 8.1a, where color is used only to code for the retinotopy maps shown on inflated cortical surfaces in Fig. 8.1c, d. Figure 8.1b shows an example of the weighted minimum-l_2-norm solution (thresholded at 0.1 of absolute maximum) of the event-related field at 100 ms poststimulus onset (lower right visual quadrant). Note how this activity is very distributed. To visualize the maps for different quadrants simultaneously and contrast them, we normalized these maps by their absolute maximum, thresholded them at 0.9, and color-coded the activity based on which quadrant had the maximal activity on each source point.

Figure 8.1c shows maps produced by distributed methods (from top to bottom): (1) weighted minimum-l_2-norm ($\kappa = 0.5$); (2) dSPM ($\kappa = 0.8$); (3) sLORETA ($\kappa = 0$); and (4) matched filter. The first three had dipole orientation constraints, but the matched filter did not. Figure 8.1d shows maps produced by sparse methods (from top to bottom): (1) SBL; (2) multiscale SBL; (3) MAP with $p = 0$; (4) multiscale MAP with Laplacian prior. In contrast to the distributed estimates, the sparse estimates were not changed much by thresholding, as expected. The maps produced by the different methods show some minor differences (related to depth-bias compensation and sparsity), but all maps show the basic expected pattern for V1/V2 retinotopy. The fact that retinotopy was discriminated with the thresholded maps suggests that thresholding can be very useful for distributed estimates since these have maxima with little localization bias. Interestingly, the simple matched filter showed a clear map consistent with the V1/V2 borders.

8.9 Conclusion

The relative strengths of different localization algorithms offer an opportunity to select the most appropriate algorithm, constraints, and priors for a given experiment. If one expects only a few focal sources, then dipole fitting algorithms may be suffi-
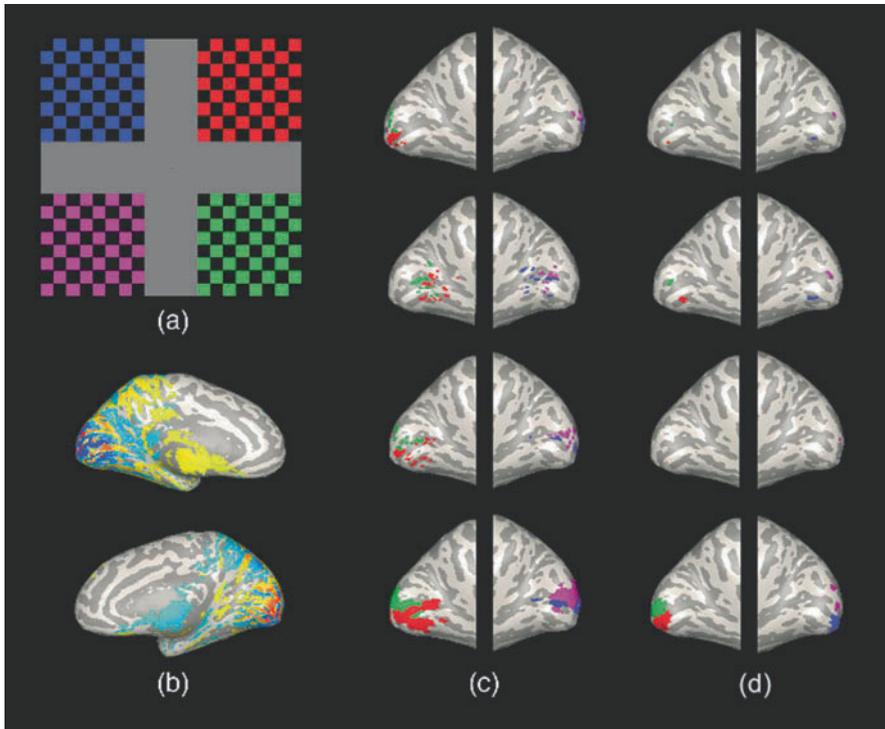
Fig. 8.1: Retinotopy of four visual quadrants. (**a**) Schematic of visual stimuli. (**b**) Weighted minimum-l_2-norm estimate thesholded only at 0.1 of maximum. (**c**) Retinotopic maps produced by distributed methods thresholded at 0.9 of maximum (colors code for visual quadrant that maximally activated that area). (**d**) Retinotopic maps produced by sparse methods. See Section 8.8.

cient. If one expects distributed sources, then distributed MAP estimation methods (e.g., using a Gaussian prior, as in MNE, dSPM, or sLORETA), spatial scans, or beamforming algorithms are appropriate. If one expects sparse sources, then SBL or MAP estimation with a Laplacian or more super-Gaussian prior may better reflect the true sources. If one expects compact distributed sources with variable levels of spatial extent, then SBL or MAP estimation (with $p \leq 1$) using a mixture model of multiscale basis functions may be optimal.

It should be noted, however, that all of these methods are expected to reveal somewhat similar functional brain maps for the same data set. If major discrepancies in terms of the brain regions involved are evidenced, this should raise a warning flag that some piece of the puzzle during source analysis has been misplaced higher up in the long chain of treatments involved. Registration with MRI is a major source of error, together with numerical errors in the computation of realistic head models using the BEM or FEM.

As in all experimental data analysis methods, one should bare in mind the balance between the sophistication of the methods involved, that should include all the prior information available to the scientist, and the robustness to deviations of the model from reality (head position, conductivity of tissues, etc.).

Multiple commercial and academic software solutions are now available to the scientist and clinician, which can help him/her grow confident of this exciting technique that images brain functions at high-temporal resolution.

References

1. Adrian, E., Mathews, B. The Berger rhythm: Potential changes from the occipital lobes in man. Brain **57,** 355–385 (1934)
2. Ahlfors, S.P., Ilmoniemi, R.J., Hamalainen, M.S. Estimates of visually evoked cortical currents. Electroencephalogr Clin Neurophysiol **82**(3), 225–236 (1992)
3. Akalin-Acar, Z., Gencer, N.G. An advanced boundary element method (BEM) implementation for the forward problem of electromagnetic source imaging. Phys Med Biol **49**(21), 5011–5028 (2004)
4. Attal, Y., Bhattacharjee, M., Yelnik, J., Cottereau, B., Lefvre, J., Okada, Y., Bardinet, E., Chupin, M., Baillet, S. Modeling and detecting deep brain activity with MEG & EEG. Conf Proc IEEE Eng Med Biol Soc, 4937–4940 (2007)
5. Auranen, T., Nummenmaa, A., Hamalainen, M.S., Jaaskelainen, I.P., Lampinen, J., Vehtari, A., Sams, M. Bayesian analysis of the neuromagnetic inverse problem with lp-norm priors. NeuroImage **26**(3), 870–884 (2005)
6. Baillet, S., Mosher, J.C., Leahy, R.M. Electromagnetic brain mapping. IEEE Signal Process Mag **18**(6), 14–30 (2001)
7. Bell, A.J., Sejnowski, T.J. An information-maximization approach to blind separation and blind deconvolution. Neural Comput **7**(6), 1129–1159 (1995)
8. Berger, H. Über das Elektroenkephalogramm des Menschen. Archiv für Psychiatrie und Nervenkrankheiten **87**, 527–570 (1929)
9. Bertrand, C., Ohmi, M., Suzuki, R., Kado, H. A probabilistic solution to the MEG inverse problem via MCMC methods: The reversible jump and parallel tempering algorithms. IEEE Trans Biomed Eng **48**(5), 533–542 (2001)
10. Bolton, J.P.R., Gross, J., Liu, A.K., Ioannides, A.A. SOFIA: Spatially optimal fast initial analysis of biomagnetic signals. Phys Med Biol **44**, 87–103 (1999)
11. Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., Knight, R.T. High gamma power is phase-locked to theta oscillations in human neocortex. Science **313**(5793), 1626–1628 (2006)
12. Cohen, D. Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. Science **161**, 784–786 (1968)
13. Cohen, D. Magnetoencephalography: Detection of the brain's electrical activity with a superconducting magnetometer. Science **175**, 664–666 (1972)
14. Cotter, S.F., Rao, B.D., Engan, K., Kreutz-Delgado, K. Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Trans Signal Process **53**(7), 2477–2488 (2005)
15. Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E. Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron **26**(1), 55–67 (2000)
16. Darvas, F., Ermer, J.J., Mosher, J.C., Leahy, R.M. Generic head models for atlas-based EEG source analysis. Hum Brain Mapp **27**(2), 129–143 (2006)
17. Dogdas, B., Shattuck, D.W., Leahy, R.M. Segmentation of skull and scalp in 3-D human MRI using mathematical morphology. Hum Brain Mapp **26**(4), 273–285 (2005)

18. Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J. Classical and Bayesian inference in neuroimaging: Theory. NeuroImage **16**(2), 465–483 (2002)

19. Fuchs, M., Kastner, J., Wagner, M., Hawes, S., Ebersole, J.S. A standardized boundary element method volume conductor model. Clin Neurophysiol **113**(5), 702–712 (2002)

20. Fuchs, M., Wagner, M., Kohler, T., Wischmann, H.A. Linear and nonlinear current density reconstructions. J Clin Neurophysiol 16(3), 267–295 (1999)

21. Gencer, N.G., Williamson, S.J. Differential characterization of neural sources with the bimodal truncated SVD pseudo-inverse for EEG and MEG measurements. IEEE Trans Biomed Eng **45**(7), 827–838 (1998)

22. George, J.S., Aine, C.J., Mosher, J.C., Schmidt, D.M., Ranken, D.M., Schlitt, H.A., Wood, C.C., Lewine, J.D., Sanders, J.A., Belliveau, J.W. Mapping function in the human brain with magnetoencephalography, anatomical magnetic resonance imaging, and functional magnetic resonance imaging. J Clin Neurophysiol **12**(5), 406–431 (1995)

23. Golub, G.H. , van Loan, C.F. Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore, MD (1996)

24. Goncalves, S.I., deMunck, J.C., Verbunt, J.P.A., Bijma, F., Heethaar, R.M., da Silva, F.L. In vivo measurement of the brain and skull resistivities using an EIT-based method and realistic models for the head. IEEE Trans Biomed Eng **50**(6), 754–767 (2003)

25. Gorodnitsky, I., Rao, B.D. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. IEEE Trans Signal Process **45**(3), 600–616 (1997)

26. Gorodnitsky, I.F., George, J.S., Rao, B.D. Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm. Electroencephalogr Clin Neurophysiol **95**(4), 231–251 (1995)

27. Grave de Peralta Menendez, R., Gonzalez Andino, S.L. Backus and Gilbert method for vector fields. Hum Brain Mapp **7**(3), 161–165 (1999)

28. Gross, J., Ioannides, A.A. Linear transformations of data space in MEG. Phys Med Biol **44**(8), 2081–2097 (1999)

29. Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., Salmelin, R. Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proc Natl Acad Sci USA **98**(2), 694–699 (2001)

30. Halchenko, Y.O., Hanson, S.J., Pearlmutter, B.A. Multimodal integration: fMRI, MRI, EEG, MEG. In: Landini, L., Positano, V., Santarelli, M.F. (eds.) *Advanced Image Processing in Magnetic Resonance Imaging, Signal Processing and Communications*, pp. 223–265. Dekker, New York (2005)

31. Hamalainen, M., Hari, R., Ilmoniemi, R., Knuutila, J., Lounasmaa, O. Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys **65**(2), 413–497 (1993)

32. Hamalainen, M., Sarvas, J. Feasibility of the homogenous head model in the interpretation of the magnetic fields. Phys Med Biol **32**, 91–97 (1987)

33. von Helmholtz, H. Ueber einige Gesetze der Vertheilung elektrischer Strome in korperlichen Leitern, mit Anwendung auf die thierisch-elektrischen Versuche. Ann Phys Chem **89**, 211–233, 353–377 (1853)

34. Hillebrand, A., Barnes, G.R. The use of anatomical constraints with MEG beamformers. NeuroImage **20**(4), 2302–2313 (2003)

35. Huang, M., Aine, C.J., Supek, S., Best, E., Ranken, D., Flynn, E.R. Multi-start downhill simplex method for spatio-temporal source localization in magnetoencephalography. Electroencephalogr Clin Neurophysiol **108**(1), 32–44 (1998)

36. Huang, M.X., Mosher, J.C., Leahy, R.M. A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG. Phys Med Biol **44**(2), 423–440 (1999)

37. Hubbard, J.I., Llinás, R.R., Quastel, D.M.J. Electrophysiological Analysis of Synaptic Transmission. Edward Arnold, London (1969)

38. Ioannides, A.A., Bolton, J.P., Clarke, C.J.S. Continuous probabilistic solutions to the biomagnetic inverse problem. Inverse Probl **6**, 523–542 (1990)

39. Jerbi, K., Mosher, J.C., Baillet, S., Leahy, R.M. On MEG forward modelling using multipolar expansions. Phys Med Biol, **47**(4), 523–555 (Feb 2002)

40. Lachaux, J.P., Rudrauf, D., Kahane, P. Intracranial EEG and human brain mapping. J Physiol (Paris) **97**, 613–628 (2003)
41. Liu, A.K., Dale, A.M., Belliveau, J.W. Monte Carlo simulation studies of EEG and MEG localization accuracy. Hum Brain Mapp **16**(1), 47–62 (2002)
42. Liu, L., Ioannides, A.A., Streit, M. Single trial analysis of neurophysiological correlates of the recognition of complex objects and facial expressions of emotion. Brain Topogr **11**(4), 291–303 (1999)
43. Luck, S.J. An Introduction to the Event-Related Potential Technique. MIT Press, Cambridge, MA (2005)
44. Mackay, D.J.C. Bayesian interpolation. Neural Comput **4**(3), 415–447 (1992)
45. MacKay, D.J.C. Comparison of approximate methods for handling hyperparameters. Neural Comput **11**(5), 1035–1068 (1999)
46. Makeig, S., Jung, T.P., Bell, A.J., Ghahremani, D., Sejnowski, T.J. Blind separation of auditory event-related brain responses into independent components. Proc Natl Acad Sci USA **94**(20), 10979–10984 (1997)
47. Makeig, S., Ramírez, R.R. Neuroelectromagnetic source imaging (NSI) toolbox and EEGLAB module. Proceedings of the 37th Annual Meeting of the Society for Neuroscience, San Diego, CA (2007)
48. Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J. Dynamic brain sources of visual evoked responses. Science **295**(5555), 690–694 (2002)
49. Matsuura, K., Okabe, Y. Selective minimum-norm solution of the biomagnetic inverse problem. IEEE Trans Biomed Eng **42**(6), 608–615 (1995)
50. Mattout, J., Phillips, C., Penny, W.D., Rugg, M.D., Friston, K.J. MEG source localization under multiple constraints: An extended Bayesian framework. NeuroImage **30**(3), 753–767 (2006)
51. Mitra, P.P., Maniar, H. Concentration maximization and local basis expansions (LBEX) for linear inverse problems. IEEE Trans Biomed Eng **53**(9), 1775–1782 (2006)
52. Mosher, J.C., Leahy, R.M. Recursive MUSIC: A framework for EEG and MEG source localization. IEEE Trans Biomed Eng **45**(11), 1342–1354 (1998)
53. Mosher, J.C., Lewis, P.S., Leahy, R.M. Multiple dipole modeling and localization from spatio-temporal MEG data. IEEE Trans Biomed Eng **39**(6), 541–557 (1992)
54. Murakami, S., Okada, Y. Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. J Physiol **575**(Pt 3), 925–936 (2006)
55. Neal, R.M. Bayesian Learning for Neural Networks. Springer, New York; Secaucus, NJ (1996)
56. Nguyen, N., Milanfar, P., Golub, G. Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement. IEEE Trans Image Process **10**(9), 1299–1308 (2001)
57. Nicholson, C., Llinas, R. Field potentials in the alligator cerebellum and theory of their relationship to Purkinje cell dendritic spikes. J Neurophysiol **34**(4), 509–531 (1971)
58. Niedermeyer, E., Lopes da Silva, F. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Williams & Wilkins, Philadelphia, PA (2005)
59. Nummenmaa, A., Auranen, T., Hamalainen, M.S., Jaaskelainen, I.P., Lampinen, J., Sams, M., Vehtari, A. Hierarchical Bayesian estimates of distributed MEG sources: Theoretical aspects and comparison of variational and MCMC methods. NeuroImage **35**(2), 669–685 (2007)
60. Nummenmaa, A., Auranen, T., Hamalainen, M.S., Jaaskelainen, I.P., Sams, M., Vehtari, A., Lampinen, J. Automatic relevance determination based hierarchical Bayesian MEG inversion in practice. NeuroImage **37**(3), 876–889 (2007)
61. Nunez, P.L., Srinivasan, R. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, New York (2006)
62. Okada, Y. Empirical bases for constraints in current-imaging algorithms. Brain Topogr **5**(4), 373–377 (1993)

63. Okada, Y.C., Wu, J., Kyuhou, S. Genesis of MEG signals in a mammalian CNS structure. Electroencephalogr Clin Neurophysiol **103**(4), 474–485 (1997)
64. Parra, L.C., Spence, C.D., Gerson, A.D., Sajda, P. Recipes for the linear analysis of EEG. NeuroImage **28**(2), 326–341 (2005)
65. Pascual-Marqui, R.D. Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. Methods Find Exp Clin Pharmacol **24** Suppl D, 5–12 (2002)
66. Pascual-Marqui, R.D., Lehmann, D., Koenig, T., Kochi, K., Merlo, M.C., Hell, D., Koukkou, M. Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. Psychiatry Res **90**(3), 169–179 (1999)
67. Penfield, W., Jasper, H.H. Epilepsy and the Functional Anatomy of the Human Brain. Little, Brown, Boston (1954)
68. Phillips, C., Mattout, J., Rugg, M.D., Maquet, P., Friston, K.J. An empirical Bayesian solution to the source reconstruction problem in EEG. NeuroImage 24(4), 997–1011 (2005)
69. Ramírez, R.R. Neuromagnetic Source Imaging of Spontaneous and Evoked Human Brain Dynamics. PhD thesis, New York University School of Medicine, New York (2005)
70. Ramírez, R.R., Makeig, S. Neuroelectromagnetic source imaging using multiscale geodesic neural bases and sparse Bayesian learning. Proceedings of the 12th Annual Meeting of the Organization for Human Brain Mapping, Florence, Italy (2006)
71. Ramírez, R.R., Makeig, S. Neuroelectromagnetic source imaging of spatiotemporal brain dynamical patterns using frequency-domain independent vector analysis (IVA) and geodesic sparse Bayesian learning (gSBL). Proceedings of the 13th Annual Meeting of the Organization for Human Brain Mapping, Chicago, IL (2007)
72. Ramírez, R.R., Makeig, S. Neuroelectromagnetic source imaging using multiscale geodesic basis functions with sparse Bayesian learning or MAP estimation. Neural Comput (In preparation) (2010)
73. Ramírez, R.R., Wipf, D., Rao, B., Makeig, S. Sparse Bayesian learning for estimating the spatial orientations and extents of distributed sources. Biomag 2006 – Proceedings of the 15th International Conference on Biomagnetism, Vancouver, BC, Canada (2006)
74. Rao, B.D., Engan, K., Cotter, S.F., Palmer, J., Kreutz-Delgado, K. Subset selection in noise based on diversity measure minimization. IEEE Trans Signal Process **51**(3), 760–770 (2002)
75. Rao, B.D., Kreutz-Delgado, K. An affine scaling methodology for best basis selection. IEEE Trans Signal Process **1**, 187–202 (1999)
76. Ribary, U., Ioannides, A.A., Singh, K.D., Hasson, R., Bolton, J.P., Lado, F., Mogilner, A., Llinas, R. Magnetic field tomography of coherent thalamocortical 40-Hz oscillations in humans. Proc Natl Acad Sci USA **88**(24), 11037–11041 (1991)
77. Sarnthein, J., Morel, A., von Stein, A., Jeanmonod, D. Thalamic theta field potentials and EEG: High thalamocortical coherence in patients with neurogenic pain, epilepsy and movement disorders. Thalamus Related Syst **2**(3), 231–238 (2003)
78. Sarvas, J. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phys Med Biol **32**(1), 11–22 (1987)
79. Sato, M., Yoshioka, T., Kajihara, S., Toyama, K., Naokazu, G., Doya, K., Kawatoa, M. Hierarchical Bayesian estimation for MEG inverse problem. NeuroImage **23**, 806–826 (2004)
80. Scherg, M., Berg, P. Use of prior knowledge in brain electromagnetic source analysis. Brain Topogr **4**(2), 143–150 (1991)
81. Schimpf, P.H., Liu, H., Ramon, C., Haueisen, J. Efficient electromagnetic source imaging with adaptive standardized LORETA/FOCUSS. IEEE Trans Biomed Eng **52**(5), 901–908 (2005)
82. Schmidt, D.M., George, J.S., Wood, C.C. Bayesian inference applied to the electromagnetic inverse problem. Hum Brain Mapp **7**(3), 195–212 (1999)
83. Sekihara, K., Nagarajan, S., Poeppel, D., Miyashita, Y. Time-frequency MEG-music algorithm. IEEE Trans Med Imaging **18**(1), 92–97 (1999)
84. Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J. Stimulus specificity of phaselocked and non-phase-locked 40 Hz visual responses in human. J Neurosci **16**(13), 4240–4249 (1996)

85. Tang, A.C., Pearlmutter, B.A., Malaszenko, N.A., Phung, D.B., Reeb, B.C. Independent components of magnetoencephalography: Localization. Neural Comput **14**(8), 1827–1858 (2002)
86. Taulu, S., Kajola, M., Simola, J. Suppression of interference and artifacts by the signal space separation method. Brain Topogr **16**(4), 269–275 (2004)
87. Taylor, J.G., Ioannides, A.A., Muller-Gartner, H.W. Mathematical analysis of lead field expansions. IEEE Trans Med Imaging **18**(2), 151–163 (1999)
88. Tesche, C.D. Non-invasive detection of ongoing neuronal population activity in normal human hippocampus. Brain Res **749**(1), 53–60 (1997)
89. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res **1**, 211–244 (2001)
90. Tuch, D.S., Wedeen, V.J., Dale, A.M., George, J.S., Belliveau, J.W. Conductivity tensor mapping of the human brain using diffusion tensor MRI. Proc Natl Acad Sci USA **98**(20), 11697–11701 (2001)
91. Ulbert, I., Halgren, E., Heit, G., Karmos, G. Multiple microelectrode-recording system for human intracortical applications. J Neurosci Methods **106**(1), 69–79 (2001)
92. Uusitalo, M.A., Ilmoniemi, R.J. Signal-space projection method for separating MEG or EEG into components. Med Biol Eng Comput **35**(2), 135–140 (1997)
93. Uutela, K., Hamalainen, M., Salmelin, R. Global optimization in the localization of neuromagnetic sources. IEEE Trans Biomed Eng **45**(6), 716–723 (1998)
94. Uutela, K., Hamalainen, M., Somersalo, E. Visualization of magnetoencephalographic data using minimum current estimates. NeuroImage **10**(2), 173–180 (1999)
95. Van Veen, B.D., van Drongelen, W., Yuchtman, M., Suzuki, A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng **44**(9), 867–880 (1997)
96. Volkmann, J., Joliot, M., Mogilner, A., Ioannides, A.A., Lado, F., Fazzini, E., Ribary, U., Llinas, R. Central motor loop oscillations in parkinsonian resting tremor revealed by magnetoencephalography. Neurology **46**(5), 1359–1370 (1996)
97. Vrba, J., Robinson, S.E. Signal processing in magnetoencephalography. Methods **25**(2), 249–271 (2001)
98. Wang, J.Z., Williamson, S.J., Kaufman, L. Magnetic source images determined by a lead-field analysis: The unique minimum-norm least-squares estimation. IEEE Trans Biomed Eng **39**(7), 665–675 (1992)
99. Wipf, D.P., Ramírez, R.R., Palmer, J.A., Makeig, S., Rao, B.D. Analysis of empirical Bayesian methods for neuroelectromagnetic source localization. In: Schlkopf, B., Platt, J., Hoffman, T. (eds.), Advances in Neural Information Processing Systems, vol. 19. MIT Press, Cambridge, MA (2007)
100. Wipf, D.P., Rao, B.D. Sparse Bayesian learning for basis selection. IEEE Trans Signal Process **52**, 2153–2164 (2004)
101. Wipf, D.P., Rao, B.D. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. IEEE Trans Signal Process **55**(7), 3704–3716 (2007)
102. Wolters, C.H., Anwander, A., Tricoche, X., Weinstein, D., Koch, M.A., MacLeod, R.S. Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: A simulation and visualization study using high-resolution finite element modeling. NeuroImage **30**(3), 813–826 (2006)
103. Zimmerman, J.E., Frederick, N.V. Miniature ultrasensitive superconducting magnetic gradiometer and its use in cardiography and other applications. Appl Phys Lett **19**(1), 16–19 (1971)
104. Zumer, J.M., Attias, H.T., Sekihara, K., Nagarajan, S.S. A probabilistic algorithm integrating source localization and noise suppression for MEG and EEG data. NeuroImage **37**, 102–115 (2007)

Chapter 9
Optimization in Brain? – Modeling Human Behavior and Brain Activation Patterns with Queuing Network and Reinforcement Learning Algorithms

Changxu Wu, Marc Berman, and Yili Liu

Abstract Here we present a novel approach to model brain and behavioral phenomena of multitask performance, which integrates queuing networks with reinforcement learning algorithms. Using the queuing network as the static platform of brain structure and reinforcement learning as the dynamic algorithm to quantify the learning process, this model successfully accounts for several behavioral phenomena related to the learning process of transcription typing and the psychological refractory period (PRP). This model also proposes brain changes that may accompany the typing and PRP practice effects that could be tested empirically with neuroimaging. All of the modeled phenomena emerged as outcomes of the natural operations of the human information processing queuing network.

9.1 Introduction

Elucidating the psychological and physiological processes that mediate cognitive and behavioral performance has been an important topic for a long period of time. This topic for many years was studied exclusively with behavioral techniques, and models of behavioral performance had to be inferred exclusively from behavioral data [13, 45]. Current researchers are now endowed with two addi-

Changxu Wu
Department of Industrial and Systems Engineering, State University of New York (SUNY), Buffalo, NY, USA, e-mail: `Changxu@buffalo.edu`

Marc Berman
Department of Psychology Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA, e-mail: `bermanm@umich.edu`

Yili Liu
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA, e-mail: `yililiu@umich.edu`

tional techniques to understand and to explain human behavioral performance: neuroimaging and computational modeling. With neuroimaging techniques, such as functional magnetic resonance imaging (fMRI [8]), positron emission tomography (PET [9]), and event-related potentials (ERP [29]), researchers can uncover the neural substrates that mediate behavioral performance. These neuroimaging techniques not only allow researchers to localize where cognitive processes reside in the brain, but also allow researchers to uncover commonalities and dissimilarities between cognitive tasks, discover individual differences, and test psychological theories and models in ways that behavioral techniques alone could not uncover [3].

Computational modeling has also been a powerful technique to simulate and compose models for how behavior is mediated. Computational models can be classified into a number of categories, including, e.g., connectionist [19, 30, 39], symbolic [24, 31], and hybrid [4, 27, 59, 51, 48, 53, 55, 54, 56, 52, 60, 58, 57, 61, 62, 63]. With these computational models, researchers are able to validate, test, and update psychological theories in ways that behavioral testing alone could not do easily.

Here we utilize computational modeling to account for changes in performance both behaviorally and neurally due to practice and learning in the context of transcription typing and the psychological refractory period (PRP; the slowing of a secondary task when it is initiated during the response of a primary task). This novel model unifies many disparate findings together into a single model without needing to make many changes to model parameters.

We chose to model the practice and learning effects in transcription typing and PRP due to the following reasons. First, transcription typing involves intricate and complex interactions of perceptual, cognitive, and motoric processes, and modeling its learning processes can help us understand the underlining quantitative mechanisms in complex motor skill acquisition. Second, there exist brain imaging data on typing and typing related behavior [17, 23] that could be modeled. In addition, human behavioral performance data, such as typing speed and typing variability, have been obtained via several experimental studies (please see the review of Salthouse [44]).

We modeled the learning effect in PRP for similar reasons. First, PRP is the simplest and one of the most basic paradigms to study multitask performance and has been used extensively as a paradigm to study multitask performance. The PRP effect has been applied in many real-world settings such as driving [25] and has been used as a measure of dual-task competency [5, 11]. Therefore, modeling the learning effects in PRP may allow us to account for the basic mechanisms in the acquisition of multitasking skills. Second, an experimental study has been conducted to study the learning effect in PRP [49], which provides important human performance data for modeling. For these reasons we found transcription typing and PRP tasks good candidates to model skill learning behavior.

9.2 Modeling Behavioral and Brain Imaging Phenomena in Transcription Typing with Queuing Networks and Reinforcement Learning Algorithms

9.2.1 Behavioral Phenomena

Salthouse [42] reviewed the major behavioral empirical results of transcription typing and summarized 29 phenomena in this area. John [22] summarized two additional behavioral phenomena found by Gentner [16] and [43]. These 31 behavioral phenomena include 12 basic phenomena, 5 error phenomena, 6 phenomena in typing units, and 8 skill learning phenomena in transcription typing. We have developed a queuing network model that successfully modeled 32 behavioral phenomena in transcription typing including 3 newly discovered eye movement phenomena and 29 of these 31 behavioral phenomena, with the exceptions being 2 phenomena related to reading and comprehension, whose modeling requires significant extensions of our model to include production systems and is a current topic of our ongoing research [48]. In this chapter we focus on modeling the learning aspects of the behavioral phenomena and brain imaging phenomena.

The first typing phenomenon that we modeled was changes in interkey response time of transcription typing, which decreases accordingly to the power law of practice [16]. For example, typing speeds of an unskilled typist (about 30 words per minute [21]) can be improved to that of a skilled typist (about 68 words per minute [42]).

The second phenomenon involved the variability of interkey intervals which decreases with the increased skill of the typist. In addition, the interquartile range of interkey intervals correlates significantly with typist's net interkey intervals ($p < 0.05$ [41]). The third behavioral phenomenon that we modeled that we will describe in this chapter was modeling the rate of repetitive tapping, which is greater among more skilled typists and the correlation between repetitive tapping speed and net typing speed is reliable ($p < 0.05$, [41]).

9.2.2 Brain Imaging Phenomena

Recently, brain imaging studies (fMRI and PET) have discovered two phenomena related to transcription typing. First, it has been found that at the beginning stages of learning a visuomotor control task, including transcription typing, the dorsal lateral prefrontal cortex (DLPFC), the basal ganglia, and the pre-SMA are highly activated [31, 40]. After practice, activation of the DLPFC disappears and strong activation is observed in the supplementary motor area (SMA), the basal ganglia, and the primary motor cortex (M1) in addition to slight activation in the somatosensory cortex (S1) [17].

Second, in the well-learned stages of typing (skilled typist in [17]), when stimuli to be typed are repetitive letters (e.g., AAA...), M1 is strongly activated, however, when stimuli to be typed are multiletter sentences (e.g., JACK AND...), M1 is strongly activated, but there is more robust activation in the SMA, the basal ganglia, and S1.

9.2.3 A Queuing Network Model with Reinforcement Learning Algorithms

9.2.3.1 The Static Portion of the Queuing Network Model

Queuing network is a mathematical discipline that is used to simulate and model a wide array of phenomena and systems including manufacturing and computer network performance. A queuing network is a network of servers that provide services to customers that wait in queues before they are serviced. Queuing networks tend to be quite flexible and can allow two or more servers to act in serial, in parallel, or in any network configuration [26, 27]. Computational models based on queuing networks have successfully integrated a large number of mathematical models of response time [26] and multitask performance [27]. A queuing network modeling architecture is called the queuing network. Model human processor (QN-MHP) has been developed and used to generate behavior in real time [28], including simple and choice reaction time [14] and driver performance [44]. The model in this chapter extends QN-MHP by integrating reinforcement learning algorithms and strengthening its long-term memory and nine motor subnetwork servers. In addition, the queuing network approach has also been used to quantify changes in brain activation for different participant populations [4].

The brain, which is an enormously complex network of interconnected systems and subsystems, acts in concert with one another to produce behavior. This idea is supported by evidence from pathway tracing studies in nonhuman primates, which revealed widely distributed networks of interconnected cortical areas, providing an anatomical substrate for large-scale parallel processing in the cerebral cortex [6]. It seems, then, that brain areas do not act in isolation from another and instead may form complex neural networks that are the basis of behavior and thought.

In addition to the widely distributed nature of the brain, each brain area may also have some level of functional specialization [9] and thus each major brain area may have certain information processing capacities and certain processing time parameters (see Table 9.1). Here we assume that the interconnections between major brain areas form a queuing network with each major brain area composing a queuing network server and that information processed at each server is a queuing network entity. In addition, neuron pathways that connect major brain areas serve as routes between our queuing network servers (see Fig. 9.1 for transcription typing routing and Fig. 3.1 a for PRP routing. Note that both networks have the same servers and overall network configurations). Therefore, it is assumed that the major brain areas form a queuing network with brain areas as the servers, information processed as

entities, and neuron pathways as routes (see Fig. 9.1). Within this general information processing structure, the major brain areas activated in the transcription typing task 10 were identified by the following fMRI and PET studies ([23, 40, 17], see Fig. 9.1).
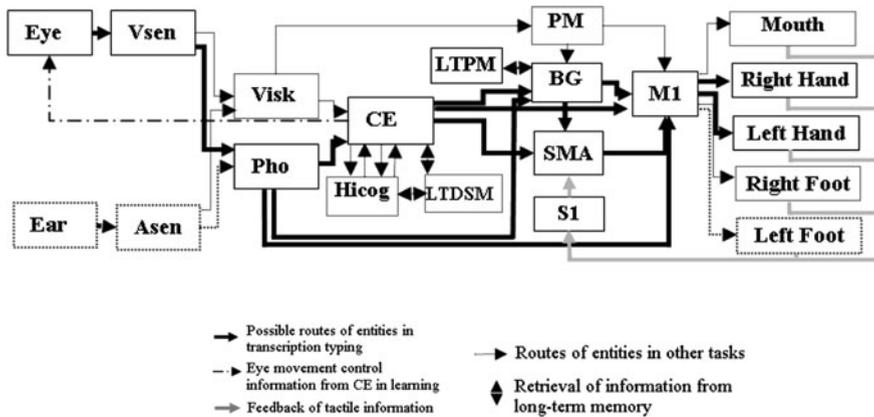

Fig. 9.1: The general structure of the queuing network model (QN-MHP) with routes and servers involved in transcription typing tasks highlighted (server names, brain structures, and processing logic and time are shown in Table 9.1).

Processing logic and time is based on the literature [10, 27, 38, 14, 37]. If we consider the network for transcription typing, as shown in Fig. 9.1, upon completing service at the Pho server, entities have numerous possible routes to follow to traverse the network: (1) At the Pho server, the entities can choose one of the three routes to depart the Pho server to the CE, BG, or M1 servers. (2) At the CE server, entities can choose to move to the BG, SMA, or M1. (3) At the BG server, entities can move to the SMA or M1 servers. Therefore, there are a total of $3 \times 3 \times 2 = 18$ possible routes for the entities to be processed in the network in transcription typing. An important question is, therefore, how the entities choose among these routes that activate (utilize) different brain areas (servers) in different learning stages or when processing different stimuli at well-learned stages? This question can be answered by the dynamic part of the model.

9.2.3.2 The Dynamic Portion of the Queuing Network Model: Self-Organization of the Queuing Network with Reinforcement Learning Algorithms

Ungerleider et al. [47] found evidence for the reorganization of brain areas with practice, which indicates that individual brain areas may change their information processing speeds in learning. Moreover, some brain areas may have error detection

Table 9.1: Server name, major function, and brain structure

Server	Brain structure	Major function (Processing logic)
Eye	Eye, LGN, SC, Visual pathway	Visual sampling and signal transmission
VSen	Distributed parallel area, superior frontal sulcus, dorsal and ventral system	Visual sensory memory and perception
Pho	Left posterior parietal cortex, inferior parietal lobe	Phonological loop to store auditoria and textual information
CE	Dorsal lateral prefrontal cortex and ACC	Mental process and response inhibition and selection
BG	Basal ganglia	Motor program retrieval
LTPM	Striatal and cerebellar systems	Long-term procedural knowledge storage
SMA	Supplementary motor area and pre-SMA	Motor program assembly, error detection, and bimanual coordination
M1	Primary motor cortex	Addressing spinal motorneourons
S1	Somatosensory cortex	Sending the sensory information to other areas
Hand	–	Execution of motor movement

functions but others may not (see Table 9.1). Because the routes of the queuing network are composed of different brain areas (servers), different routes chosen by the entities may lead to different information processing speeds or errors. If the entities try to maximize response time performance, they may choose an optimal route that maximizes speed, but may not minimize error. Some routes, however, may maximize both performance measures. Therefore, in different situations, different routes may be chosen by the entities which activate different brain areas (servers). This ability to have different routes becoming active forms the dynamic, self-organization aspect of the queuing network. Consequently, there are two levels of learning within the queuing network: (1) learning processes at the individual server level and (2) self-organization or routes of the queuing network that change depending on the stages of learning or the type of stimuli presented.

Learning Processes of the Individual Servers

In the motor learning process, the basal ganglia, striatal, and cerebellar systems (BG and LTPM servers) play a major role in procedural knowledge acquisition [2]. Therefore, the current model focuses on the BG and the LTPM servers in quantifying the learning processes of individual servers. It is assumed that the time for the BG server to retrieve a motor program from the LTPM decreases exponentially as a function of the number of practice trials (see Equation 9.1). Because the exponential function fits learning processes of memory search, motor learning, visual search,

mathematic operation tasks better than the power law [19] and has been applied in modeling long-term memory retrieval [1] we used it to model our individual server learning processes:

$$1/\mu_{BG} = A_{BG} + B_{BG}\exp-\alpha_{BG}N_{BG}, \tag{9.1}$$

$1/\mu_{BG}$: motor program retrieving time; A_{BG}: the minimal of processing time of BG server after practice (314 ms, [35]); B_{BG}: the change of expected value of processing time from the beginning to the end of practice ($2 \times 314 = 628$ ms, assumed). α_{BG}: the learning rate of server BG (0.00142, [18]); N_{BG}: number of digraphs (letter pairs excluding the space key) processed by server BG, which is implemented as a matrix of diagraph frequency recorded in LTPM server.

Self-Organization of the Queuing Network

If the entities traversing the network try to maximize their information processing speed and minimize error, it is appropriate to apply reinforcement learning algorithms to quantify this dynamic process. Reinforcement learning is a computational approach able to quantify how an agent tries to maximize the total amount of reward it receives in interacting with a complex, uncertain environment [46]. Reinforcement learning has also been applied in modeling motor learning in neuroscience [33] and, therefore, may be appropriately applied to model brain network organization. To integrate the reinforcement learning algorithms with the queuing network approach, it is necessary to define the state, transitions, and reward values of reinforcement learning with the concepts of queuing networks. Below are the definitions:

1. *State:* the status that an entity is in server i.
2. *Transition:* An entity routed from server i to j.
3. *Time-saving reward (r'_t):* $r'_t = (1/w_q) + \mu_{j,t}$ (2)
   w_q: time the entity spent waiting in the queuing of the server; $\mu_{j,t}$: processing speed of the entity at that server.
4. *Error-saving reward (r''_t):* $r''_t = 1/(\text{Nerror}_{j,t} + 1)$ (3)

$\text{Nerror}_{j,t}$: number of action errors of the previous entities made in the next server j at tth transition. Q online learning algorithms in reinforcement learning are used to quantify the processes that are used by entities to choose different routes based on rewards of different routes.

1. *Q* **online learning algorithm of time-saving reward**

$$Q_T^{t+1}Q_T^t(i,j) + \varepsilon\{r'_t + \gamma\max_k[Q_T^t(j,k)] - Q_T^t(i,j)\}, \tag{9.2}$$

ε: learning rate of Q online learning ($0 < \varepsilon < 1$, $\varepsilon = 0.99$);
γ: discount parameter of routing to next server ($0 < \gamma < 1$, $\gamma = 0.3$);
$Q_T^t + 1(i,j)$: online Q value if entity routes from server i to server j in $t+1$th transition based on time-saving reward;

$\max_k[Q_T^t(j,k)]$: maximum Q value routing from server j to the next k server(s) ($k \geq 1$).

Equation (9.2) updates a Q value of a backup choice of routes ($Q_T^t(i,j)$) based on the Q value which maximizes over all those routes possible in the next state ($\max_k[QtT(j,k)]$). In each transition, entities will choose the next server according to the updated $Q_T^t(i,j)$.

2. **Q online learning algorithm of error-saving reward**

$$Q_E^{t+1}Q_E^t(i,j) + \varepsilon\{r_t'' + \gamma\max_k[Q_E^t(j,k)] - Q_E^t(i,j)\}. \tag{9.3}$$

3. **Trade-off of the two Q values**

The choice of routes is determined by the trade-off between the two Q values. Currently, it is assumed that $Q_E^{t+1}(i,j)$ of error-saving reward has the higher priority than the $Q_T^{t+1}(i,j)$ of time-saving reward: if $Q_E^{t+1}(i,j) > Q_E^{t+1}(i,k)$, the entity will choose the next server j whatever the value of $Q_T^{t+1}(i,j)$; if $Q_E^{t+1}(i,j) = Q_E^{t+1}(i,k)$, entity will choose the next server with greater Q_T^{t+1}; if $Q_E^{t+1}(i,j) = Q_E^{t+1}(i,k)$ and $Q_T^{t+1}(i,j) = Q_T^{t+1}(i,k)$, entity will choose next server randomly. With these equations, we were able to successfully integrate queuing networks with reinforcement learning algorithms.

9.2.4 Model Predictions of three Skill Learning Phenomena and two Brain Imaging Phenomena

The three skill learning phenomena and the two brain imaging phenomena of transcription typing described earlier in this chapter can be predicted by the queuing network model with reinforcement learning.

9.2.4.1 Predictions of the three Skill Learning Phenomena

We assume that the processing times of the CE, BG, and SMA servers follow the exponential distribution (see Table 9.1 and Fig. 9.1) and are independent from one another. Therefore, if $Y_1 \cdots Y_k$ are k independent exponential random variables representing the processing times of the servers in our network, their sum X follows an Erlang distribution. Based on features of Erlang distributions, we have

$$X = \sum_{i=1}^{k} Y_i, \tag{9.4}$$

$$E[X] = E\left[\sum_{i=1}^{k} Y_i\right] = \sum_{i=1}^{k} E[Y_i] = k\frac{1}{\lambda}, \tag{9.5}$$

$$\text{Var}[X] = \text{Var}\left[\sum_{i=1}^{k} Y_i\right] = \sum_{i=1}^{k} \text{Var}[Y_i] = k\frac{1}{\lambda^2}. \tag{9.6}$$

These mathematical results can be used to predict the skill learning phenomena, together with the prediction described below that entities may learn to skip certain server(s). First, because $E[X] = k(1/\lambda)$, if $k' < k$, then it follows that $E[X'] < E[X]$. This may be one of the reasons that the skipping of server(s) can explain a reduction in interkey time in typing normal text (the first skill learning phenomenon in this chapter) and repetitive letters (the third skill learning phenomenon). Second, skipping some of the servers will decrease the variance of the Erlang distribution because if $k' < k$, then $\text{Var}[X'] < \text{Var}[X]$. This is one possible reason why skipping over server(s) can account for the reduction in the variability of interkey time in the learning process (the second skill learning phenomenon).

9.2.4.2 Predictions of the First Brain Imaging Phenomenon

At the Pho server during the initial stages of learning, entities can go through the CE server for eye movement control to locate the specific position of a target key on the keyboard ([12], see Fig. 9.1) and for response selection and inhibition. Entities can also traverse the route from Pho to BG, but it takes longer than going through the CE because the BG may not work effectively in retrieving the motor program from LTPM [2] and its Q value of time-saving reward is smaller than that of CE. Entities can also choose the route from Pho→ M1 directly. However, the occurrence of typing errors will decrease the Q value of error-saving reward from 18 Pho→M1. As the number of practice trials increases, the route Pho→BG is selected by the majority of the entities because the functions of CE are gradually replaced by the BG with less process time based on parallel cortico-basal ganglia mechanisms [33].

Second, at the CE server, entities can traverse one of the routes from CE to BG, SMA, or M1. If entities select the first route, the correct motor program will be retrieved without decreasing the $Q_E^{t+1}(i, j)$ value. If the second or the third route is chosen, its $Q_E^{t+1}(i, j)$ value will decrease because no correct motor program is retrieved.

The third prediction involves the BG server. Since stimuli keep changing in typing multidigit sentences, entities can go from the BG directly to M1 skipping SMA whose function is motor program assembling [36]. However, ensuring movement accuracy for error detection [17] will decrease $Q_E^{t+1}(i, j)$ in route BG...M1. In sum, at the beginning of the learning process, entities will go through Pho→CE→BG→SMA→M1. After learning, the majority of entity will travel Pho→BG→SMA→M1.

9.2.4.3 Predictions of the Second Brain Imaging Phenomenon

If stimuli change from repetitive letters to regular words in the same task, the entities will change routes from Pho→M1 to Pho→BG→SMA→M1 because the error-saving reward decreases in route Pho...M1 without the motor program functions of

BG and the sequencing functions of SMA. This is our second prediction of changes in neural processing with learning.

9.2.5 Simulation of the three Skill Learning Phenomena and the two Brain Imaging Phenomena

9.2.5.1 The First and the Second Skill Learning Phenomena

Simulation results showed that the simulated interkey interval in the learning process followed the power law of practice (R square $= 0.8$, $p < 0.001$). The simulated interkey interval also improved from 385 to 180 ms, which was consistent with existing experimental data about performance changes from the unskilled typist (interkey time 400 ms) with estimation error 3.75% (estimation error $= |YX|/X \times 100\%$, Y: simulation result; X: experiment result) to the skilled typist (177 ms interkey time) with estimation error 1.69% (see Fig. 9.1).

As shown in Fig. 9.2, the change of the quartile range (75% quartile–25% quartile) is significantly correlated with the change of the simulated speed ($p < 0.05$), which is consistent with the experimental results of Salthouse [41]. This was one of the phenomena not covered by TYPIST [22].
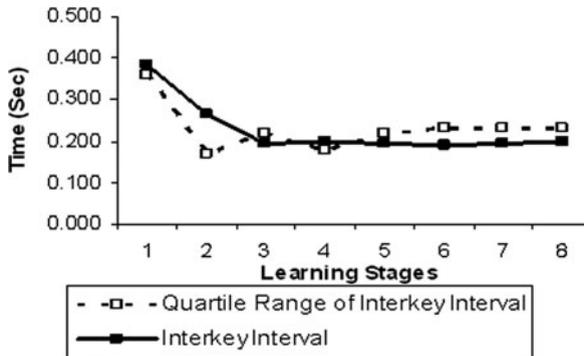

Fig. 9.2: Simulated variability of interkey interval and interkey interval in the learning process. Each stage represents 352,125 keystrokes.

9.2.5.2 The Third Skill Learning Phenomena

The simulated tapping rate (interkey interval in typing repetitive letters) and typing speed of text (interkey interval in typing multidigit sentence) during the learning process were found to be strongly correlated ($p < 0.05$), which is consistent with the experimental results of Salthouse [41] who found the significant correlation between the two variables ($p < 0.01$). Therefore, our model successfully modeled these behavioral phenomena with very high accuracy.

9.2.5.3 The First Brain Imaging Phenomena

As shown in Fig. 9.3, at the beginning of practice, the CE (including DLPFC) and the BG servers are highly utilized, while the SMA server (including pre-SMA) (3%) and M1 and two hand servers (15%) are less utilized. After $352,125 \times 8$ trials of practice, the CE server (DLPFC) decreased its utilization greatly to 0%. Percentage of utilization of SMA server is increased by 47%. M1 and two hand servers and S1 also increased their percentage of utilization during the learning process by 85% and 22%, respectively. These simulation results are consistent with the experimental results in PET and fMRI studies [23, 40, 17] who found similar patterns of increases and decreases in brain activity.
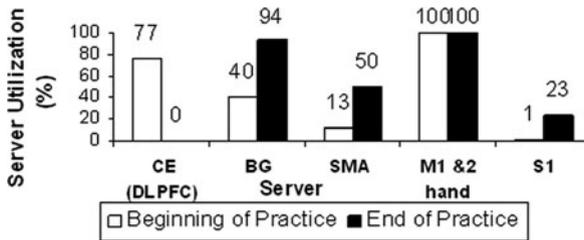

Fig. 9.3: Server utilization at the beginning and end of practice in learning to type multidigit sentence.

9.2.5.4 The Second Brain Imaging Phenomena

After the model finished its learning process, it was able to simulate the second brain imaging phenomenon of the skilled typist in typing different stimuli. The 1,600 letters to be typed by the model changed following this pattern: $1st - 800th$ letters: repetitive letters; $801st - 1,600$ letters: multidigit sentence.

Figure 9.4 shows the percentage of utilization of the major servers in the different stimulus conditions. When the model is typing repetitive letters, mainly M1 and two hand servers are utilized. When the stimuli changed from repetitive letters to multidigit sentences the utilization of SMA, BG, and S1 increased by 49, 90, and
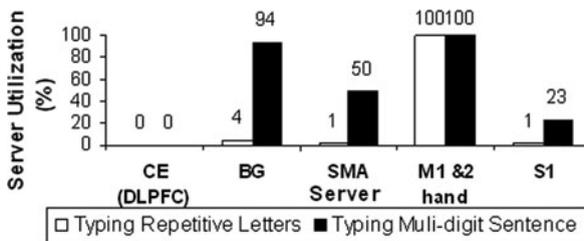

Fig. 9.4: Server utilization when stimuli presented changed in the well-learned transcription typing situation.

22%, respectively. The model demonstrated that fewer entities travel from Pho to M1 directly when the stimuli presented changes from repetitive letters to multidigit sentences. These results are consistent with the fMRI results of [17].

In practice, because our queuing network model was built with a general structure with common brain regions, it can be easily transformed to model other task situations, e.g., PRP [50]. Moreover, the current model can generate behavioral results by the interaction of the queuing network servers without drawing complex scheduling charts. These unique features offer great potential of the model for learning and can easily be used by researchers in cognitive modeling and human factors.

9.3 Modeling the Basic PRP and Practice Effect on PRP with Queuing Networks and Reinforcement Learning Algorithms

PRP (Psychological Refractory Period) is one of the most basic and simple forms of dual-task situations and has been studied extensively in the laboratory for half a century [31]. In the basic PRP paradigm, two stimuli are presented to subjects in rapid succession and each requires a quick response. Typically, responses to the first stimulus (Task 1) are unimpaired, but responses to the second stimulus (Task 2) are slowed by 300 ms or more . In the PRP paradigm of Selst et al. [44], task 1 required subjects to discriminate tones into high or low pitches with vocal responses (audio-vocal responses); in task 2 subjects watched visually presented characters and performed a choice reaction time task with manual responses (visual-motor responses). They found that practice dramatically reduced dual-task interference in PRP.

The basic PRP effect has been modeled by several major computational cognitive models based on production rules, notably EPIC [31] and ACT-R/PM [7]. Based on its major assumption that production rules can fire in parallel, EPIC successfully modeled the basic PRP effect by using complex lock and unlock strategies in central processes to solve the time conflicts between perceptual, cognitive, and motor processing [31]. However, neither EPIC nor ACT-R/PM modeled the practice effect on PRP.

Here we modeled PRP effects with the same model that modeled typing phenomena and integrated queuing network theory [26, 27] with reinforcement learning algorithms [46]. Model simulation results were compared with experimental results of both the basic PRP paradigm and the PRP practice effects [49]. All of the simulated human performance data were derived from the natural interactions among servers and entities in the queuing network without setting up lock and unlock strategies or drawing complex scheduling charts.

9.3.1 Modeling the Basic PRP and the Practice Effect on PRP with Queuing Networks

Figure 9.5 shows the queuing network model that was used to model PRP effects. The model architecture is identical to the model that was used to model typing
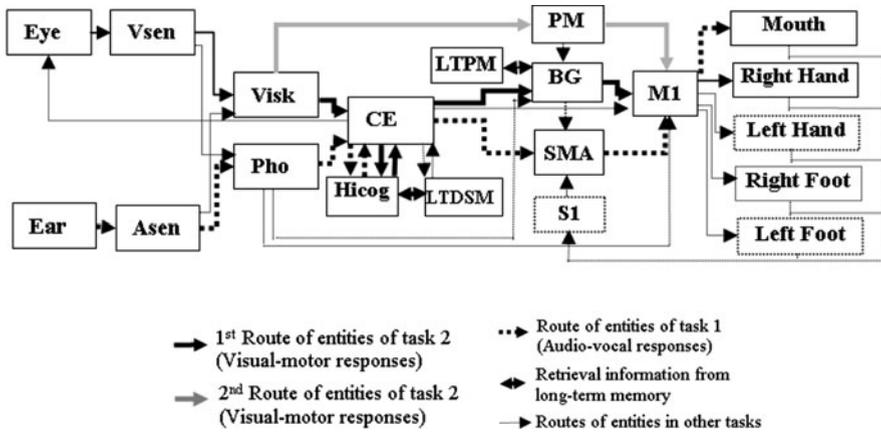
Fig. 9.5: The general structure of the queuing network model (QN-MHP) with servers and routes involved in the PRP task highlighted (server names, brain structures, major functions, and processing time are shown in Table 9.1).

phenomena. However, entities traverse different routes while performing PRP tasks than they traversed when performing typing tasks.

Because the PRP effect prior to or at the beginning of learning (the basic PRP) is a special case of the PRP effect during the learning process, the two phenomena of PRP (basic and learned) are modeled with the same mechanisms in our queuing network model. The experimental tasks and data of Van Selst et al. [49] were used to test the model.

Brain areas (servers) and their routes related to the two PRP tasks in Van Selste's study were identified within the general queuing network structure based on recent neuroscience findings [32, 15, 2], see Fig. 9.5). When exploring Fig. 9.5 entities of task 1 (audio-vocal responses) cannot bypass the Hicog server because the phonological judgment function is mainly mediated by the Hicog server, and thus there is only one possible route for the entities of task 1 (see the dotted thick line in Fig. 9.5) to traverse. However, the function of movement selection in task 2 (visual-motor responses) is located not only in the Hicog server but also in the PM server. Therefore, there are two possible routes for the entities of task 2 starting at Visk server (see the gray and black solid lines in Fig. 9.5).

However, how might the entities of task 2 choose one of the two alternative routes in the network? What is the behavioral impact of this choice on PRP and the practice effect on PRP? These questions can be answered by integrating queuing networks with reinforcement learning algorithms. Before exploring the mechanism with which entities of task 2 select from one of the two routes, it is necessary to understand the learning process of individual brain areas. It was discovered that each individual brain area reorganizes itself during the learning process and increases its processing speed [44]. For example, for the simplest network with two routes (see Fig. 9.6), if servers 2 and 3 change their processing speeds, different routes chosen by an entity (1→3→4 or 1→2→4) will lead to different performance. Without con-

sidering the effect of error, entities will choose the optimal route with the shortest processing time if they want to maximize the reward of performance.

Consequently, to model learning, it is first necessary to quantify the learning process in individual servers. Based on that, the condition under which an entity switches between the two routes shown in Fig. 9.6 can be established and proved by integrating queuing network with reinforcement learning. Finally, this quantitative condition of route switching can be applied to the more complex model of 18 servers with two routes (see Fig. 9.5) to generate the basic PRP and the reduction of PRP during the learning process.
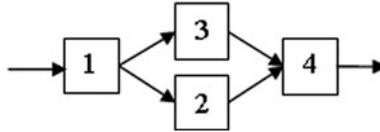

Fig. 9.6: The simplest queuing network with two routes.

9.3.1.1 Learning Process in Individual Servers

Based on the functions of the servers in Table 9.1, the two long-term memory servers (LTDSM and LTPM) play the major roles in learning phonological judgments (task 1) and choice reaction (task 2) [2]. Because the learning effects of long-term memory are represented as speed of retrieval of production rules and motor programs from the two long-term memory servers at the Hicog and the BG servers, it is important to quantify the processing time of the Hicog and the BG servers. In addition, because the premotor cortex (PM) server is activated in learning visuomotor associations [32], changes in the processing speed of the PM server is also to be considered in the learning process of the model.

Because the exponential function fits the learning processes in memory search, motor learning, visual search, and mathematic operation tasks better than the power law [18], it was again applied to model the learning process in the individual servers here

$$1/\mu_i = A_i + B_i \text{Exp}(-\alpha_i N_i), \tag{9.7}$$

μ_i: processing speed of the server i; $(1/\mu_i)$ is its processing time; A_i: the minimal of processing time of server i after intensive practice; B_i: the change of expected value of processing time of server i from the beginning to the end of practice; α_i: learning rate of server i; N_i: number of customers processed by server i.

For the BG server, $1/\mu_{\text{BG}}$: motor program retrieving time; ABBGB: the minimal of processing time of BG server after practice (314 ms, [35]); B_{BG}: the change of expected value of processing time from the beginning to the end of practice (2 × 314 = 628 ms, assumed); α_{BG}: the learning rate of server BG (0.00142, [18]); N_{BG}:

number of entities processed by server BG which is implemented as a matrix of frequency recorded in LTPM server.

For the Hicog and PM servers, to avoid building an ad hoc model and using the result of the experiment to be simulated directly, nine parameters in the Hicog and the PM servers were calculated based on previous studies (see Appendix 1).

9.3.1.2 Learning Process in the Simplest Queuing Network with two Routes

Based on the learning process of individual servers, the condition under which an entity switches between the two routes in the simplest form of queuing networks with two routes (each capacity equals 1) (from route $1 \ldots 2 \ldots 4$ to route $1 \ldots 3 \ldots 4$, see Fig. 9.6) was quantified and proved by the following mathematical deduction.

1. *Q* **online learning equation [46]**

$$Q^{t+1}(i,j)Q^t(i,j) + \varepsilon\{r_t + \gamma\max_k[Q^t(j,k) - Q^t(i,j)], \qquad (9.8)$$

where $Q^{t+1}(i,j)$ is the online Q value if entity routes from server i to server j in $t + 1$th transition; $\max_k[Q(j,k)]$ represents maximum Q value routing from server j to the next k server(s) ($k \leq 1$); $r_t = \mu_{j,t}$ is the reward and is the processing speed of the server j if entity enters it at tth transition; N_{jt} represents number of entities go to server j at tth transition; ε is the learning rate of Q online learning ($0 < \varepsilon < 1$); γ is the discount parameter of routing to next server ($0 < \gamma < 1$); and p is the probability of entity routes from server 1 to server 3 does not follow the Q online learning rule if $Q(1,3) > Q(1,2)$. For example, if $p = 0.1$, then 10% of entity will go from server 1 to server 2 even though $Q(1,3) > Q(1,2)$.
   State is the status that an entity is in server i; transition is defined as an entity routed from server i to j. Equation (9.8) updates a Q value of a backup choice of routes ($Q^{(t+1)}(i,j)$) based on the Q value which maximizes over all those routes possible in the next state ($\max_k[Q(j,k)]$). In each transition, entities will choose the next server according to the updated $Q^t(i,j)$. If $Q(1,3) > Q(1,2)$, more entity will go from server 1 to server 3 rather than go to server 2.
2. **Assumption**

   - ε is a constant which does not change in the current learning process ($0 < \varepsilon < 1$) .
   - Processing speed of server 4 (μ_4) is constant.

3. **Lemma 9.1.** At any transition state t ($t \neq 0$), if $1/\mu_{2,t} < 1/\mu_{3,t}$ then $Q^{t+1}(1,2) > Q^{t+1}(1,3)$
   Proof of Lemma 9.1 (see Appendix 2).
   Based on Lemma 9.1 and Equation (9.7), we got Lemma 9.2:
4. **Lemma 9.2.** At any transition state t ($t \neq 0$), if $A_2 + B_2\text{Exp}(\alpha_2 N_{2t}) < A_3 + B_3\text{Exp}(-\alpha_3 N_{3t})$ then $Q^{t+1}(1,2) > Q^{t+1}(1,3)$.

9.3.2 Predictions of the Basic PRP and the Practice Effect on PRP with the Queuing Network Model

Based on Equation (9.7) and Lemmas 9.1 and 9.2, we can predict the simulation results of the basic PRP effect and the PRP practice effect. For the entities in task 2 (see Fig. 9.5), at the beginning of the practice phrase, because the visual-motor mappings are not established in PM [32], PM takes a longer time to process the entities than the CE and the Hicog servers. Thus, the Q value from Visk to PM ($Q(1,3)$) is lower than the Q value from Visk to CE ($Q(1,2)$). According to Lemma 9.1, the majority of the entities will go to the CE and Hicog server at the beginning of the learning process in dual tasks. Consequently, entities from task 1 also go through the CE and Hicog server thus producing a bottleneck at the Hicog server which produces the basic PRP effect. This bottleneck is similar in theory to that of Pashler [34].

During the learning process, the CE will send entities which increase the processing speed of PM based on the parallel learning mechanisms between the visual loop (including CE) and the motor loop (including PM) ([33], see Table 9.1). Therefore, when the Q value of the 2P and P route of task 2 increases, an increasing number of entities of task 2 will travel on the 2nd route and form an automatic process, which creates two parallel routes that could be traversed in this dual-task situation. However, because the learning rate of PM server (1/16,000) is lower than that of the Hicog server for the entities in task 2 (1/4,000), the majority of the entities will still go through the Hicog server.

9.3.3 Simulation Results

Figure 9.7 shows the simulation results of the basic PRP effect compared to the empirical results. The linear regression function relating the simulation and
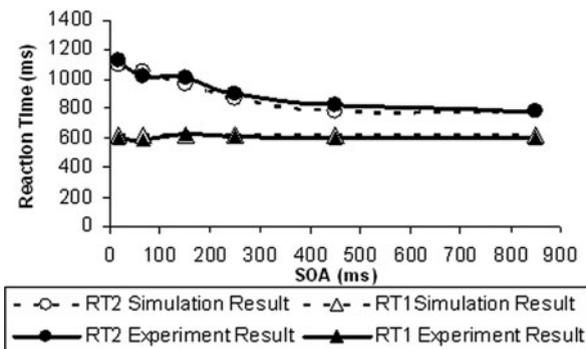

Fig. 9.7: Comparison of simulation and experiment results at the beginning of practice (basic PRP effect).

experimental results 32 is: $Y = 1.057X - 58$ (Y: experiment result; X: simulated result; R square $= 0.984$, $p < 0.001$;). Therefore, our model fits the data well.

Figure 9.8 compares of simulation and experiment results of the PRP effect at the end of practice (after 7,200 fs trials). The linear regression function relating the simulated results and experiment results is: $Y = 1.03X + 105$ (R square $= 0.891$, $p < 0.001$), therefore, our model accurately captures learning effects related to the PRP effect.
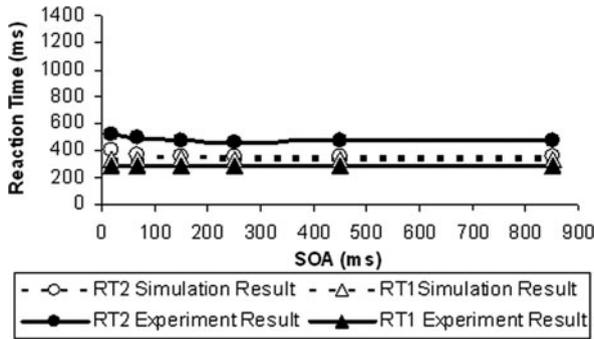
Fig. 9.8: Comparison of the simulation and experiment results at the end of practice.

Lastly, Fig. 9.9 shows the comparison of the simulation and experimental results during the practice process (7,200 trials). The linear regression function relating the simulated results and experiment results is: $Y = 0.965X + 10$ (R square $= 0.781$, $p < 0.001$). Moreover, it was found that the Q value of the second route of task 2 never exceeded that of the first route of task 2 during the practice process as

Fig. 9.9: Comparison of simulation and experiment results during the practice process (7,200 trials).

the majority of entities of task second went through the first route rather than the second route. In some ways this is supported by recent neuroimaging work on PRP by [20]. Those authors found little 33 differences in activations/neural networks in the PRP task when performance was assesse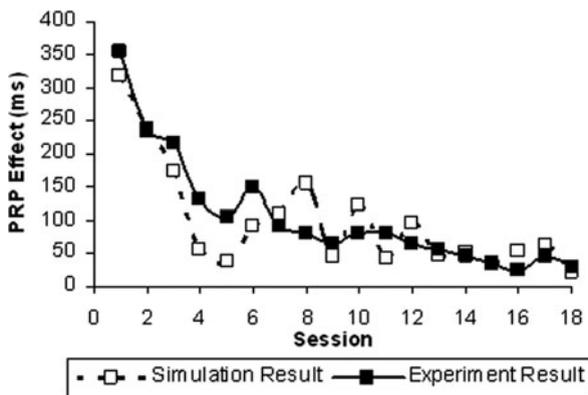d at long and short SOAs. Such large activation differences between short and long SOAs would be predicted by active monitoring theories of the PRP effect. However, Jiang et al. [20] contend that their data suggest that the PRP effect reflects passive queuing and not active monitoring. This is yet other evidence supporting the queuing network architecture and structure of our model as we did not find much difference in performance in the Hicog server before and after practice and at short and long SOAs. In addition, routes are chosen passively with Q learning and are not subject to active monitoring processes.

With the formation of an automatic process during learning, two parallel routes were formed in the dual-task situation, which partially eliminated the bottleneck at the Hicog server. The PRP effect is reduced greatly with the decrease in the processing time in both the Hicog and the PM server. However, since the majority of the entities of the two tasks still went through the Hicog server, the effect of the automatic process on PRP reduction does not exceed the effect of the reduction of RT 1 on the PRP effect. This is consistent with the result of Van Selst et al. [49] that the automatic process does grow from weak to strong but only weakly contributes to PRP reduction.

9.4 Discussion

In the previous sections of this chapter, we described the modeling of brain activation patterns as well as the behavioral phenomena in learning of two basic perceptual-motor tasks (transcription typing and PRP). In modeling the phenomena in typing, reinforcement algorithms guided how the entities traversed through different routes before and after learning. The brain areas activated both before and after learning are consistent with neuroimaging findings. In modeling PRP practice effects, we used the same simulation model to quantify the formation of automatic processes (reduction of the visual-motor task 2) during the learning processes in Van Selst et al. [49] study.

There are several questions to be answered by future research utilizing our model. First, neuroscience evidence has shown that many brain areas have overlapping functionality which was not captured by the current model, which assumed discrete brain areas with specific functions. This will increase the difficulty in modeling the cooperation of information processes in the different brain areas. Second, the traveling of entities from one server to another does not necessarily indicate the activation of two brain areas. Brain area activation as uncovered with fMRI studies is based on brain hemodynamics, which is an indirect measure of neural activity and thus has poor temporal resolution. Therefore, using fMRI data to guide modeling of processing times is somewhat tenuous. Therefore, 35 caution should be taken in comparing the simulation results of the model with the results of fMRI studies.

We are currently developing a computational model of the human cognitive system which is able to account for experimental findings in both neuroscience and

behavioral science. It is one step further to understanding the quantitative mechanisms of complex cognition and provides an alternative way to connect the brain's function with overt behavioral phenomena. We believe this current model is a firm step in this direction.

Appendix 1

Parameters setting at Hicog and PM server

- $A_{\text{Hicog-symbol}}$: minimal value of the processing time of task 2 entity in Hicog server. Since choice reaction time (RT) of four alternatives can be reduced to RT of two alternatives with practice, after intensive practice, RT of eight alternative choices in Van Selst's experiment will reduce to RT of four alternatives without intensive practice. $A_{\text{Hicog-symbol}}$ equals the RT of four alternatives (Hick's Law, intercept:150 ms, slope:170 ms/bit, Schmidt, 1988) minus one average perception cycle (100 ms), two cognitive cycles (2×70 ms), and one motor cycle (70 ms) [10]. Therefore, $A_{\text{Hicog-symbol}} = 150 + 170 \times \text{Log2}(4) - 100 - 2 \times 70 - 70 = 180$ ms.
- $B_{\text{Hicog-symbol}}$: change of processing time of task 2 entity in Hicog server at the beginning and end of practice. At the beginning of the practice in single task 2, RT of the eight alternatives (Hick's Law, intercept:150 ms, slope:170 ms/bit) is composed of one perception cycle (100 ms), maximum processing time at Hicog ($A_{\text{Hicog-symbol}} + B_{\text{Hicog-symbol}}$), and one motor cycle (70 ms) [10]. Therefore, $B_{\text{Hicog-symbol}} = 150 + 170 \times \text{Log2}(8) - 100 - A_{\text{Hicog-symbol}} - 70 = 170$ ms.
- $\alpha_{\text{Hicog-symbol}}, \alpha_{\text{Hicog-tone}}$: learning rate of Hicog server in processing the task 2 and task 1 entities. Based on $\alpha = 0.001$ approximately in Heathcote et al.'s [18] study, learning difficulty increased four times because of the four incompatible alternatives. Thus, $\alpha_{\text{Hicog-symbol}} = \alpha_{\text{Hicog-tone}} = 0.001/4 = 1/4,000$.
- $A_{\text{Hicog-tone}}$: minimal value of the processing time of task 1 entity in central executive. After intensive practice, the discrimination task of the two classes of tones in Van Selst's (1999) experiment can be simplified into a choice reaction time of two alternatives, requiring the minimum value of one cognitive cycle (25 ms) [10].
- $B_{\text{Hicog-tone}}$: change of processing time of task 1 entity in Hicog at the beginning and end of practice. At the beginning of the single task 1, the reaction time to discriminate the two classes of tone is 642 ms, which is composed of one perception cycle (100 ms), two cognitive cycles (70×2 ms), ($A_{\text{Hicog-tone}} + B_{\text{Hicog-tone}}$), and one motor cycle (70 ms). Therefore, $B_{\text{Hicog-tone}} = 642 - 100 - 2 \times 70 - A_{\text{Hicog-tone}} - 70 = 307$ ms.
- $A_{\text{PM-symbol}}$: minimal value of the processing time of task 2 entity in PM. After intensive practice, RT of the eight alternative choices in Van Selst's experiment will transform to RT of eight most compatible alternatives (RT = 217 ms, Schmidt, 1988) which is composed of one perception cycle and one motor cycle. Therefore, $A_{\text{PM-symbol}} = 217 - 100 - 70 = 47$ ms.

- $B_{\text{PM-symbol}}$: change of processing time of task 2 entity in PM at the beginning and end of practice. At the beginning of practice in single task 2, RT of eight alternative choice reaction time (Hick's law: 50 ms, slope: 170 ms/bit) is composed of one average perception cycle (100 ms), $(A_{\text{PM-symbol}} + B_{\text{PM-symbol}})$, one motor cycle (70 ms). Thus, $B_{\text{PM-symbol}} = 150 + 170 \times \text{Log2}(8) - 100 - A_{\text{PM-symbol}} - 70 = 443$ ms.
- $\alpha_{\text{PM-symbol}}$: learning rate of PM in processing the task 2 entity. The speed of formation of the automatic process in PM is slower than Hicog because it receives the entities from CE server via the indirect parallel learning mechanism with the four incompatible alternatives [33]. Thus, $\alpha_{\text{PM-symbol}} = (0.001/4)/4 = 1/16,000$.

Appendix 2

Lemma 9.1. At any transition state t ($t \neq 0$), if $1/\mu_{2,t}, t < 1/\mu_{3,t}$, then $Q_{t+1}(1,2) > Q_{t+1}(1,3)$

Proof. Using mathematic deduction method

(i) At $t = 0$: $Q^1(1,3) = Q^1(1,2) = Q^1(2,4) = Q^1(3,4) = 0$.
(ii) At $t = 1$: Using the online Q learning formula: $Q^2(1,3) = Q^1(1,3) + \varepsilon[r_t + \gamma Q^1(3,4) - Q^1(1,3)] = \varepsilon \mu_{3,1}$.

Note: because entity routes to only one server (server 4) $\max_b Q^t(S_t + 1, b) = Q(3,4), Q^2(1,2) = \varepsilon \mu_{2,1}, Q^2(3,4) = \varepsilon \mu_4, Q^2(2,4) = \varepsilon \mu_4$; If $1/\mu_{2,1} < 1/\mu_{3,1}$ then $\varepsilon \mu_{3,1} < \varepsilon \mu_{2,1}$ (given $0 < \varepsilon < 1$), i.e., $Q^2(1,2) > Q^2(1,3)$. Thus, lemma is proved at $t = 1$.

iii According to mathematic deduction method, Lemma 9.1 is correct: i.e., at transition state $t = k$: if $1/\mu_{2,k} < 1/\mu_{3,k}$ then $Q^{k+1}(1,2) > Q^{k+1}(1,3)$. We want to prove at transition state $k + 1$, lemma is still correct: i.e., At transition state $t = k + 1$:
if $1/\mu_{2,k+1} < 1/\mu_{3,k+1}$, then $Q^{k+2}(1,2) > Q^{k+2}(1,3)$ At $t = k + 1$: $Q^{k+2}(1,2) = Q^{k+1}(1,2) + \varepsilon[\mu_{2,k+1} + \gamma \varepsilon \mu_4 - Q^{k+1}(1,2)]$

$$Q^{k+2}(1,3) = Q^{k+1}(1,3) + \varepsilon[\mu_{3,k+1} + \gamma \varepsilon \mu_4 - Q^{k+1}(1,3)], \qquad (9.9)$$

$$\begin{aligned} Q^{k+2}(1,2) &- Q^{k+2}(1,3) \\ &= Q^{k+1}(1,2) + \varepsilon[\mu_{2,k+1} + \gamma \varepsilon \mu_4 - Q^{k+1}(1,2)] \\ &\quad - Q^{k+1}(1,3) + \varepsilon[\mu_{3,k+1} + \gamma \varepsilon \mu_4 - Q^{k+1}(1,3)] \\ &= (1 - \varepsilon)[Q^{k+1}(1,2) - Q^{k+1}(1,3)] + (\varepsilon \mu_{2,k+1} - \varepsilon \mu_{3,k+1}) \end{aligned} \qquad (9.10)$$

With Equation (9.3) and $0 < \varepsilon < 1$, we have

$$(1 - \varepsilon)[Q^{k+1}(1,2) - Q^{k+1}(1,3)] > 0. \qquad (9.11)$$

Given $1/\mu_{2,k+1} < 1/\mu_{3,k+1}$ and $0 < \varepsilon < 1$, then $(\varepsilon\mu_{2,k+1} - \varepsilon, \mu_{3,k+1}) > 0$, i.e., $Q^{k+2}(1,3) - Q^{k+2}(1,2) > 0$

Thus, Lemma 9.1 is correct at $t = k+1$. Lemma 9.1 is proved.

References

1. Anderson, J., Lebiere, C. The Atomic Components of Thought. Erlbaum, Mahwah, NJ (1998)
2. Bear, M., Connors, B., Paradiso, M. Neuroscience: Exploring the Brain. Lippincott Williams & Wilkins Publisher, Baltimore, MD (2001)
3. Berman, M., Jonides, J., Nee, D. Studying mind and brain with fMRI. Soc Cogn Affect Neurosci **1**(2), 158–161 (2006)
4. Berman, M., Liu, Y., Wu, C. A 3-node queuing network template of cognitive and neural differences as induced by gray and white matter changes. In: Proceedings of the 8th International Conference on Cognitive Modeling, pp. 175–180. Ann Arbor, MI (2007)
5. Bherer, L., Kramer, A., Peterson, M., Colcombe, S., Erickson, K., Becic, E. Testing the limits of cognitive plasticity in older adults: Application to attentional control. Acta Psychologica **123**(3), 261–278 (2006)
6. Bressler, S. Large-scale cortical networks and cognition. Brain Res Rev **20**, 288–304 (1995)
7. Byrne, M., Anderson, J. Serial modules in parallel: The psychological refractory period and perfect time-sharing. Psychol Rev **108**(4), 847–869 (2001)
8. Cabeza, K. Handbook of Functional Neuroimaging of Cognition, 2nd edn. MIT Press, Cambridge, MA (2006)
9. Cabeza, R., Nyberg, L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. J Cogn Neurosci **12**(1), 1–47 (2000)
10. Card, S., Moran, T., Newell, A.N. Handbook of perception and human performance, chap. The Model Human Processor: An Engineering Model of Human Performance. Wiley, New York (1986)
11. Dell'Acqua, R., Sessa, P., Pashler, H. A neuropsychological assessment of dual-task costs in closed-head injury patients using Cohen's effect size estimation method. Psychol Res-Psychologische Forschung **70**(6), 553–561 (2006)
12. Deseilligny, C., Muri, M. Cortical control of ocular saccades in humans: A model for motricity. In: Neural Control of Space Coding and Action Production, Progress in Brain Research, Vol. 142, pp. 1–19. Elsevier, New York (2003)
13. Donders, F. Attention and Performance 2, chap. Over de snelheid van psychische processen [On the speed of mental processes]. North-Holland, Amsterdam (1969). (Original work published in 1869)
14. Feyen, R., Liu, Y. Modeling task performance using the queuing network model human processor (qnmhp). In: Proceedings of the 4th International Conference on Cognitive Modeling, pp. 73–78 (2001)
15. Fletcher, P., Henson, R. Frontal lobes and human memory – insights from functional neuroimaging. Brain **124**, 849–881 (2001)
16. Gentner, D. The acquisition of typewriting skill. Acta Psychol **54**, 233–248 (1983)
17. Gordon, A., Soechting, J. Use of tactile afferent information in sequential finger movements. Exp Brain Res **107**, 281–292 (1995)
18. Heathcote, A., Brown, S., Mewhort, D. The power law repealed: The case for an exponential law of practice. Psychol Bull **7**(2), 185–207 (2000)
19. Hinton, B. Using coherence assumptions to discover the underlying causes of the sensory input. In: Connectionism: Theory and Practice (Vancouver Series in Cognitive Science) (1992). (Paperback – Aug 20, 1992)
20. Jiang, Y., Saxe, R., Kanwisher, N. Functional magnetic resonance imaging provides new constraints on theories of the psychological refractory period. Psychol Sci **15**(6), 390–396 (2004)

21. John, B. Typist: A theory of performance in skilled typing. Hum Comput Interact **11**, 321–355 (1996)
22. John, B. Contributions to engineering models of human-computer interaction. Ph.D. Thesis, Department of Psychology, Carnegie-Mellon University (1988)
23. Jueptner, M., Weiller, C. A review of differences between basal ganglia and cerebellar control of movements as revealed by functional imaging studies. Brain **121**, 1437–1449 (1998)
24. Laird, J., Newell, A., Rosenbloom, P. Soar: An architecture for general intelligence. Artif Intell **33**, 1–64 (1987)
25. Levy, J., Pashler, H., Boer, E. Central interference in driving – is there any stopping the psychological refractory period? Psychol Sci **17**(3), 228–235 (2006)
26. Liu, Y. Queuing network modeling of elementary mental processes. Psychol Rev **103**, 116–136 (1996)
27. Liu, Y. Queuing network modeling of human performance of concurrent spatial and verbal tasks. IEEE Trans Syst, Man, Cybern **27**, 195–207 (1997)
28. Liu, Y., Feyen, R., Tsimhoni, O. Queuing network-model human processor (QN-MHP):A computational architecture for multi-task performance in humanmachine systems. ACM Trans Comput-Hum Interact **13**(1), 37–70 (2006)
29. Luck, S. An Introduction to the Event-Related Potential Technique, The MIT Press, Boston, MA (2005)
30. McCloskey, M. Networks and theories – the place of connectionism in cognitive science. Psychol Sci **2**(6), 387–395 (1991)
31. Meyer, D., Kieras, D. A computational theory of executive cognitive processes and multiple-task performance.1. basic mechanisms. Psychol Rev **104**(1), 3–65 (1997)
32. Mitz, A., Godschalk, M., Wise, S. Learning-dependent neuronal-activity in the premotor cortex – activity during the acquisition of conditional motor associations. J Neurosci **11**(6), 1855–1872 (1991)
33. Nakahara, H., Doya, K., Hikosaka, O. Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences. J Cogn Neurosci **13**(5), 626–647 (2001)
34. Pashler, H. Processing stages in overlapping tasks – evidence for a central bottleneck. J Exp Psychol-Hum Percept Perform **10**(3), 358–3 (1984)
35. Rektor, I., Kanovsky, P., Bares, M. A SEEG study of ERP in motor and premotor cortices and in the basal ganglia. Clin Neurophysiol **114**, 463–471 (2003)
36. Roland, P. Brain Activation. John Wiley & Sons, New York (1993)
37. Romero, D., Lacourse, M., Lawrence, K. Event-related potentials as a function of movement parameter variations during motor imagery and isometric action. Behav Brain Res **117**, 83–96 (2000)
38. Rudell, A., Hu, B. Does a warning signal accelerate the processing of sensory information? evidence from recognition potential responses to high and low frequency words. Int J Psychophysiol **41**, 31–42 (2001)
39. Rumelhart, D., McClelland, J. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, Cambridge, MA (1986)
40. Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Sasaki, Y., Putz, B. Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. J Neurosci **18**(5), 1827–1840 (1998)
41. Salthouse, T. Effects of age and skill in typing. J Exp Psychol: General **113**, 345–371 (1984)
42. Salthouse, T. Perceptual, cognitive, and motoric aspects of transcription typing. Psychol Bull **99**(3), 303–319 (1986)
43. Salthouse, T., Saults, J. Multiple spans in transcription typing. J Appl Psychol **72**(2), 187–196 (1987)
44. Selst, M.V., Ruthruff, E., Johnston, J. Can practice eliminate the psychological refractory period effect? J Exp Psychol: Hum Percept Perform **25**(5), 1268–1283 (1999)
45. Sutton, R., Barto, A. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA (1998)

46. Tsimhoni, O., Liu, Y. Modeling steering with the queuing network-model human processor (QN-MHP). In: Proceedings of the 47th Annual Conference of the Human Factors and Ergonomics Society, pp. 81–85 (2003)
47. Ungerleider, L.G., Doyon, J., Karni, A. Imaging brain plasticity during motor skill learning. Neurobiol Learn Mem **78**, 553–564 (2002)
48. Wu, C., Liu, Y. Modeling behavioral and brain imaging phenomena in transcription typingwith queuing networks and reinforcement learning algorithms. In: Proceedings of the 6th International Conference on Cognitive Modeling (ICCM-2004), pp. 314–319. Pittsburgh, PA (2004)
49. Wu, C., Liu, Y. Modeling human transcription typing with queuing networkmodel human processor. In: Proceedings of the 48th Annual Meeting of Human Factors and Ergonomics Society, Vol. **48**(3), pp. 381–385. New Orleans, Louisiana (2004)
50. Wu, C., Liu, Y. Modeling psychological refractory period (PRP) and practice effect on 41 PRP with queueing networks and reinforcement learning algorithms. In: Proceedings of the 6th International Conference on Cognitive Modeling (ICCM-2004), pp. 320–325. Pittsburgh, PA (2004)
51. Wu, C., Liu, Y. Modeling psychological refractory period (PRP) and practice effect on PRP with queuing networks and reinforcement learning algorithms. In: Proceedings of the Sixth International Conference on Cognitive Modeling, pp. 320–325. Pittsburgh, PA (2004)
52. Wu, C., Liu, Y. Modeling fMRI bold signal and reaction time of a dual task with a 42 queuing network modeling approach. In: 28th Annual Conference of the Cognitive Science Society. Vancouver, BC, Canada (2006)
53. Wu, C., Liu, Y. Queuing network modeling of a real-time psychophysiological index of mental workload.p300 amplitude in event-related potential (ERP). In: Paper presented at the 50th Annual Conference of the Human Factors and Ergonomics Society. San Francisco, CA (2006)
54. Wu, C., Liu, Y. Queuing network modeling of age differences in driver mental workload and performance. In: 50th Annual Conference of the Human Factors and Ergonomics Society. San Francisco, CA (2006)
55. Wu, C., Liu, Y. Queuing network modeling of driver workload and performance. In: 50th Annual Conference of the Human Factors and Ergonomics Society. San Francisco, CA (2006)
56. Wu, C., Liu, Y. Queuing network modeling of reaction time, response accuracy, and stimulus-lateralized readiness potential onset time in a dual task. In: 28th Annual Conference of the Cognitive Science Society. Vancouver, BC, Canada (2006)
57. Wu, C., Liu, Y. A new software tool for modeling human performance and mental workload. Q Hum Factors Appl: Ergon Des **15**(2), 8–14 (2007)
58. Wu, C., Liu, Y. Queueing network modeling of transcription typing. ACM Trans Comput-Hum Interact **15**(1), Article No.:6 (2007)
59. Wu, C., Liu, Y. Queuing network modeling of driver workload and performance. IEEE Trans Intell Transport Syst **8**(3), 528–537 (2007)
60. Wu, C., Liu, Y., Tsimhoni, O. Application of schedulingmethods in designing multimodal in-vehicle systems. In: Society of Automobile Engineers (SAE) World Congress. SAE, Detroit, MI (2008)
61. Wu, C., Liu, Y., Walsh, C. Queuing network modeling of a real-time psychophysiological index of mental workload–p300 in event-related potential (ERP). IEEE Trans Syst, Man, Cybern (Part A) **38**(5), pp. 1068–1084 (2007)
62. Wu, C., Tsimhoni, O., Liu, Y. Development of an adaptive workload management system using queuing network-model of human processor. In: The 51st Annual Conference of the Human Factors and Ergonomics Society. Baltimore, MD (2007)
63. Wu, C., Tsimhoni, O., Liu, Y. Development of an adaptive workload management system using queueing network-model of human processor. IEEE Intell Transp Syst **9**(3), 463–475 (2008)

Chapter 10
Neural Network Modeling of Voluntary Single-Joint Movement Organization I. Normal Conditions

Vassilis Cutsuridis

Abstract Motor learning and motor control have been the focus of intense study by researchers from various disciplines. The neural network model approach has been very successful in providing theoretical frameworks on motor learning and motor control by modeling neural and psychophysical data from multiple levels of biological complexity. Two neural network models of voluntary single-joint movement organization under normal conditions are summarized here. The models seek to explain detailed electromyographic data of rapid single-joint arm movement and identify their neural substrates. The models are successful in predicting several characteristics of voluntary movement.

10.1 Introduction

Voluntary movements are goal-directed movements triggered either by internal or external cues. Voluntary movements can be improved with practice as one learns to anticipate and correct for environmental obstacles that perturb the body. Single-joint rapid (ballistic) movements are goal-directed movements performed in a single action, without the need for corrective adjustments during its course. They are characterized by a symmetric bell-shaped velocity curve, where the acceleration (the time from the start to the peak velocity) and deceleration (the time from the peak velocity to the end of movement) times are equal [3]. Similar velocity profiles have also been observed in multi-joint movements [11].

The electromyographic (EMG) pattern of single-joint rapid voluntary movements in normal subjects is also very characteristic. It is characterized by alternating bursts of agonist and antagonist muscles [28]. The first agonist burst provides the impulsive force for the movement, whereas the antagonist activity provides the braking force

Vassilis Cutsuridis
Centre for Memory and Brain, Boston University, Boston, MA, USA
e-mail: vcut@bu.edu

to halt the limb. Sometimes a second agonist burst is needed to bring the limb to the final position [1, 4, 5, 6, 23, 24, 25, 26, 27, 36]. The combination of the agonist–antagonist–agonist bursts is known as the triphasic pattern of muscle activation [28]. An excellent review on the properties of the triphasic pattern of muscle activation and the produced movement under different experimental conditions can be found in Berardelli and colleagues [2].

The origin of the triphasic pattern and whether it is controlled by the nervous system has been long debated [33]. In a review paper by Berardelli and colleagues [2], three conclusions were made: (1) the basal ganglia output plays a role in the scaling of the first agonist burst size, (2) the corticospinal tract has a role in determining spatial and temporal recruitment of motor units, and (3) the proprioceptive feedback is not necessary to the production of the triphasic pattern, but it contributes to the accuracy of both the trajectory and the end point of ballistic movements. That means that the origin of the triphasic pattern of muscle activation *may* be a central one, but afferent inputs can also modulate the voluntary activity.

10.2 Models and Theories of Motor Control

Motor learning and motor control have been the focus of intense study by researchers from various disciplines. The experimental researchers interested in motor learning investigate how practice facilitates skill acquisition and improvement. The theoretical/computational researchers interested in motor control have investigated which movement variables are controlled during movement from the nervous system [33]. Many computational models of motor control have been advanced over the years [14]. These models include the equilibrium point hypothesis [20], dynamical system theory [32], the pulse-step model [22], the impulse-timing model [35], the dual-strategy hypothesis [14], models about minimizing movement variables [34], and neural network models [8, 9, 10, 13, 17, 15, 16, 18].

The neural network model approach has been very successful in providing theoretical frameworks on motor learning and motor control by modeling neural and psychophysical data from multiple levels of biological complexity. In particular, the vector integration to endpoint (VITE) and factorization of muscle length and muscle tension (FLETE) neural network models of Bullock, Grossberg, and colleagues [7,8,9,10,13] have provided qualitative answers to questions such as how can a limb be rotated to and stabilized at a desired angle? How can movement speed from an initial to a desired final angle be controlled under conditions of low joint stiffness? How can launching and braking forces be generated to compensate from inertial loads? The VITE model was capable of generating single-joint arm movements, whereas the FLETE model afforded independent voluntary control of joint stiffness and joint position, and incorporated second-order dynamics, which played a large role in realistic limb movements. Variants of the FLETE model [9] have been successful in producing realistic transient muscle activations, such as the triphasic pattern of muscle activation observed during rapid, self-terminated movements.

Despite their successes, the VITE and FLETE models have several limitations. First, in an attempt to simulate the joint movement and joint stiffness, Bullock and Grossberg speculated the presence of the two partly independent cortical processes [30], a reciprocal signal of antagonist muscles responsible for the joint rotation, and a co-contraction signal of antagonist muscle responsible for joint stiffness. However, neither the VITE-FLETE model studies [9] nor the Humphrey and Reed [30] experimental study has identified the exact neural correlates (i.e., cell types) for the reciprocal activation and co-contraction of antagonist muscles.

Second, they failed to provide functional roles of experimentally identified neurons in primary motor cortex (area 4) and parietal cortex (area 5), such as the phasic
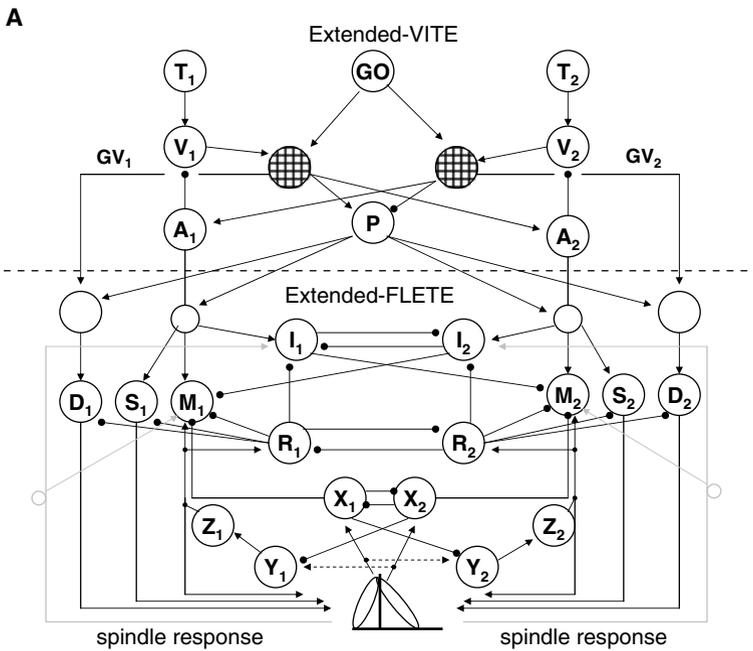
Fig. 10.1: Extended VITE–FLETE models without dopamine (DA). (A and B) *Top*: Extended-VITE model for variable-speed trajectory generation. *Bottom*: Extended-FLETE model of the opponent processing spinomuscular system. *Arrow lines*: excitatory projections; *solid dot lines*: inhibitory projections; *dotted arrow lines*: feedback pathways from sensors embedded in muscles. *GO*: basal ganglia output signal; *P*: bidirectional co-contractive signal; *T*: target position command; *V*: DV activity; *GV*: DVV activity; *A*: current position command; *M*: alpha motoneuronal (MN) activity; *R*: renshaw cell activity; *X, Y, Z*: spinal inhibitory interneuron (IN) activities; I_a: spinal type a inhibitory IN activity; *S*: static gamma MN activity; *D*: dynamic gamma MN activity; *1,2*: antagonist cell pair.

Fig. 10.1: (continued)

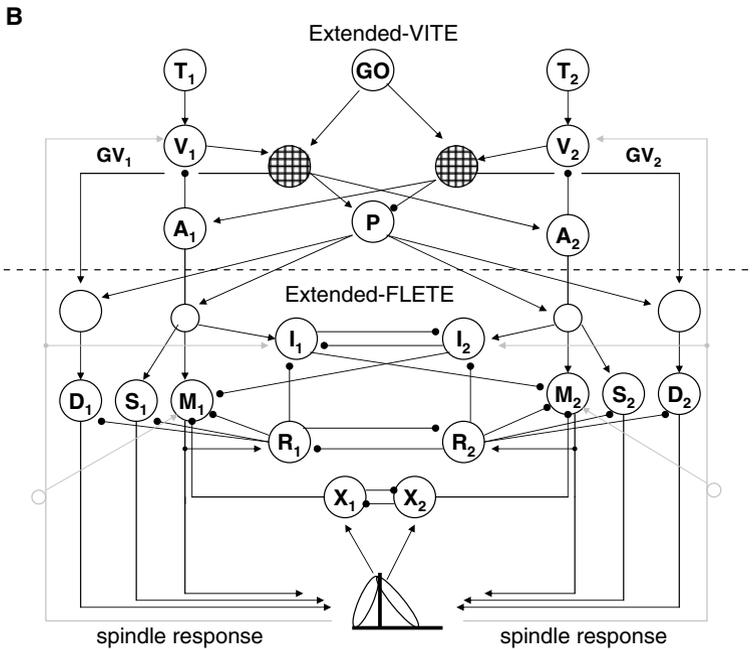cells, the tonic cells, the reciprocal cells, and the bidirectional cells known to play a role in voluntary movement initiation and control [19, 21].

Third, they failed to identify the site of origin of the triphasic pattern of muscle activation [2]. Is the triphasic pattern cortically or subcortically originated [2]? If cortically originated, are the agonist and antagonist bursts generated from experimentally identified cortical cell types? Does the afferent feedback from the muscle spindles to the spinal cord play any role in maintenance of this pattern? Does the feedback from the muscle spindles to the cortex play a role in the generation of the second agonist burst?

These limitations were addressed successfully by the extended VITE–FLETE with dopamine models of Cutsuridis and Perantonis [18] and Cutsuridis [15, 16, 17]. These models have answered issues concerning voluntary movement and proprioception in normal and Parkinsonian conditions. The temporal development of these models in normal conditions (i.e., without dopamine) is reviewed in detail in the next section.

10.3 The Extended VITE–FLETE Models Without Dopamine

Figures 10.1a, b depict the extended VITE–FLETE models without dopamine of voluntary movement and proprioception [15, 16, 17, 18]. Both models were based

on known corticospinal neuroanatomical connectivity. Detailed description and complete mathematical formalism of the models can be found in Cutsuridis and Perantonis [18] and Cutsuridis [15, 17]. Both extended VITE–FLETE without dopamine models, while they preserved the original VITE–FLETE model's capability of generating rapid single-joint movements and affordance of independent voluntary control of joint stiffness and joint movement, they were extended it in three fundamental ways.

In a behavioral neurophysiology task, Doudet and colleagues [19] trained monkeys to perform fast flexion and extension elbow movements while they recorded from their primary motor cortex. Three classes of movement-related neurons were identified according to their activity during the movement: (1) neurons showing a reciprocal discharge pattern for flexion and extension movements (reciprocal neurons), (2) neurons whose activity changed for only one direction (unidirectional neurons), and (3) neurons whose activity decreased or increased for both directions of movement (bidirectional neurons). In the extended VITE–FLETE with dopamine model of Figure 10.1a [15, 16, 18] functional roles to the cortically identified reciprocal [19], bidirectional [19], phasic MT and tonic neurons were assigned. An arm movement difference vector (DV) was computed in parietal area 5 from a comparison of a target position vector (TPV) with a representation of the current position called perceived position vector (PPV). The DV signal then projected to area 4, where a desired velocity vector (DVV) and a nonspecific co-contractive signal (P) [30] were formed. A voluntarily scalable GO signal multiplied (i.e., gated) the DV input to both the DVV and the P in area 4, and thus volitional sensitive velocity and nonspecific co-contractive commands were generated, which activated the lower spinal centers. The DVV and P signals corresponded to two partly independent neuronal systems with the motor cortex [30].

The output of the basal ganglia (BG) system, which represented the activity of the GPi was modeled by a GO signal:

$$G(t) = G_0(t - \tau_i)^2 u[t - \tau_i]/(\beta + \gamma(t - \tau_i)^2), \tag{10.1}$$

where G_0 amplified the G signal, i was the onset time of the ith volitional command, β and γ are free parameters, and $u[t]$ was a step function that jumped from 0 to 1 to initiate movement. The difference vector (DV), which represented cortical area's 5 phasic cell activity, was described by

$$\frac{dV_i}{dt} = 30(-V_i + T_i - A_i), \tag{10.2}$$

where T_i was the target position command and A_i was the current limb position command.

In contrast to the original VITE–FLETE model [9, 10], in the extended VITE–FLETE models [15, 16, 17, 18], the desired velocity vector (DVV) represented the activity of cortical area's 4 phasically activated reciprocal neurons [19], and it was organized for the reciprocal activation of antagonist muscles. It was defined by

$$u_i = [G(V_i - V_j) + B_u]^+, \tag{10.3}$$

where i, j designated opponent neural commands and B_u was the baseline activity of the phasic-MT area 4 cell activity.

The co-contractive vector (P) represented area's 4 phasic activity of bidirectional neurons (i.e., neurons whose activity decreases or increases for both directions of movement [19]), and it was organized for the co-contraction of antagonist muscles (see columns 1 and 3 of Fig. 10.2). It was given by

$$P = [G(V_i - V_j) + B_P]^+. \tag{10.4}$$

Fig. 10.2: Comparison of peristimulus time histograms (PSTH) of reciprocally organized neurons (column 1; reproduced with permission from [19, Fig. 4A, p. 182], Copyright Springer-Verlag) in area 4, simulated area's 4 reciprocally organized phasic (DVV) cell activities (column 2), PSTH of area's 4 bidirectional neurons (column 3; reproduced with permission from [19, Fig. 4A, p. 182], Copyright Springer-Verlag) and simulated area's 4 co-contractive (P) cells activities (column 4) for a flexion (row 1) and extension (row 2) movements in normal monkey. The vertical bars indicate the onset of movement. Note a clear triphasic AG1-ANT1-AG2 pattern marked with arrows is evident in PSTH of reciprocally and bidirectionally organized neurons. The second AG2 burst is absent in simulated DVV cell activities.

While the reciprocal pattern of muscle activation served to move the joint from an initial to a final position, the antagonist co-contraction served to increase the apparent mechanical stiffness of the joint, thus fixing its posture or stabilizing its course of movement in the presence of external force perturbations. The Renshaw population cell activity was modelled by

$$\frac{dR_i}{dt} = \phi(\lambda B_i - R_i)z_i \max(M_i, 0) - R_i(1.5 + \max(R_j, 0)), \tag{10.5}$$
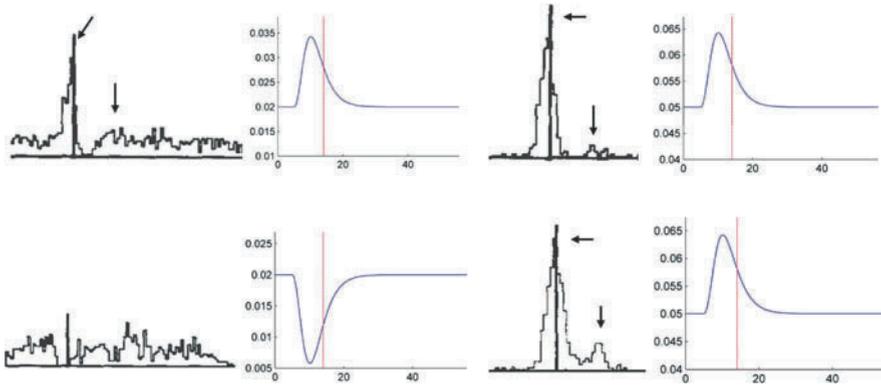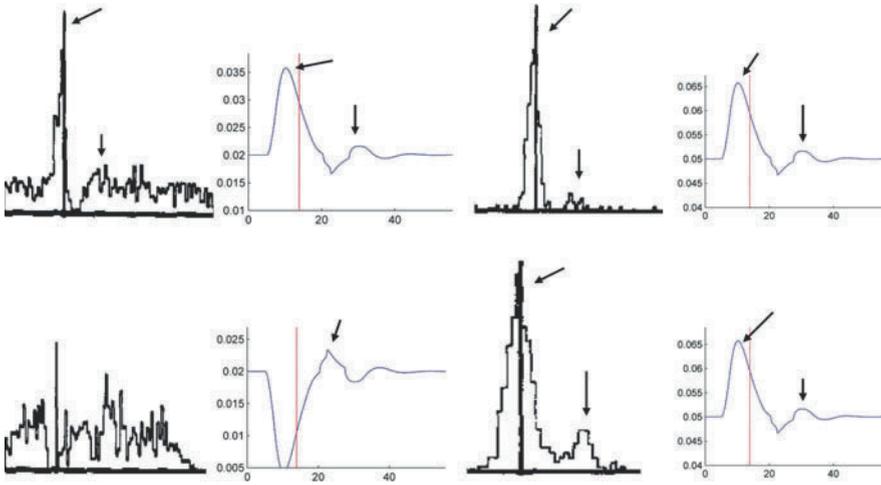
Fig. 10.3: Comparison of peristimulus time histograms (PSTH) of reciprocally organized neurons (column 1; reproduced with permission from [19, Fig. 4A, p. 182], Copyright Springer-Verlag) in area 4, simulated area's 4 reciprocally organized phasic (DVV) cell activities (column 2), PSTH of area's 4 bidirectional neurons (column 3; reproduced with permission from [19, Fig. 4A, p. 182], Copyright Springer-Verlag), and simulated area's 4 co-contractive (P) cells activities (column 4) for a flexion (row 1) and extension (row 2) movements in normal monkey. The vertical bars indicate the onset of movement. Note a clear triphasic AG1-ANT1-AG2 pattern marked with arrows is evident in PSTH of reciprocally and bidirectionally organized neurons. The same triphasic pattern is evident in simulated DVV cell activities. The second peak in simulated activities marked with an arrow arises from the spindle feedback input to area's 5 DV activity.

whereas the $\alpha - MN$ population activity was described by

$$\frac{dM_i}{dt} = \phi(\lambda B_i - M_i) \cdot (A_i + P + \chi \cdot E_i) - (M_i + 2) \cdot (1 + \Omega \cdot \max(R_i, 0)$$
$$+ \rho \cdot \max(X_i, 0) + \max(I_j', 0)), \tag{10.6}$$

where X_i was the type I_b interneuron (I_bIN) force feedback, E_i was the stretch feedback, and I_j was the type I_a interneuron. The type I_a interneuron (I_aIN) population activity was defined as

$$\frac{dI_i}{dt} = \phi \cdot (15 - I_i) \cdot (A_i + P + \chi E_i) - I_i(1 + \Omega \cdot \max(R_i, 0) + \max(I_j', 0)). \tag{10.7}$$

The I_bIN population activity without dopamine was given by

$$\frac{dX_i}{dt} = \phi \cdot (15 - X_i)F_i - X_i \cdot (0.8 + 2.2 \max(X_j, 0)), \tag{10.8}$$

where F_i was the feedback activity of force-sensitive Golgi tendon organs.

While the extended model was successful in simulating the neuronal activity of the reciprocal and bidirectional cells and proposed for the functional roles in joint movement and stiffness, it failed to simulate the second agonist burst of both the reciprocal and the bidirectional neurons (see columns 1 and 3 of figures 10.2 and 10.3). Due to this failure a biphasic (not triphasic) pattern of α-motorneuronal activation is produced (see Fig. 10.4A). As mentioned earlier, the role of the second agonist burst of the triphasic pattern is to clamp the limb to its final position [29].

To simulate the second observed burst in the reciprocal and bidirectional discharge patterns as well as in the α-MN activities, the extended VITE–FLETE model of Fig.10.1a [15, 16, 18] was further extended (see Fig. 10.1B) by incorporating the effect of the neuroanatomically observed muscle spindle feedback to the cortex [17]. To model this effect, equation (10.2) was changed to

$$\frac{dV_i}{dt} = 30(-V_i + T_i - A_i + a_w \cdot (W_i(t - \tau) - W_j(t - \tau))), \qquad (10.9)$$

where T_i was the target position command, A_i was the current limb position command, a_w was the gain of spindle feedback, and $W_{i,j}$ were the spindle feedback signals from the antagonist muscles. A clear triphasic AG_1-ANT_1-AG_2 reciprocal pattern of cellular activity is evident in figure (column 1 of figure 10.3). Similarly, the activity of bidirectional neurons tuned to both directions of movement is also shown (column 3 of figure 10.3). The simulated marked by an arrow first peak of extension and second peak of flexion reciprocal cells is primarily due to spindle feedback input to DV activity (a feature lacking in [18]). This cortical triphasic pattern of neuronal activation then drives the antagonist α-MNs and produces at their level a similar triphasic pattern of muscle activation (see Fig. 10.4B).
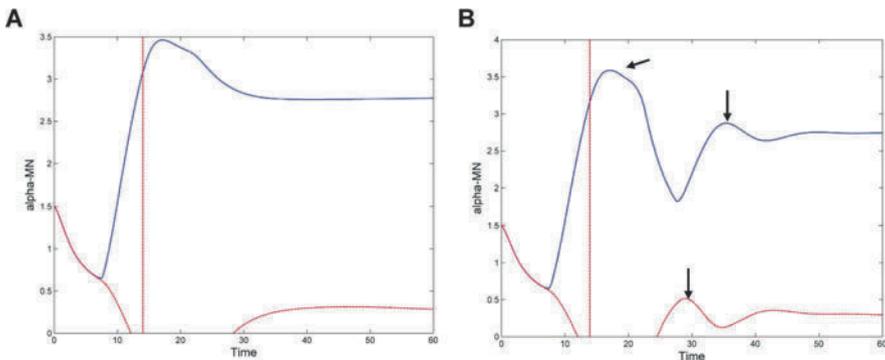

Fig. 10.4: (A) Simulated α-MN activity when the muscle spindle feedback is absent from the cortex. Note a pronounced biphasic AG1-ANT1 pattern of muscle activation. The second AG2 bursts are absent. (B) Simulated α-MN activity when the muscle spindle feedback is present in the cortex. Note a clear triphasic AG1-ANT1-AG2 pattern of muscle activation.

10.4 Conclusion

This chapter has focused on two neural network models of voluntary movement and proprioception under normal conditions. The models seek to explain detailed electromyographic data of rapid single-joint arm movement and identify their neural substrates. The models were successful in providing answers to the questions detailed in the previous sections as well as predicting several characteristics of voluntary movement:

- The reciprocal and bidirectional neurons in primary motor cortex [19] are the two partly independent cortical processes [30] for the reciprocal activation and co-contraction of antagonist muscles in the control of joint rotation and joint stiffness.
- The origin of the triphasic pattern of muscle activation in normal conditions is predicted to be cortical.
- The neural substrates of the triphasic pattern of muscle activation in normal conditions are predicted to be the neuronal discharge patterns of the reciprocal neurons in primary motor cortex.
- The afferent feedback from the muscle spindles to the cortex is responsible for the generation of second agonist burst in the neuronal and EMG activities that clamp the limb to its final position.

Many more predictions regarding voluntary movement control under normal conditions can be found in [18,15,16,17]. In the next chapter, issues regarding *voluntary movement disorganization in Parkinson's disease* will be addressed. In particular, what role, if any, does dopamine depletion in key cortical and spinal cord sites play in the initiation, execution, and control of voluntary movements in Parkinson's disease patients? Does dopamine depletion in basal ganglia, cortex, and spinal cord have any effect on the triphasic pattern of muscle activation? How do the neuronal and EMG variables change when dopamine is depleted?

References

1. Berardelli, A., Dick, J., Rothwell, J., Day, B., Marsden, C. Scaling of the size of the first agonist EMG burst during rapid wrist movements in patients with Parkinson's disease. J Neurol Neurosurg Psych **49**, 1273–1279 (1986)
2. Berardelli, A., Hallett, M., Rothwell, J., Agostino, R., Manfredi, M., Thompson, P., Marsden, C. Single-joint rapid arm movement in normal subjects and in patients with motor disorders. Brain **119**, 661–674 (1996)
3. Britton, T., Thompson, P., Day, B., Rothwell, J., Findley, L., Marsden, C. Rapid wrist movements in patients with essential tremor. The critical role of the second agonist burst. Brain **117**, 39–47 (1994)
4. Brown, S., Cooke, J. Initial agonist burst duration depends on movement amplitude. Exp Brain Res **55**, 523–527 (1984)
5. Brown, S., Cooke, J. Movement related phasic muscle activation I. Relations with temporal profile of movement. J Neurophys **63**(3), 455–464 (1990)

6. Brown, S., Cooke, J. Movement related phasic muscle activation II. Generation and functional role of the triphasic pattern. J Neurophysiol **63**(3), 465–472 (1990)

7. Bullock, D., Grossberg, S. Neural dynamics of planned arm movements: Emergent inva-riants and speed-accuracy properties during trajectory formation. Psychol Rev **95**, 49–90 (1988)

8. Bullock, D., Grossberg, S. VITE and FLETE: Neural modules for trajectory formation and tension control. In: Volitional Action. North-Holland, Amsterdam, The Netherlands (1989). 253–277

9. Bullock, D., Grossberg, S. Adaptive neural networks for control of movement trajectories invariant under speed and force rescaling. Hum Mov Sci **10**, 3–53 (1991)

10. Bullock, D., Grossberg, S. Emergence of triphasic muscle activation from the nonlinear inter-actions of central and spinal neural networks circuits. Hum Mov Sci **11**, 157–167 (1992)

11. Camarata, P., Parker, R., Park, S., Haines, S., Turner, D., Chae, H., Ebner, T. Effects of MPTP induced hemiparkinsonism on the kinematics of a two-dimensional, multi-joint arm movement in the rhesus monkey. Neuroscience **48**(3), 607–619 (1992)

12. Chapman, C.E., Spidalieri, G., Lamarre, Y. Discharge properties of area 5 neurons during arm movements triggered by sensory stimuli in the monkey. Brain Res **309**, 63–77 (1984)

13. Contreras-Vidal, J., Grossberg, S., Bullock, D. A neural model of cerebellar learning for arm movement control: Cortico-spino-cerebellar dynamics. Learn Mem **3**(6), 475–502 (1997)

14. Corcos, D.M., Jaric, S., Gottlieb, G. Electromyographic analysis of performance enhance-ment. In: Zelaznik, H.N. (ed.) Advances in Motor Learning and Control. Human Kinetics, Vancouver, BC (1996)

15. Cutsuridis, V. Neural model of dopaminergic control of arm movements in Parkin-son's dis-ease Bradykinesia. Artificial Neural Networks, *LNCS*, Vol. 4131, pp. 583–591. Springer-Verlag, Berlin (2006)

16. Cutsuridis, V. Biologically inspired neural architectures of voluntary movement in nor-mal and disordered states of the brain. Ph.D. Thesis (2006). Unpublished Ph.D. dissertation. http://www.cs.stir.ac.uk/ vcu/papers/PhD.pdf

17. Cutsuridis, V. Does reduced spinal reciprocal inhibition lead to co-contraction of antagonist motor units? a modeling study. Int J Neural Syst **17**(4), 319–327 (2007)

18. Cutsuridis, V., Perantonis, S. A neural model of Parkinson's disease bradykinesia. Neural Netw **19**(4), 354–374 (2006)

19. Doudet, D., Gross, C., Arluison, M., Bioulac, B. Modifications of precentral cortex dis-charge and EMG activity in monkeys with MPTP induced lesions of DA nigral lesions. Exp Brain Res **80**, 177–188 (1990)

20. Feldman, A. Once more on the equilibrium-point hypothesis (λ model) for motor control. J Mot Behav **18**, 17–54 (1986)

21. Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., Massey, J.T. On the relations between the direction of two dimensional arm movements and cell discharge in primate motor cortex. J Neurosci **2**, 1527–1537 (1982)

22. Ghez, C. Integration in the nervous system, chap. Contributions of Central Programs to Rapid Limb Movement in the Cat, pp. 305–319. Igaku-Shoin, Tokyo (1979)

23. Ghez, C., Gordon, J. Trajectory control in targeted force impulses. I. Role in opposing muscles. Exp Brain Res **67**, 225–240 (1987)

24. Ghez, C., Gordon, J. Trajectory control in targeted force impulses. II. Pulse height control. Exp Brain Res **67**, 241–252 (1987)

25. Ghez, C., Gordon, J. Trajectory control in targeted force impulses. III. Compensatory adjust-ments for initial errors. Exp Brain Res **67**, 253–269 (1987)

26. Gottlieb, G., Latash, M., Corcos, D., Liubinskas, A., Agarwal, G. Organizing principle for sin-gle joint movements: I. agonist-antagonist interactions. J Neurophys **13**(6), 1417–1427 (1992)

27. Hallett, C.M., Marsden, G. Ballistic flexion movements of the human thumb. J Physiol **294**, 33–50 (1979)

28. Hallett, M., Shahani, B., Young, R. EMG analysis of stereotyped voluntary movements in man. J Neurol Neurosurg Psychiatr **38**, 1154–62 (1975)

29. Hannaford, B., Stark, L. Roles of the elements of the triphasic control signal. Exp Neurol **90**, 619–634 (1985)

30. Humphrey, D., Reed, D. Separate cortical systems for control of joint movement and joint stiffness: Reciprocal activation and coactivation of antagonist muscles. In: Desmedt, J.E. (ed.) Motor Control Mechanisms in Health and Disease, pp. 347–372. Raven Press, New York (1983)

31. Kalaska, J.F., Cohen, D.A.D., Prud'homme, M.J., Hude, M.L. Parietal area 5 neuronal activity encodes movement kinematics, not movement dynamics. Exp Brain Res **80**, 351–364 (1990)

32. Schoner, G., Kelso, G. Dynamic pattern generation in behavioural and neural systems. Science **239**, 1513–1520 (1988)

33. Stein, R. What muscle variable(s) does the nervous system control in limb movements? Behav Brain Sci **5**, 535–577 (1982)

34. Uno, Y., Kawato, M., Suzuki, R. Formation and control of optimum trajectory in human multijoint arm movement-minimumtorque-changemodel. Biol Cybern **61**, 89–101 (1989)

35. Wallace, S. An impulse-timing theory for reciprocal control of muscular activity in rapid, discrete movements. J Mot Behav **13**, 1144–1160 (1981)

36. Wierzbicka, M.,Wiegner, A., Shahani, B. Role of agonist and antagonist muscles in fast arm movements in man. Exp Brain Res **63**, 331–340 (1986)

Chapter 11
Neural Network Modeling of Voluntary Single-Joint Movement Organization II. Parkinson's Disease

Vassilis Cutsuridis

Abstract The organization of voluntary movement is disrupted in Parkinson's disease. The neural network models of voluntary movement preparation and execution presented in the previous chapter are extended here by studying the effects of dopamine depletion in the output of the basal ganglia and in key neuronal types in the cortex and spinal cord. The resulting extended DA–VITE–FLETE model offers an integrative perspective on corticospinal control of Parkinsonian voluntary movement. The model accounts for most of the known empirical signatures of Parkinsonian willful action.

11.1 Introduction

Parkinson's disease (PD) is a disabling motor disorder that affects all kinds of movements. In the early stages of PD, patients have difficulty with walking, speaking, or getting in and out of chairs [33]. As the disease progresses, all movements are affected resulting at the end of the disease a complete inability to move. Patients require intense concentration to overcome the apparent inertia of the limbs that exists even for the simplest motor tasks. Movement initiation is particularly impaired when novel movements or sequences of movements are attempted [16, 3, 41].

The lack of understanding of the causes of PD and the problems associated with its treatment have led to the search for appropriate animal models. Over the years, two experimental methods have been employed to induce Parkinsonism in animals: (1) application of reserpine, alpha-methyl-p-tyrosine (AMPT) [14], and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) [12] resulting in dopamine depletion in the brain and (2) stereotaxic lesions with focal injections of 6-hydroxydopamine

V. Cutsuridis
Centre for Memory and Brain, Boston University, Boston, MA, USA,
e-mail: vcut@bu.edu

(6-OHDA) into the substantia nigra or medial forebrain bundle to destroy the ascending dopamine tracts. Depending on the method used, the effects vary and can be temporary or permanent.

MPTP administration in primates produces three distinct phases of altered motor activity: acute phase, subacute phase, and chronic phase [12]. In the acute phase after administration, animals appear to go to sleep and fall slowly from their perches to the floor of the cage; their eyes remain open, but with a vacant gaze [12]. Occasionally, wild running or exaggerated startle response events are observed [12]. The acute effects of MPTP last approximately 0.5–1.0 h and then disappear until subsequent administration [12].

During the subacute phase after MPTP administration, persistent motor deficits develop. Animals become increasingly akinetic and show rigidity of the limbs, freezing episodes, postural abnormalities, and loss of vocalization and blink reflex [12]. Compulsive climbing behavior can also occur at this stage, causing animals to damage their heads and faces [12]. This spectrum of behavioral effects lasts for some weeks, but the animals slowly recover. In subsequent weeks the motor deficits stabilize, and the animals enter the chronic phase of MPTP action. They show less spontaneous movements, although when challenged, they can move freely in the home cage [12]. Complex movements are poorly coordinated and clumsily executed. Hesitation prior to movement is apparent, and the range of movements observed appears limited [12].

Postmortem studies of PD in humans [22] and MPTP-treated rhesus monkeys [44] have shown that the toxin destroys the cells of the substantia nigra pars compacta, but not of the locus coereleus, dorsal raphe, and substantia nigra innominata [12]. Within the substantia nigra, the cells in the centrolateral area of the SNc are damaged more extensively than those in the medial portion of the SNc [32]. Administration of MPTP to young primates causes a profound ($> 90\%$) persistent loss of caudate-putamen dopamine content that is irreversible by any form of medication. The ventral tegmental area (VTA) adjacent to the substantia nigra shows limited and variable damage to tyrosine hydroxylase-containing cells in MPTP-induced Parkinsonism [12].

11.2 Brain Anatomy in Parkinson's Disease

The difficulty in understanding and treating Parkinson's disease is because there are multiple brain areas and pathways affected from the sites of neuronal degeneration all the way to the muscles. Figure 11.1 depicts three of these pathways: (1) the pathway from the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) to the striatum and from there to the thalamic nuclei and the frontal cortex through the substantia nigra pars reticulata (SNr) and the globus pallidus internal segment (GPi), (2) the pathway from the SNc and the VTA to the striatum and from there to the brainstem through the SNr and GPi, and (3) the pathway from
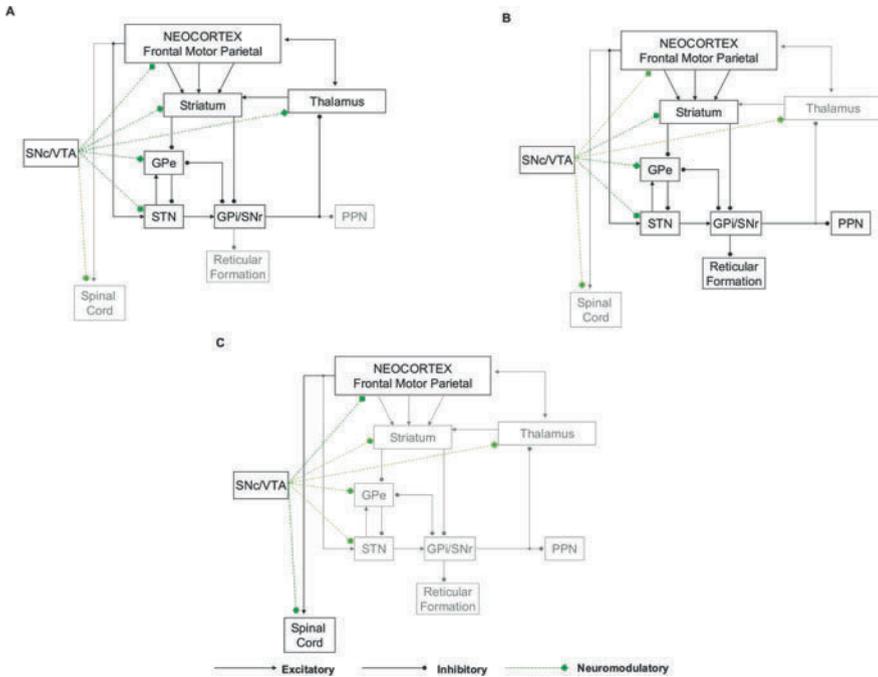
Fig. 11.1: Brain anatomical pathways in Parkinson's disease. (**A**) Pathways from the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) to the striatum and from there to the thalamic nuclei and the frontal cortex through the substantia nigra pars reticulata (SNr) and the globus pallidus internal segment (GPi). (**B**) Pathway from the SNc and the VTA to the striatum and from there to the brainstem through the SNr and GPi. (**C**) Pathway from the SNc/VTA to cortical areas such as the supplementary motor area (SMA), the parietal cortex, and the primary motor cortex (M1), and from there to the spinal cord.

the SNc/VTA to cortical areas such as the supplementary motor area (SMA), the parietal cortex, and the primary motor cortex (M1), and from there to the spinal cord.

The most popular view is that cortical motor centers are inadequately activated by excitatory circuits passing through the basal ganglia (BG) [1]. As a result, inadequate facilitation is provided to the otherwise normally functioning motor cortical and spinal cord neuronal pools and hence movements are small and weak [1]. Recently, a new view has been introduced by the modeling studies of Cutsuridis and Perantonis [21] and Cutsuridis [18, 19, 20]. According to this view, the observed delayed movement initiation and execution in PD is due to altered activity of motor cortical and spinal cord centers because of disruptions to their input from the basal ganglia structures and to their dopamine (DA) modulation. The main hypothesis is that depletion of DA modulation from the SNc disrupts, via several pathways,

the buildup of the pattern of movement-related responses in the primary motor and parietal cortex and results in a loss of directional specificity of reciprocal and bidirectional cells in the motor cortex as well as in a reduction in their activities and their rates of change. These changes result in delays in recruiting the appropriate level of muscle force sufficiently fast and in an inappropriate scaling of the dynamic muscle force to the movement parameters. A repetitive triphasic pattern of muscle activation is sometimes needed to complete the movement. All of these disruptions result in an increase of mean reaction time and a slowness of movement.

11.3 Empirical Signatures

The validity of the model's hypothesis is based on the existence of a widespread dopaminergic innervation in not only the basal ganglia, but also in cortex and spinal cord as well as on its effects on movement, muscular, and neuronal parameters of Parkinson's disease patients and MPTP-lesioned animals.

11.4 Is There Dopaminergic Innervation of the Cortex and Spinal Cord?

A widespread dopaminergic innervation from the substantia nigra (SN), the VTA, and the retrorubral area (RRA) to the cortex and spinal cord exists [6, 54]. A schematic diagram of the dopaminergic innervation of the neocortex is depicted in Fig. 11.2. DA afferents are densest in cortical areas 24, 4, 6, and SMA, where they display a trilaminar pattern of distribution, predominating in layers I, IIIa, and V–VI [5, 54, 25, 27, 28]. In the granular prefrontal (areas 46, 9, 10, 11, 12), parietal (areas 1, 2, 3, 5, 7), temporal (areas 21, 22), and posterior cingulate (area 23) cortices, DA afferents are less dense and show a bilaminar pattern of distribution in the depth of layers I and V–VI [5, 42, 43, 27, 28, 46]. Area 17 has the lowest DA density, where the DA afferents are mostly restricted to layer I [5].

In addition to the DAergic innervation of the neocortex, the presence of dopaminergic fibers in the dorsal and ventral horns of the spinal cord has been observed by several groups [7, 8]. In the dorsal horn, DA fibers are localized in the superficial layers and in the laminae III–V and X. In ventral horn, DA fibers are found in layers VII, VIII, and laminae IX [51]. The sources of the dorsal DAergic innervation are the posterior and dorsal hypothalamic areas and the periventricular gray matter of the caudal thalamus, whereas of the ventral dopaminergic innervation is the caudal hypothalamus A11 cell group [47]. Finally, an uncrossed nigrospinal DAergic pathway has been documented by anatomical methods [15].
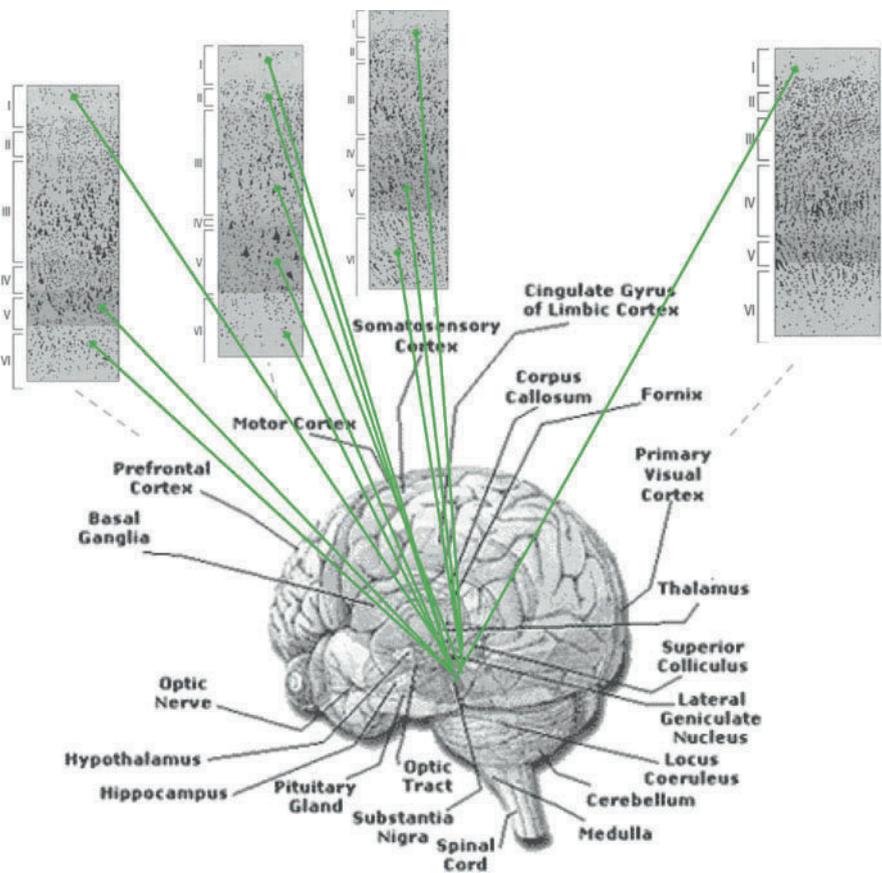
Fig. 11.2: Dopaminergic laminar innervation of the prefrontal, motor, somatosensory, and primary visual cortices from the substantia nigra pars compacta. *Diamond solid green lines*: dopamine projection from the substantia nigra.

11.5 Effects of Dopamine Depletion on Neuronal, Electromyographic, and Movement Parameters in PD Humans and MPTP Animals

The effects of dopamine depletion on neuronal, electromyographic, and movement parameters in PD humans and in MPTP-treated animals are briefly summarized below.

11.5.1 Cellular Disorganization in Cortex

Doudet and colleagues [23] trained monkeys to perform fast flexion and extension elbow movements while they recorded from their primary motor cortex before and

after 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine (MPTP) injection. A reduction in the number of reciprocally organized cells (neurons showing a reciprocal discharge pattern for flexion and extension movements; 49% in the pre-MPTP state and 18% in the post-MPTP state) and an increase in the number of unidirectional cells (cells whose activities change in only one direction; 19% in the pre-MPTP state and 50% in the post-MPTP state) without an alteration of the overall excitation were reported. It was suggested that there was a lift of inhibition from cells that are normally inhibited during movement resulting in an extra-imposed load on the limb [23].

11.5.2 Reduction of Neuronal Intensity and of Rate of Development of Neuronal Discharge in the Primary Motor Cortex

Watts and Mandir [50] examined the effects of MPTP-induced Parkinsonism on the primary motor cortex task-related neuronal activity and motor behavior of monkeys. Two monkeys were trained in the pre-MPTP state with the help of visual cues, delivered via a panel of light-emitting diodes (LEDs), to make fast, wrist flexion movements of 60. Once the animals were fully trained on the task and the M1 neuronal and EMG activities were recorded, intracarotid injection of MPTP was administered to induce a stable state of Parkinsonism. Single neuronal recordings were repeated during the experimentally induced Parkinsonian state for many months. They reported a decrease in the percentage of movement onset-related neurons and an increase in the latency between the start of M1 neuronal activity and the movement onset and in the duration of after-discharge following movement onset in the hemi-Parkinsonian state.

Similarly, Gross and colleagues [36] trained monkeys to perform a rapid elbow movement (> 30) of extension or flexion in response to an auditory signal. The unit activity of the primary motor cortical cells was recorded 500 ms before and 1,500 ms after the beginning of the auditory signal, before and after an electrolytic lesion of the substantia nigra pars compacta (SNc). They reported that the maximum discharge frequency in lesioned animals was lower than in normal animals.

Doudet and colleagues [23] observed a similar change in discharge rate of primary motor cortical cells as well as a prolongation of their total response duration. They reported that the time between the start of the alterations in the neuronal discharge and the onset of movement was increased by 20%.

11.5.3 Significant Increase in Mean Duration of Neuronal Discharge in Primary Motor Cortex Preceding and Following Onset of Movement

In the experimental paradigm described earlier, Gross and colleagues [36] observed that the latency between the onset of neuronal discharge and the beginning of fore-
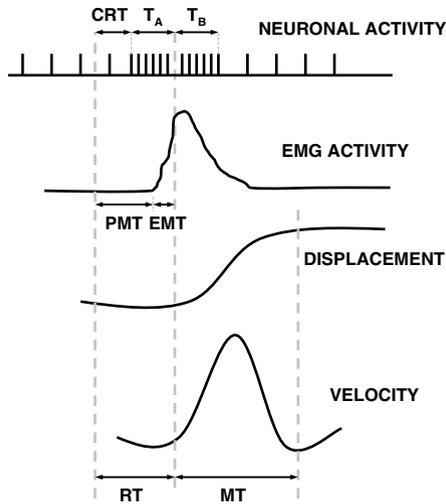
Fig. 11.3: Schematic representation of the neuronal, electromyographic, and kinematic variables. CRT: cellular reaction time; PMT: premotor time; EMT: electromechanical time; RT: reaction time; MT: movement times; T_A: time of neuronal discharge prior to movement onset; T_B: duration of neuronal discharge after movement onset.

arm displacement and the duration of the neuronal activity from the onset of movement and the time where the level of activity returned to resting levels (see Fig. 11.3 for a schematic description) were increased. Similarly, Doudet and colleagues [23] reported that the mean duration of neuronal discharge in area 4 preceding the onset of movement was slightly affected in the MPTP-treated animals, whereas the mean duration of neuronal discharge following the onset of movement was significantly increased.

11.5.4 Prolongation of Behavioral Simple Reaction Time

Benazzouz et al. [2] trained monkeys to perform a rapid elbow movement (> 40) of extension or flexion in response to an auditory signal. EMG activity was recorded with intramuscular electrodes 500 ms before and 1,500 ms after the beginning of the auditory signal, before and after an MPTP lesion of the substantia nigra pars compacta (SNc). They reported that the behavioral simple reaction time (cellular reaction time + mean duration of neuronal discharge before movement onset (TA; see Fig. 11.3) after a nigral MPTP lesion was significantly increased for both extension and flexion movements. Similarly, Doudet et al. [24, 23] and Gross et al. [36] observed a significant change in the mean values of the simple reaction time (RT)

for both flexion and extension movements in the MPTP-treated and electrolytical lesioned animals. Weiss et al. [52] investigated the kinematic organization of discrete elbow movements of different amplitudes to targets of various sizes of young, elderly, and PD subjects. The investigators reported a significant increase in the simple reaction time between young, elderly, and PD subjects over all conditions.

11.5.5 Repetitive Triphasic Pattern of Muscle Activation

Hallett and Khoshbin [37] asked healthy and Parkinson's disease (PD) patients to make rapid accurate elbow flexion movements of different angular distances (10, 20, and 40) while they recorded their EMG activities and their elbow angles with surface electrodes. They reported that healthy subjects exhibited a triphasic (agonist–antagonist–agonist) EMG pattern. However, the EMG patterns in the muscles of Parkinson's disease patients differed from those of the healthy subjects in that the bursts of EMG activity in the agonist muscle did not increase in magnitude for the larger amplitude movements. Hallett and Khoshbin [37] interpreted their results indicating that patients with PD are unable to sufficiently activate agonist muscles during movements made as quickly as possible. They showed that an apparent compensation for the decreased muscular activation was to evoke more cycles of activity to complete the movement. Doudet et al. [23] reported that in order for MPTP-treated animals to achieve the full amplitude of the required movement, additional successive bursts of lower amplitude and duration were needed.

11.5.6 Electromechanical Delay Time Is Increased

Electromechanical delay time (EMT; time between the onset of modification of agonist EMG activity and the onset of movement (OM); see Fig. 11.3) is significantly increased in MPTP-treated animals. Benazzouz et al. [2] study showed that monkeys display a significant increase in the EMD time. Doudet et al. [24, 23] in the exact same experimental paradigm observed a similar delay in EMT.

11.5.7 Depression of Rate of Development and Peak Amplitude of the First Agonist Burst of EMG Activity

Godaux and colleagues [34] conducted experiments with control and PD patients seated facing a target button. Subjects were instructed to switch off the target button when it lit, by pressing it as rapidly as possible. The activities of anterior deltoid, biceps brachii, triceps brachii, and extensor indicis muscles were recorded using sur-

face electrodes as the subjects were performing the task. Godaux et al. [34] found that the amplitudes of the peak EMG activity were reduced and the rates of development of muscle activity in both flexors and extensors were depressed in PD patients.

Corcos et al. [17] measured the maximum elbow flexor and extensor muscle strength in patients with PD during on and off anti-PD medication. Patients were tested in two maximally produced muscle isometric contractions and two flexion contractions equal to 50% of their maximal voluntary contractions. In all four conditions, the patients were seated on a table with fully supinated right forearm flexed 90° with respect to the arm and positioned vertically. The forearm was attached to a stiff steel bar and changes in torque were measured by strain gauges. EMG signals were recorded with surface electrodes. Corcos and colleagues reported a reduction in the peak torque and in the rate of change of torque.

Watts and Mandir [50] trained PD patients and age-matched controls to perform a rapid, wrist flexion task. Their hands were hidden from their view. Visual cues were used to instruct the subjects where and when to move. The subjects were advised to move as quickly and as accurately as possible once they were given the go-signal. Their flexor and extensor electromyographic (EMG) activities were recorded using surface electrodes during the trials. They noted decreased average amplitude of EMG activity for the patients with Parkinson's disease. Doudet et al. [24, 23] reported that the rate of development and peak amplitude of the first agonist burst of EMG activity were depressed. Similarly, Hallett and Khoshbin [37] observed, in patients with Parkinson's disease, there was a similar reduction in the activity of the first agonist burst as if it has reached a ceiling.

11.5.8 Movement Time Is Significantly Increased

Rand et al. [45] trained PD patients and age-matched controls to make rapid arm movements with or without accuracy constraints. Subjects were seated in front of a horizontal digitizer and held a stylus. The subject was required to move the stylus from a home position to a target position after an auditory signal. In the spatial accuracy condition, the subjects were required to move the stylus to the defined target and stop on it, whereas in the *n*-spatial accuracy condition, the subjects were asked to move toward the target without stopping precisely on it. The subjects were asked to make their movements as fast and as accurate as possible. Rand et al. [45] reported that the movements of patients were slower than those of the controls in both the acceleration phase and the deceleration phase. The prolonged deceleration phase for the patients was more pronounced in the target condition. In addition, the kinematics of PD patients were abnormal, characterized by a higher number of acceleration zero crossings indicating that their movements were segmented and that the first zero crossing occurred much earlier in the movement. Weiss et al. [52] trained and tested young, elderly, and PD subjects in making discrete elbow movements with varying amplitudes to targets of varying sizes. They reported that

both the acceleration and the deceleration times were increased. Doudet et al. [24, 23] and Benazzouz et al. [2] reported a 25–30% increase in movement duration in monkeys treated with MPTP compared to normal monkeys. Watts and Mandir [50] showed that both MPTP-treated animals and Parkinson's disease patients take longer time to complete the required movements.

11.5.9 Reduction of Peak Velocity

In the experimental study described earlier, one of Godaux et al. [34] findings was a profound decrease in the peak velocity of movement of PD patients. Camarata et al. [13] reported that in the MPTP-treated animals, the velocity profiles appeared less smooth and the amplitude of the velocity profile decreased and delayed in time at most distances and directions tested. Weiss et al. [52] observed a similar decrease in the peak velocity of movement of PD patients. Further, Benazzouz et al. [2] and Doudet et al. [24, 23] after treating monkeys with MPTP found a significant decrease in the amplitude of their velocity profiles. Rand et al. [45] reported a significant reduction of the peak velocity in both accuracy and no-accuracy movement conditions.

11.5.10 Reduction of Peak Force and Rate of Force Production

Stelmach et al. [48] examined the preparation and the production of isometric force in Parkinson's disease. PD patients, elderly, and young subjects were asked to generate a percentage of their maximum force levels. PD patients showed a similar progression of force variability and dispersion of peak forces to that of control subjects. Force production impairments were seen at the within-trial level. PD patients were substantially slower in initiating a force production and their peak forces were reduced.

11.5.11 Movement Variability

Camarata et al. [13] trained monkeys to make two-joint movements on a horizontal plane by moving a manipulandum in six different directions (30, 90, 150, 210, 270, and 330) at five distances from a central start box. Velocity and acceleration profiles were calculated for both pre- and post-MPTP states. They reported a marked variability in the onset, peak velocity, and time course of the velocity profile of MPTP-treated monkeys. Similarly, Stelmach et al. [48] reported variability in the force profile of Parkinson's disease patients.

11.6 The Extended VITE–FLETE Models with Dopamine

Figure 11.4 depicts the extended VITE–FLETE with dopamine model of voluntary movement preparation and execution in Parkinson's disease. In the previous chapter, the temporal development of the model without dopamine was discussed. The model under normal conditions successfully predicted the origin of the triphasic pattern of muscle activation and its neural substrates. In this chapter and although a much larger set of experimental data has been briefly described in the previous section, I will describe how the triphasic pattern and its neural and EMG substrates change when dopamine is depleted in basal ganglia, cortex, and spinal cord. Detailed descriptions of the model and its complete mathematical formalism
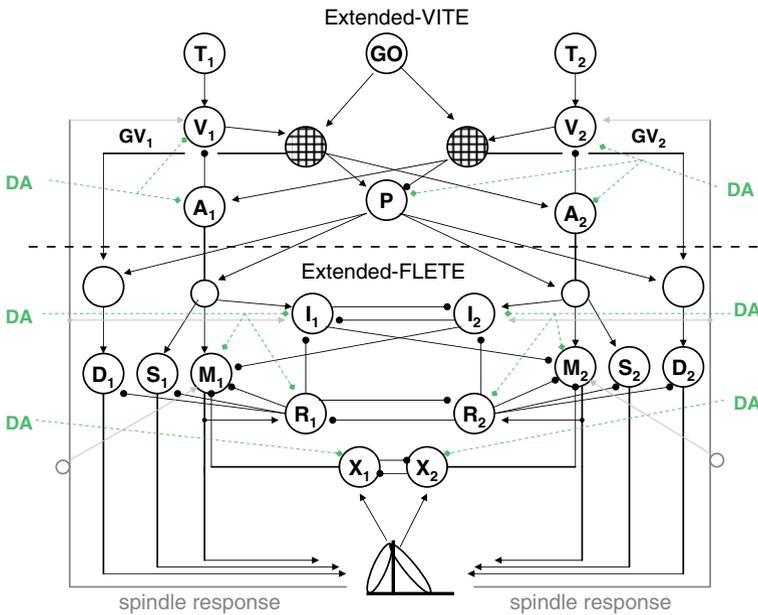
Fig. 11.4: Extended VITE–FLETE models with dopamine (DA). *Top*: DA–VITE model for variable-speed trajectory generation. *Bottom*: DA–FLETE model of the opponent processing spinomuscular system. *Arrow lines*: excitatory projections; *solid dot lines*: inhibitory projections; *diamond-dotted green lines*: dopamine modulation; *dotted arrow lines*: feedback pathways from sensors embedded in muscles. *GO*: basal ganglia output signal; *P*: bidirectional co-contractive signal; *T*: target position command; *V*: DV activity; *GV*: DVV activity; *A*: current position command; *M*: alpha motoneuronal (MN) activity; *R*: renshaw cell activity; *X, Y, Z*: spinal inhibitory interneuron (IN) activities; I_a: spinal type a inhibitory IN activity; *S*: static gamma MN activity; *D*: dynamic gamma MN activity; *1,2*: antagonist cell pair.

can be found in Cutsuridis and Perantonis [21] and Cutsuridis [18,19,20]. As in the extended model without dopamine (see previous chapter), the GO signal was defined by

$$G(t) = G_0(t - \tau_i)^2 u[t - \tau_i]/(\beta + \gamma(t - \tau_i)^2), \tag{11.1}$$

where G_0 amplifies the G_0 signal, i is the onset time of the ith volitional command, β and γ are free parameters, and $u[t]$ is a step function that jumps from 0 to 1 to initiate movement. The difference vector (DV) with dopamine was described by

$$\frac{dV_i}{dt} = 30(-V_i + T_i - DA_1 A_i + DA_1 a_w(W_i(t - \tau) - W_j(t - \tau))), \tag{11.2}$$

where T_i is the target position command, A_i is the current limb position command, a_w is the gain of the spindle feedback, $W_{i,j}$ are the spindle feedback signals from the antagonist muscles, and DA_1 is the modulatory effect of dopamine on area 4's PPV inputs to DV cell activity. Dopamine's values ranged from 0 (lesioned) to 1 (normal). The desired velocity vector (DVV) with dopamine which represented area's 4 reciprocally activated cell activity was defined by

$$u_i = \left[G(DA_2 V_i - DA_3 V_j + \frac{B_u}{DA_4} \right]^+, \tag{11.3}$$

where i, j designate opponent neural commands, B_u is the baseline activity of the phasic-MT area 4 cell activity, and DA_2, DA_3 are the modulatory effects of dopamine on DV inputs to DVV cell activity and DA_4 is the effect of dopamine on DVV baseline activity. The reader can notice that parameter DA_1 modulates the PPV input to area's 5 phasic (DV) cell activity (Equation 11.2), whereas parameters DA_2, DA_3, and DA_4 modulate the DV inputs to DVV and P cell activity (area's 4 reciprocal and bidirectional activities) and to DVV baseline activity (Equations 11.3 and 11.4), respectively. This is, as we explained in a previous section, because DA afferents are densest in area 4 than they are in area 5. So, the effect of DA depletion would be stronger in area 4 than in area 5. Also, the DV flexion (V_i) cell is modulated by a different DA parameter DA_2) from the DV extension (V_j) cell (DA_3). The latter is supported by the experimental findings of Doudet and colleagues [23] (for comparison see Figs. 11.4 and 11.5, where the firing intensity of the flexion cells is affected (reduced) more than the firing intensity of the extension cells). The co-contractive vector (P) with dopamine which was represented by area's 4 bidirectional neuronal activity was given by

$$u_i = \left[G(DA_2 V_i - DA_3 V_j + \frac{B_P}{DA_4} \right]^+, \tag{11.4}$$

whereas the present position vector (PPV) dynamics was defined by

$$\frac{dA_i}{dt} = G[DA_2.V_i]^+ - G[DA_3.V_j]^+. \tag{11.5}$$

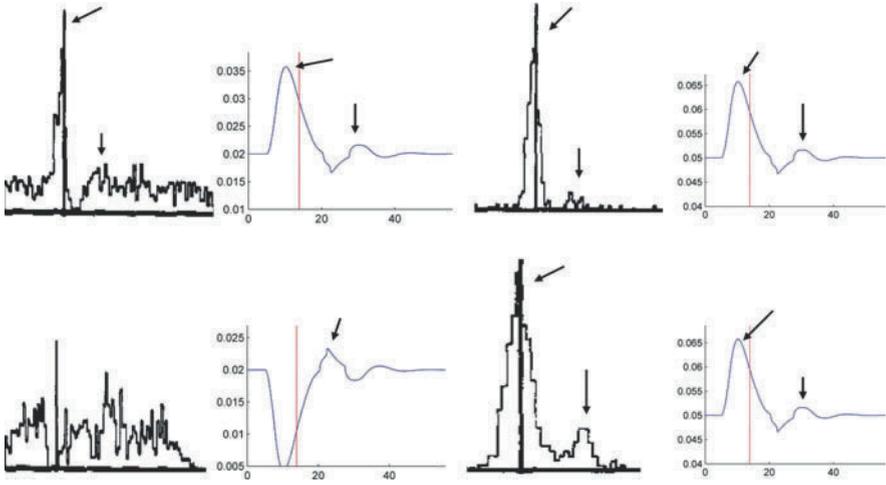The renshaw population cell activity with dopamine was modeled by

Fig. 11.5: Comparison of peristimulus time histograms (PSTH) of reciprocally organized neurons (column 1; reproduced with permission from Doudet et al. [23, Fig. 4A, p. 182], Copyright Springer-Verlag) in area 4, simulated area's 4 reciprocally organized phasic (DVV) cell activities (column 2), PSTH of area's 4 bidirectional neurons (column 3; reproduced with permission from [23, Fig. 4A, p. 182], Copyright Springer-Verlag) and simulated area's 4 co-contractive (P) cells activities (column 4) for a flexion (row 1) and extension (row 2) movements in normal monkey. The *vertical bars* indicate the onset of movement. Note a clear triphasic AG1-ANT1-AG2 pattern marked with *arrows* is evident in PSTH of reciprocally and bidirectionally organized neurons. The same triphasic pattern is evident in simulated DVV cell activities. The second peak in simulated activities marked with an arrow arises from the spindle feedback input to area's 5 DV activity.

$$\frac{dR_i}{dt} = \phi(\lambda B_i - R_i)\mathrm{DA}_5 z_i \max(M_i, 0) - \mathrm{DA}_6 R_i(1.5 + \max(R_j, 0)), \qquad (11.6)$$

whereas the $\alpha - \mathrm{MN}$ population activity with dopamine was described by

$$\frac{M_i}{dt} = \phi \ (\lambda B_i - M_i)\mathrm{DA}_7(A_i + P + \chi E_i) - (M_i + 2)\mathrm{DA}_8(1 + \Omega \max(R_i, 0)$$
$$+ \rho \max(X_i, 0) + \max(I_j, 0)) \qquad (11.7)$$

where X_i is the type I_b interneuron ($I_b\mathrm{IN}$) force feedback, E_i is the stretch feedback, and I_j is the type I_a interneuron. The type I_a interneuron ($I_a\mathrm{IN}$) population activity with dopamine was defined as

$$\frac{dI_i}{dt} = \phi \ (15 - I_i)\mathrm{DA}_9(A_i + P + \chi E_i) - \mathrm{DA}_{10} I_i(1 + \Omega \max(R_i, 0)$$
$$+ \max(I_j, 0)). \qquad (11.8)$$

The I_bIN population activity with dopamine was given by

$$\frac{dX_i}{dt} = \phi A_{11}(15 - X_i)F_i - X_i DA_{11}(0.8 + 2.2 \max(X_j, 0)), \qquad (11.9)$$

where F_i is the feedback activity of force-sensitive Golgi tendon organs.

11.7 Simulated Effects of Dopamine Depletion on the Cortical Neural Activities

Figures 11.5 and 11.6 show qualitative comparisons of experimental and simulated neuronal discharges of reciprocal and bidirectional neurons in normal and dopamine-depleted conditions, respectively. It is clearly evident an overall reduction of firing intensity [23, 36], a reduced rate of change of neuronal discharge [23, 36], a disorganization of neuronal activity (neuronal direction specificity is markedly reduced) [23], and an increase in baseline activity (in the normal case the baseline activity was 0.05, whereas in dopamine depleted the baseline activity increased to 0.07) [23]. Figure 11.8 shows a qualitative comparison of abnormal cellular responses of GPi neurons to striatal stimulation in MPTP-treated monkeys (column 1 of Fig. 11.8) and simulated GPi neuronal responses (column 2 of Figure 11.8).
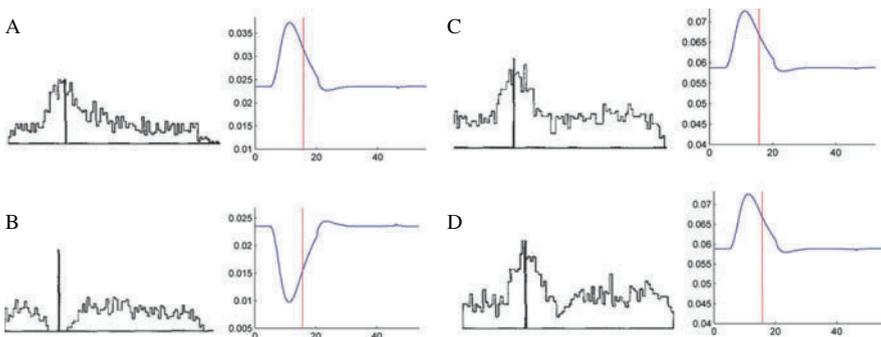
Fig. 11.6: Comparison of peristimulus time histograms (PSTH) of reciprocally organized neurons (column 1; reproduced with permission from [23, Fig. 4A, p. 182], Copyright Springer-Verlag) in area 4, simulated area's 4 reciprocally organized phasic (DVV) cell activities (column 2), PSTH of area's 4 bidirectional neurons (column 3; reproduced with permission from [23], Fig. 4A, p. 182, Copyright Springer-Verlag) and simulated area's 4 co-contractive (P) cells activities (column 4) for a flexion (**A** and **C**) and extension (**B** and **D**) movements in MPTP-treated monkey. The *vertical bars* indicate the onset of movement. Note that the triphasic pattern is disrupted: Peak AG1 and AG2 bursts have decreased, and ANT pause is shortened.

In their study, Tremblay and colleagues [49] observed an abnormal oscillatory GPi response, but failed to offer a functional role for it oscillatory responses. We propose that such GPi oscillatory responses (repetitive GO signal), comprising of at least two inhibitory–excitatory sequences, gate (multiply) the DV signal and generate repetitive volitional motor commands (DVV signals; not shown), which in turn generate repetitive agonist–antagonist muscle bursts (see row 2, column 3 of Figure 11.8) needed sometimes by PD patients to complete the full amplitude of the movement.

11.8 Simulated Effects of Dopamine Depletion on EMG Activities

As mentioned in the previous chapter, single ballistic movements at a joint in normal individuals are made with a single biphasic (sometimes triphasic) pattern of EMG activity in agonist and antagonist muscles [39, 4, 9, 10, 11, 35, 29, 30, 31, 53, 38]. In PD patients, the size of the first agonist burst is reduced. Up to a certain size, movements might be performed by a single agonist-antagonist pattern of muscle activation [26], but there are times that movements would require additional bursts of EMG activity [37, 2, 23] in order for the limb to reach the target. The extended DA–VITE–FLETE model has offered a plausible hypothesis of why PD EMG agonist burst activity is reduced and why sometimes multiple bursts of AG-ANT-
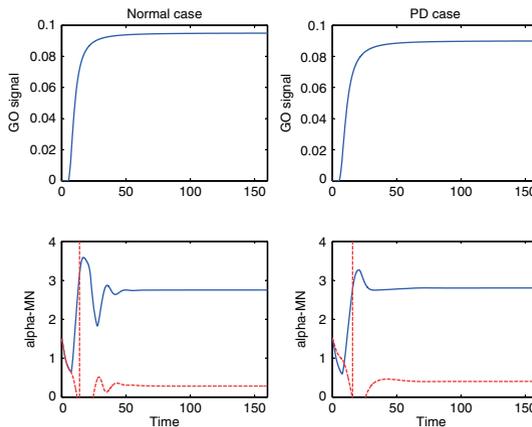
Fig. 11.7: Comparison of simulated GO signals (row 1) and α-MN activities (row 2) in normal (column 1) and dopamine-depleted (column 2) conditions. (Row 2) *Blue solid curve*: agonist α-MN activity; *Red-dashed curve*: antagonist α-MN activity. Note in PD case, the triphasic pattern is disrupted and it is replaced by a biphasic pattern of muscle activation. Also, the peaks of agonist and antagonist bursts are decreased.

AG are needed to complete the movement. According to the model, disruptions of the GO signal and dopamine depletion in the cortex and spinal cord disrupt the reciprocal organization of M1 neurons, reduce their activity, increase their rate of change, and hence result in the downscaling of the size of the first agonist burst and in the increase of its rate of change. So, in order for the subject to complete the movement and reach the target, additional EMG bursts are required. Figure 11.7 shows a qualitative comparison of the normal (column 1) and dopamine-depleted (column 2) simulated alpha motoneuronal (MN) activities of the agonist and antagonist muscles. A significant reduction in the peak agonist and antagonist amplitude as well as of their rate of development is evident [50, 23, 37, 17]. In contrast to some PD studies [2, 24, 40], a single and non co-contractive agonist-antagonist pattern of muscle activation is observed (column 2 of figure 11.7). Figure 11.8 shows a qualitative comparison of the experimental (column 1) and simulated (column 2) GPi discharge patterns (GO signal) and α-MN activity (column 3) in normal (row 1) and PD (row 2) large amplitude movement conditions. An abnormal oscillatory GO signal and DA depletion in the cortex and spinal cord result in a repetitive biphasic pattern of muscle activation (indicated by the arrows) necessary to complete the movement [37]. In the model, the generation of such repetitive biphasic pattern of
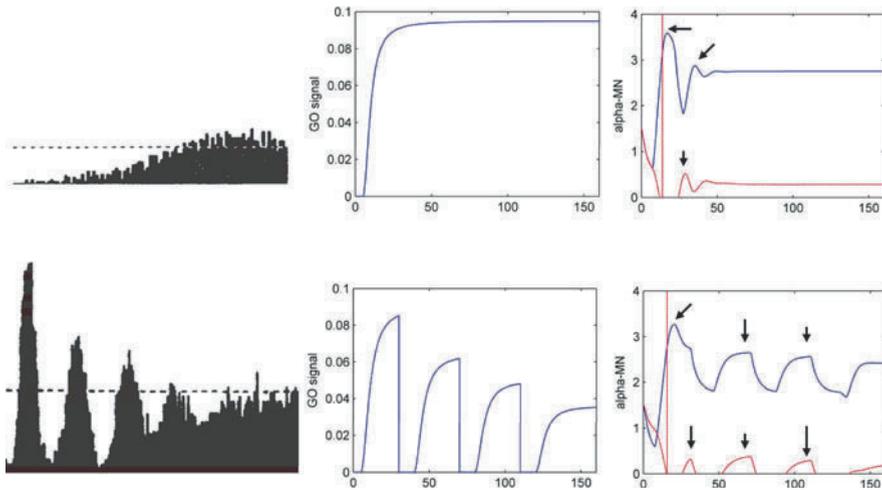
Fig. 11.8: Comparison of the experimental GPi PSTH (column 1), GO signals (column 2), and α-MN activities (column 3) in normal (row 1) and dopamine-depleted (row 2) conditions. (*Column 3, rows 1 and 2*) *Blue-colored solid curve*: agonist α-MN unit; *Red-colored dashed curve*: antagonist α-MN unit. Note in dopamine-depleted case the α-MN activity is disrupted and replaced by repetitive and co-contractive agonist–antagonist bursts (*row 2, column 3*). (*Column 1, row 1*) GPi PSTH in intact monkey reproduced with permission from Tremblay et al. [49, Fig. 4, p. 6], Copyright Elsevier. (*Column1, row 2*) GPi PSTH in MPTP monkey reproduced with permission from [49, Fig. 2, p. 23], Copyright Elsevier.

muscle activation is the result of the gating of the DV signal by multiple inhibition–excitation sequences of abnormal GO signal for the generation of multiple volitional motor cortical commands sent down to the spinal cord for the completion of the movement.

11.9 Conclusion

This chapter has focused on how the smooth organization of voluntary movement observed in normal individuals is disrupted in Parkinson's disease. The neural network model of voluntary movement preparation and execution presented in the previous chapter was extended by studying the effects of dopamine depletion in the output of the basal ganglia and in key neuronal types in cortex and spinal cord. The resulting extended DA–VITE–FLETE model offered an integrative perspective on corticospinal control of Parkinsonian voluntary movement. The model accounted for some of the known empirical signatures of Parkinsonian willful action:

- Cellular disorganization in cortex
- Increases in neuronal baseline activity
- Reduction of firing intensity and firing rate of cells in primary motor cortex
- Abnormal oscillatory GPi response
- Disinhibition of reciprocally tuned cells
- Repetitive bursts of muscle activation
- Reduction in the size and rate of development of the first agonist burst of EMG activity
- Repetitive triphasic pattern of muscle activation
- Non co-contraction of antagonist MN units in small amplitude movements
- Co-contraction of antagonist MN units in large amplitude movements

The interested reader should refer to the modeling studies of Cutsuridis and Perantonis [21] and Cutsuridis [18, 19, 20], where additional empirical signatures of PD kinematics have been successfully simulated:

- Increased duration of neuronal discharge in area 4 preceding and following onset of movement
- Prolongation of premotor and electromechanical delay times
- Asymmetric increase in the time-to-peak and deceleration time
- Decrease in the peak value of the velocity trace
- Increase in movement duration
- Movement variability

All these results provided sufficient evidence to support the main hypothesis of the model, which stated that "elimination of DA modulation from the SNc disrupts, via several pathways, the buildup of the pattern of movement-related responses in the primary motor and parietal cortex, and results in a loss of directional specificity of reciprocal and bidirectional cells in the motor cortex as well as in a reduction in their activities and their rates of change. These changes result in delays in recruiting

the appropriate level of muscle force sufficiently fast and in an inappropriate scaling of the dynamic muscle force to the movement parameters. A repetitive triphasic pattern of muscle activation is sometimes needed to complete the movement. All of these result in an increase of mean reaction time and a slowness of movement" [21].

References

1. Albin, R., Young, A., Penney, J. The functional anatomy of basal ganglia disorders. Trends Neurosci **12**, 366–375 (1989)
2. Benazzouz, A., Gross, C., Dupont, J., Bioulac, B. MPTP induced hemiparkinsonism in monkeys: Behavioral, mechanographic, electromyographic and immunohistochemical studies. Exp Brain Res **90**, 116–120 (1992)
3. Benecke, R., Rothwell, J., Dick, J. Performance of simultaneous movements in patients with Parkinson's disease. Brain **109**, 739–757 (1986)
4. Berardelli, A., Dick, J., Rothwell, J., Day, B., Marsden, C. Scaling of the size of the first agonist EMG burst during rapid wrist movements in patients with Parkinson's disease. J Neurol Neurosurg Psych **49**, 1273–1279 (1986)
5. Berger, B., Trottier, S., Verney, C., Gaspar, P., Alvarez, C. Regional and laminar distribution of dopamine and serotonin innervation in the macaque cerebral cortex: A radioautographic study. J Comp Neurol **273**, 99–119 (1988)
6. Bjorklund, A., Lindvall, O. Dopamine containing systems in the CNS. Classical Transmitters in the CNS: Part 1, Handbook of Chemical Neuroanatomy, Vol. 2, pp. 55–121. Elsevier, Amsterdam (1984)
7. Bjorklund, A., Skagerberg, G. Evidence of a major spinal cord projection from the diencephalic A11 dopamine cell group in the rat using transmitter-specific fluoroscence retrograde tracing. Brain Res **177**, 170–175 (1979)
8. Blessing, W., Chalmers, J. Direct projection of catecholamine (presumably dopamine)-containing neurons from the hypothalamus to spinal cord. Neurosci Lett **11**, 35–40 (1979)
9. Brown, S., Cooke, J. Initial agonist burst duration depends on movement amplitude. Exp Brain Res **55**, 523–527 (1984)
10. Brown, S., Cooke, J. Movement related phasic muscle activation I. Relations with temporal profile of movement. J Neurophys **63**(3), 455–464 (1990)
11. Brown, S., Cooke, J. Movement related phasicmuscle activation II. Generation and functional role of the triphasic pattern. J Neurophysiol **63**(3), 465–472 (1990)
12. Burns, R., Chiueh, C., Markey, S., Ebert, M., Jacobowitz, D., Kopin, I. A primate model of parkinsonism: Selective destruction of dopaminergic neurons in the pars compacta of the substantia nigra by n-methyl-4-phenyl-1,2,3,6-tetrahydropyridine. Proc Natl Acad Sci USA **80**, 4546–4550 (1983)
13. Camarata, P., Parker, R., Park, S., Haines, S., Turner, D., Chae, H., Ebner, T. Effects of MPTP induced hemiparkinsonism on the kinematics of a twodimensional, multi-joint arm movement in the rhesus monkey. Neuroscience **48**(3), 607–619 (1992)
14. Carlsson, A., Lindquist, M., Magnusson, T. 3,4-dihydroxyphenylalanine and 5-hydroxytryptophan as reserpine antagonists. Nature **180**, 1200 (1957)
15. Commissiong, J., Gentleman, S., Neff, N. Spinal cord dopaminergic neurons: Evidence for an uncrossed nigrostriatal pathway. Neuropharmacology **18**, 565–568 (1979)
16. Connor, N., Abbs, J. Task-dependent variations in parkinsonianmotor impairments. Brain **114**, 321–332 (1991)
17. Corcos, D., Jaric, S., Gottlieb, G. Electromyographic analysis of performance enhancement. Advances in Motor Learning and Control. Human Kinetics, Champaign, IL (1996)
18. Cutsuridis, V. Neural model of dopaminergic control of arm movements in Parkinson's disease Bradykinesia. Artificial Neural Networks, *LNCS*, Vol. 4131, pp. 583–591. Springer-Verlag, Berlin (2006)

19. Cutsuridis, V. Biologically inspired neural architectures of voluntary movement in normal and disordered states of the brain. Ph.D. Thesis (2006). Unpublished Ph.D. dissertation. http://www.cs.stir.ac.uk/ vcu/papers/PhD.pdf

20. Cutsuridis, V. Does reduced spinal reciprocal inhibition lead to co-contraction of antagonist motor units? a modeling study. Int J Neural Syst **17**(4), 319–327 (2007)

21. Cutsuridis, V., Perantonis, S. A neural model of Parkinson's disease bradykinesia. Neural Netw **19**(4), 354–374 (2006)

22. Davis, G., Williams, A., Markey, S., Ebert, M., Calne, E., Reichert, C., Kopin, I. Chronic parkinsonism secondary to intravenous injection of meperidine analogues. Psychiatr Res **1**, 249–254 (1979)

23. Doudet, D., Gross, C., Arluison, M., Bioulac, B. Modifications of precentral cortex discharge and EMG activity in monkeys with MPTP induced lesions of DA nigral lesions. Exp Brain Res **80**, 177–188 (1990)

24. Doudet, D., Gross, C., Lebrun-Grandie, P., Bioulac, B. MPTP primate model of Parkinson's disease: A mechanographic and electromyographic study. Brain Res **335**, 194–199 (1985)

25. Elsworth, J., Deutch, A., Redmond, D., Sladek, J., Roth, R. MPTP reduces dopamine and norepinephrine concentrations in the supplementary motor area and cingulate cortex of the primate. Neurosci Lett **114**, 316–322 (1990)

26. Flowers, K. Visual "closed-loop" and "open-loop" characteristics of voluntary movement in patients with parkinsonism and intention tremor. Brain **99**(2), 269–310 (1976)

27. Gaspar, P., Duyckaerts, C., Alvarez, C., Javoy-Agid, F., Berger, B. Alterations of dopaminergic and noradrenergic innervations in motor cortex in Parkinson's disease. Ann Neurol **30**, 365–374 (1991)

28. Gaspar, P., Stepniewska, I., Kaas, J. Topography and collateralization of the dopaminergic projections to motor and lateral prefrontal cortex in owl monkeys. J Comp Neurol **325**, 1–21 (1992)

29. Ghez, C., Gordon, J. Trajectory control in targeted force impulses. I. Role in opposing muscles. Exp Brain Res **67**, 225–240 (1987)

30. Ghez, C., Gordon, J. Trajectory control in targeted force impulses. II. Pulse height control. Exp Brain Res **67**, 241–252 (1987)

31. Ghez, C., Gordon, J. Trajectory control in targeted force impulses. III. Compensatory adjustments for initial errors. Exp Brain Res **67**, 253–269 (1987)

32. Gibb, W., Lees, A., Jenner, P., Marsden, C. MPTP: Effects of MPTP in the mid-brain of marmoset. A Neurotoxin Producing A Parkinsonian Syndrome, pp. 607–614. Academic Press, New York (1986)

33. Gibberd, F. The management of Parkinson's disease. Practitioner **230**, 139–146 (1986)

34. Godaux, E., Koulischer, D., Jacquy, J. Parkinsonian bradykinesia is due to depression in the rate of rise of muscle activity. Ann Neurol **31**(1), 93–100 (1992)

35. Gottlieb, G., Latash, M., Corcos, D., Liubinskas, A., Agarwal, G. Organizing principle for single joint movements: I. agonist-antagonist interactions. J Neurophys **13**(6), 1417–1427 (1992)

36. Gross, C., Feger, J., Seal, J., Haramburu, P., Bioulac, B. Neuronal activity of area 4 and movement parameters recorded in trained monkeys after unilateral lesion of the substantia nigra. Exp Brain Res Suppl. **7**, 181–193 (1983)

37. Hallett, M., Khoshbin, S. A physiological mechanism of bradykinesia. Brain **103**, 301–314 (1980)

38. Hallett, M., Marsden, G. Ballistic flexion movements of the human thumb. J Physiol **294**, 33–50 (1979)

39. Hallett, M., Shahani, B., Young, R. EMG analysis of stereotyped voluntary movements. J Neurol Neurosurg Psychiatr **38**, 1154–62 (1975)

40. Hayashi, A., Kagamihara, Y., Nakajima, Y., Narabayashi, H., Okuma, Y., Tanaka, R. Disorder in reciprocal innervation upon initiation of voluntary movement in patients with Parkinson's disease. Exp Brain Res **70**, 437–440 (1988)

41. Lazarus, J., Stelmach, G. Inter-limb coordination in Parkinson's disease. Mov Disord 7, 159–170 (1992)

42. Lewis, D., Morrison, J., Goldstein, M. Brainstem dopaminergic neurons project to monkey parietal cortex. Neurosci Lett **86**, 11–16 (1988)
43. Lidow, M., Goldman-Rakic, P., Gallager, D., Geschwind, D., Rakic, P. Distribution of major neurotransmitter receptors in the motor and somatosensory cortex of the rhesus monkey. Neuroscience **32**(3), 609–627 (1989)
44. Pifl, C., Bertel, O., Schingnitz, G., Hornykiewitz, O. Extrastriatal dopamine in symptomatic and asymptotic rhesus monkey treated with 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). Neurochem Int **17**, 263–270 (1990)
45. Rand, M., Stelmach, G., Bloedel, J. Movement accuracy constraints in Parkinson's disease patients. Neuropsychologia **38**, 203–212 (2000)
46. Scatton, B., Javoy-Agid, F., Rouquier, L., Dubois, B., Agid, Y. Reduction of cortical dopamine, noradrenaline, serotonin and their metabolites in Parkinson's disease. Brain Res **275**, 321–328 (1983)
47. Shirouzou, M., Anraku, T., Iwashita, Y., Yoshida, M. A new dopaminergic terminal plexus in the ventral horn of the rat spinal cord. Immunohistochemical studies at the light and the electron microscopic levels. Experientia **46**, 201–204 (1990)
48. Stelmach, G., Teasdale, N., Phillips, J., Worringham, C. Force production characteristics in parkinson's disease. Exp Brain Res **76**, 165–172 (1989)
49. Tremblay, L., Filion, M., Bedard, P. Responses of pallidal neurons to striatal stimulation in monkeys with MPTP-induced parkinsonism. Brain Res **498**(1), 17–33 (1989)
50. Watts, R., Mandir, A. The role of motor cortex in the pathophysiology of voluntary movement deficits associated with parkinsonism. Neurol Clin **10**(2), 451–469 (1992)
51. Weil-Fugazza, J., Godefroy, F. Dorsal and ventral dopaminergic innervation of the spinal cord: Functional implications. Brain Res Bull **30**, 319–324 (1993)
52. Weiss, P., Stelmach, G., Adler, C., Waterman, C. Parkinsonian arm movements as altered by task difficulty. Parkinsonism Relat Disord **2**(4), 215–223 (1996)
53. Wierzbicka, M., Wiegner, A., Shahani, B. Role of agonist and antagonist muscles in fast arm movements. Exp Brain Res **63**, 331–340 (1986)
54. Williams, S., Goldman-Rakic, P. Widespread origin of the primate mesofrontal dopamine system. Cereb Cortex **8**, 321–345 (1998)

Chapter 12
Parametric Modeling Analysis of Optical Imaging Data on Neuronal Activities in the Brain

Shigeharu Kawai, Yositaka Oku, Yasumasa Okada, Fumikazu Miwakeichi, Makio Ishiguro, and Yoshiyasu Tamura

Abstract An optical imaging technique using a voltage-sensitive dye (voltage imaging) has been widely applied to the analyses of various brain functions. Because optical signals in voltage imaging are small and require several kinds of preprocessing, researchers who use voltage imaging often conduct signal averaging of multiple trials and correction of signals by cutting the noise near the baseline in order to improve the apparent signal–noise ratio. However, a noise cutting threshold level that is usually set arbitrarily largely affects the analyzed results. Therefore, we aimed to develop a new method to objectively evaluate optical imaging data on neuronal activities. We constructed a parametric model to analyze optical time series data. We have chosen the respiratory neuronal network in the brainstem as a representative system to test our method. In our parametric model we assumed an optical signal of each pixel as the input and the inspiratory motor nerve activity of the spinal cord as the output. The model consisted of a threshold function and a delay transfer function. Although it was a simple nonlinear dynamic model, it could provide precise

Shigeharu Kawai
The Graduate University for Advanced Studies, Minato-ku, Tokyo, Japan,
e-mail: `kawai@ism.ac.jp`

Yositaka Oku
Hyogo College of Medicine, Nishinomiya, Hyogo, Japan, e-mail: `yoku@hyo-med.ac.jp`

Yasumasa Okada
Keio University Tsukigase Rehabilitation Center, Izu, Shizuoka, Japan,
e-mail: `yasumasaokada@1979.jukuin.keio.ac.jp`

Fumikazu Miwakeichi
Chiba University, Inage-ku, Chiba, Japan,
e-mail: `miwake1@faculty.chiba-u.jp`

Makio Ishiguro
The Institute of Statistical Mathematics, Minato-ku, Tokyo, Japan,
e-mail: `ishiguro@ism.ac.jp`

Yoshiyasu Tamura
The Institute of Statistical Mathematics, Minato-ku, Tokyo, Japan, e-mail: `tamura@ism.ac.jp`

estimation of the respiratory motor output. By classifying each pixel into five types based on our model parameter values and the estimation error ratio, we obtained detailed classification of neuronal activities. The parametric modeling approach can be effectively employed for the evaluation of voltage-imaging data and thus for the analysis of the brain function.

12.1 Introduction

Electrical neuronal activity in the brain, on the scale of tissue or multiple cellular levels, has been investigated classically with a multielectrode technique. Although this technique enables us to analyze spatiotemporal profiles of neuronal activities as multiple spike trains [25, 16, 2], the spatial resolution is generally low (e.g., 20 recording points/$200\,mm^2$) due to the limited number and density of microelectrodes. Further, because a multielectrode technique requires insertion of multiple microelectrodes into brain tissue, it could cause mechanical tissue damage especially in the brain of small animals.

On the contrary, optical recording techniques do not have such drawbacks. Optical recording techniques were first reported in 1968 [3, 28], have been steadily improved, and have now become more popular than a multielectrode technique in the analysis of neuronal activity of the brain. Among various optical recording methods, a technique using a voltage-sensitive dye (voltage imaging) enables us to noninvasively analyze membrane potential changes of multiple neurons in a region of interest (ROI) [29, 17, 23, 11, 10, 18, 19, 22, 21]. Although the temporal resolution of voltage imaging is generally lower than that of a multielectrode technique, voltage imaging provides much higher spatial resolution than a multielectrode technique (e.g., 10,000 recording points/$10\,mm^2$).

Optical imaging data give us copious information especially in the spatial domain. However, the data obtained with this technique must be cautiously evaluated. This is because optical signals are small and thus usually require cycle triggered signal averaging (e.g., 50 times) and noise cutting near the baseline using an arbitrarily set threshold in order to improve the apparent signal-to-noise ratio. Further, optical signals can be affected by photobleaching and thus may need correction of the deviated baseline. Through such preprocessing, the timing of activity occurrence in different regions cannot be directly compared, as an example indicates in Fig. 12.1.

Several researchers have developed more sophisticated analytical methods, which have fled from such threshold problems. Fukunish and Murai [12] were pioneers of statistical analysis of voltage-imaging data. They analyzed the spatiotemporal patterns of neuronal activities and oscillatory neural activity transfer by applying a multivariable autoregressive model to the voltage-imaging signals of the guinea pig primary auditory cortex. Fisher et al. [9] conducted voltage-imaging experiments in the intra-arterially perfused in situ rat preparation and applied a correlation coefficient imaging technique to extract and classify respiratory related signals from optical images. They calculated the correlation for each pixel with a given correlation function; they used five different functions that approximated activities of
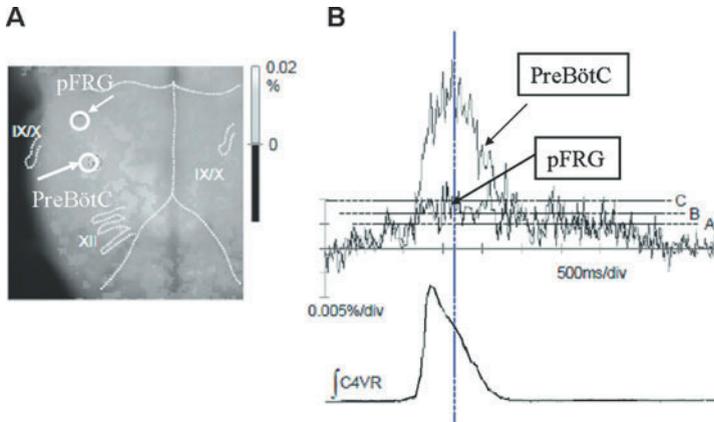
Fig. 12.1: Examples of cases with arbitrarily set threshold that may mislead the timing of the onset of optical signals. **a**: Respiratory related optical signals on the ventral surface of the medulla in a rat brainstem–spinal cord preparation. Region of interests (ROIs) were set in the pFRG area and the pre-BötC area. The anterior inferior cerebellar artery, the basilar artery, and the vertebral artery are demarcated with *white dotted lines*. IX/X, XII, cranial nerves. **b**: Integrated C4VR activity ($\int$C4VR) and optical signal waveforms in the pFRG area and the pre-BötC area, which correspond to the ROI on the photograph on panel A. The *vertical line* indicates the timing at which the voltage image was computed. *Horizontal dotted lines* represent different thresholds. If the threshold level is the *horizontal dotted line* a, then optical signals in both pFRG and pre-BötC areas appear simultaneously at the preinspiratory period. If the threshold level is b, then only optical signals in the pFRG area appear at the preinspiratory period. If the threshold level is c, then optical signals first appear in the pre-BötC area at the onset of inspiration, and subsequently signals appear in the pFRG area. Therefore, images could mislead the timing of the onset of optical signals.

basic types of respiratory neurons. Oku et al. [22] have developed a method to identify respiratory related pixel areas by calculating the cross-correlation between the forth cervical spinal cord (C4) ventral root (C4VR) inspiratory output activity and the optical time series data in each pixel in the neonatal rat brainstem–spinal cord preparation. In this method, by estimating the maximum correlation coefficient and the lag at which the maximum correlation coefficient is given, functional characteristics of the neurons in the two respiratory rhythm generators (RRGs) could be clearly discriminated. Recently, Yoshida et al. [30] applied an independent component analysis and correlation analysis to voltage-imaging data obtained from the guinea pig brain, and found that ongoing and spontaneous activities in the auditory cortex exhibit anisotropic spatial coherence extending along the isofrequency bands.

Although optical imaging with cycle triggered signal averaging has been widely used as explained above, the ability of such evaluation is limited within grasping

qualitative characteristics as power distribution ratios or correlation coefficients. So far, the quantitative analysis, e.g., estimation of respiratory output using optical imaging data, has not been reported. In the present study, we aimed to develop a new method to objectively and quantitatively evaluate voltage-imaging data. For this purpose, we intended to construct a parametric model to analyze optical time series data.

We have chosen the respiratory neuronal network in the brainstem as a representative system to test our method. Because the respiratory neuronal network in the brainstem consists of functionally and anatomically distinct neuronal groups and also forms motor nerve activity as the neural output, the respiratory neuronal network is ideal as a model system for our analysis. The essential current knowledge on the respiratory neuronal network in the brainstem is as follows. The respiratory rhythm and motor patterns are generated by neuronal aggregates that are distributed bilaterally in a columnar manner in the ventrolateral and dorsolateral reticular formation (for review see [24, 7, 6, 8]). In neonatal animals, two respiratory related ventrolateral medullary regions, the parafacial respiratory group (pFRG) [23] and the pre-Bötzinger complex (pre-BötC) [26], have been identified as putative RRGs. However, the detailed function and anatomy of these RRGs have not been clarified.

12.2 Methods

12.2.1 Recording of Optical Signals and Preprocessing

We recorded respiratory neuronal activities in the brainstem by voltage imaging in isolated brainstem–spinal cord preparations. For principles and general techniques of voltage imaging, refer to the reviews by Cohen and Salzberg [4], Kamino [15], Ebner and Chen [5], and Baker et al. [1]. Briefly, preparations were made of neonatal Sprague-Dawley rats ($n = 19$, 0–1 day old) under deep anesthesia as described elsewhere [27, 20, 18, 19, 22]. Experimental protocols were approved by the Animal Research Committee of Hyogo College of Medicine. Preparations were stained with a voltage-sensitive dye (di-2-ANEPEQ) [19, 22]. Inspiratory burst activity was monitored from C4VR using a glass suction electrode. Activity of respiratory neurons in the ventral medulla was analyzed using an optical recording system (MiCAM Ultima, BrainVision, Tokyo). Preparations were illuminated with a tungsten–halogen lamp (150 W) through a band-pass excitation filter ($\lambda = 480$–550 nm). Epifluorescence through a long-pass barrier filter ($\lambda > 590$ nm) was detected with a CMOS sensor array. Magnification of the microscope was adjusted to $2.8 \times 3.3 \times$ depending on the size of the brainstem. One pixel corresponded to $30 \times 30 - 35 \times 35 \mu m$, and the image sensor covered a total of $3 \times 3 - 3.5 \times 3.5$ mm^2. A total of 256 frames, 50 frames/s, were recorded starting at 1.28 s before the onset of C4VR activity.

As shown in Fig. 12.2b, c, raw optical signals had poor signal-to-noise ratios and respiratory related signals were not obvious. The correlation coefficient values
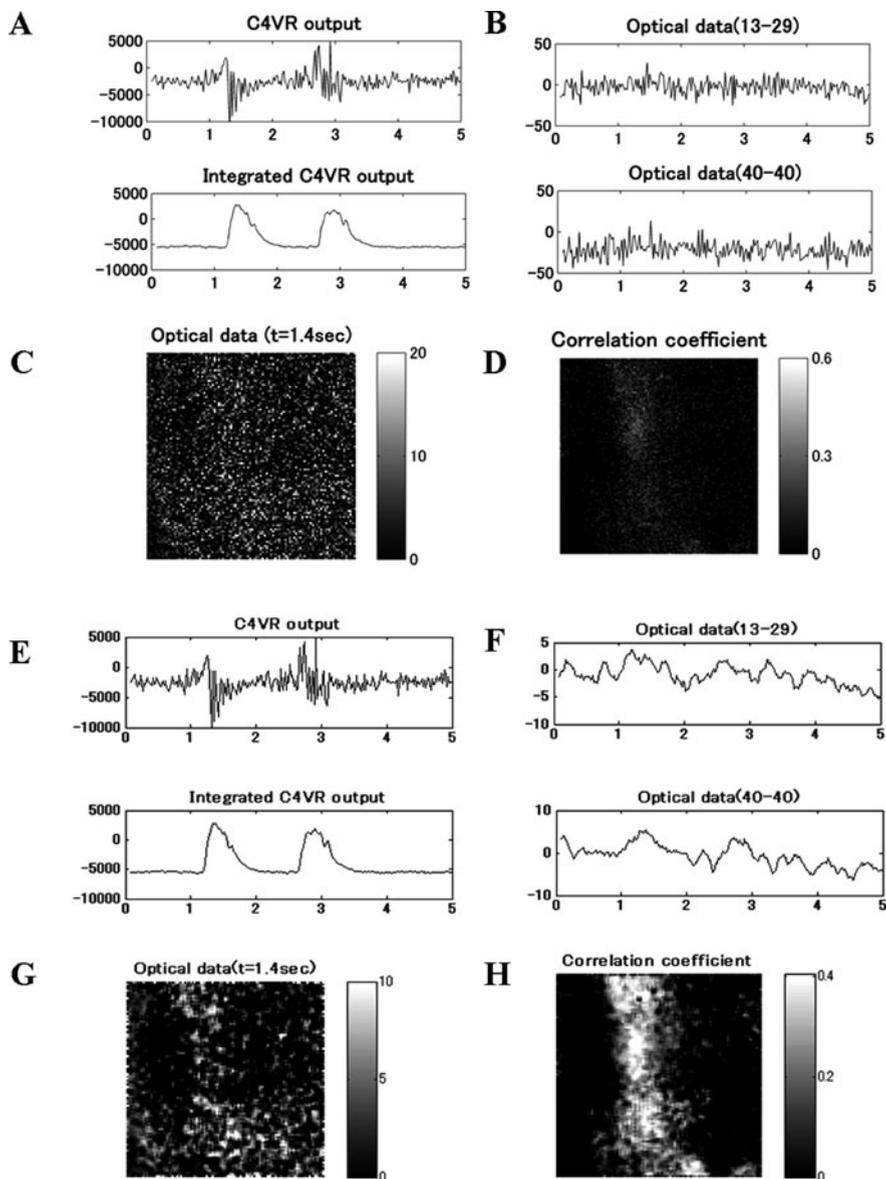
Fig. 12.2: Comparison between raw data and data after preprocessing. **a**: Time series of C4VR output raw data and integrated data of C4VR (raw data). **b**: Time series of imaging data in two pixels (raw data). **c**: Spatial distribution of imaging data value (raw data). **d**: Correlation coefficient (raw data). **e**: Time series of C4VR output raw data and integrated data of C4VR (data after preprocessing). **f**: Time series of imaging data in two pixels (data after preprocessing). **g**: Spatial distribution of imaging data value (data after preprocessing). **h**: Correlation coefficient (data after preprocessing).

between integrated C4VR output and each imaging data were plotted on the photo image of the ventral medullary surface (Fig. 12.2d), and there were few pixels whose values were over 0.4. In contrast, when signals were moving time averaged (bin width $= 7$) and spatially averaged by 3×3 pixels, respiratory related activities became visible (Fig. 12.2f, g). Many pixels whose correlation coefficient values were more than 0.5 were seen widely (Fig. 12.2h).

12.2.2 Modeling

Let us consider a model which estimates the respiratory motor output from respiratory related optical signals. It is natural to assume that the respiratory motor output is the sum of an estimation function of optical signals derived from pixels involved in respiratory neuronal activities:

$$y^*(t) = \sum_{i=1}^{N} g_i(x_i(t)).$$ (12.1)

Here, y^*, g_i, x_j, and N represent the estimated C4VR motor output, the estimation function, optical time series data in pixel i, and the number of respiratory related pixels, respectively. However, the number of respiratory related pixels is several hundreds, and it is not practical to deal with (12.1) because too many numbers of parameters must be determined. Instead, we consider a model where we estimate the respiratory motor output from optical signals derived from a specific set of respiratory related pixels. The number of pixels taken into account N_0 is very small as compared to N, and in the extreme case, N_0 can be 1.

$$y^*(t) = \sum_{j=1}^{N_0} f_j(x_j(t)).$$ (12.2)

In this case, it is essential to develop methods to determine the estimation function and to select the specific set of pixels. First, let us consider the case where N_0 is 1, i.e., a single-input single-output (SISO) model. The model must satisfy the following conditions:

(1) The respiratory motor output is not activated unless the optical signal within the pre-BöC exceeds a certain threshold.

(2) The pre-BötC region is activated earlier and deactivated later than the respiratory motor output.

To satisfy these conditions, we consider a nonlinear dynamic model consisting of a sigmoid function and a delayed first-order transfer function (STF model; sigmoid and transfer function model),

$$y^*(s) = \frac{Ke^{-Ls}}{1 + Ts} \times \frac{1}{1 + e^{-(x(s)-a)}},$$ (12.3)

where a, K, L, and T represent threshold, gain, dead time (delay) and time constant, respectively. The parameter values were determined so that the variance of estimation error was minimized. Given that the sampling interval is Δt, the dead time L is an integer multiple of Δt, i.e., where l is an integer. For (12.3) is rewritten in a discrete form:

$$y^*(n) = e^{\Delta t/T} y*(n-1) + (1 - e^{\Delta t/T}) \frac{K}{1 + e^{-(x(n-l-1)-a)}}. \tag{12.4}$$

For a given l, the variance of estimation error $\sigma(e)^2$ is expressed as

$$\sigma_l^2(e) = \frac{1}{N} \sum_{k=1}^{N} (y(n) - y^*(n))^2. \tag{12.5}$$

Then, $a_l C T_l$ and K_l that minimize $\sigma_l^2(e)$ are estimated by one of the nonlinear optimization method, the sequential quadratic programing method [14, 13]. Finding l^* that minimizes σ_l^2 gives the optimal set of parameter values for a, K, L, and T as a_l^Λ, K_l^Λ, T_l^Λ, $\Delta t * l^*$.

Next, we consider a multi-input single-output (MISO) model where the output is estimated by the weighted sum of STF model estimates applied to optical time series data at each pixel, expecting the improvement of the estimation.

$$y* = \sum_{i=1}^{I} w_i \frac{1}{1 + e^{-(x_i - a_i)}} \times \frac{K_i e^{-L_i s}}{1 + T_i s} x_i(s). \tag{12.6}$$

Here we estimate w_i, the weight coefficient, using the same nonlinear optimization method so that the estimation error variance is minimized.

12.2.3 Classification of Optical Signals Based on Activation Timing

We classified optical signals into five categories based on the timing of the onset of activation, the timing when the activation reached its peak, the timing when the activation subsided to the resting state, and the magnitude of variation (Fig. 12.3a). The timings were evaluated relative to the respiratory motor activity. Figure 12.3b exemplifies the five activation patterns with relation to the respiratory motor activity. Note that the respiratory motor activity and optical time series data were artificially composed in these examples. Type-1 pixels correspond to pixels within the pFRG, whereas Type-2 pixels correspond to those within the pre-BötC, which more directly contribute to the respiratory motor output. Type-3 and Type-4 are also respiratory related pixels, but are assumed to poorly contribute to the respiratory motor activity. Type-5 is pixels that show by chance behaviors similar to the respiratory motor activity.

A

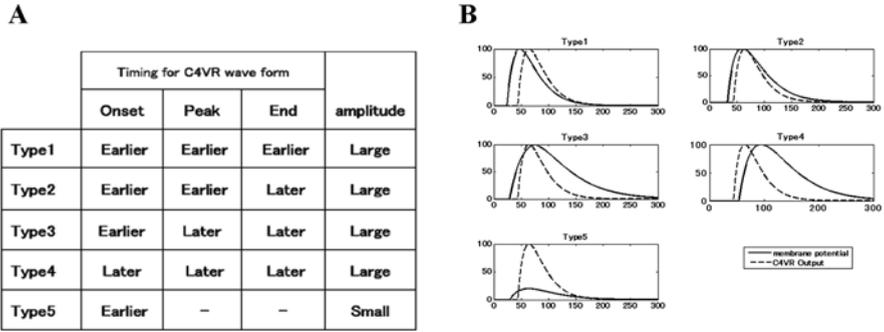|	Timing for C4VR wave form			
	Onset	**Peak**	**End**	**amplitude**
Type1	Earlier	Earlier	Earlier	Large
Type2	Earlier	Earlier	Later	Large
Type3	Earlier	Later	Later	Large
Type4	Later	Later	Later	Large
Type5	Earlier	–	–	Small

B

Fig. 12.3: Types of activity patterns for C4VR output. **a**: Characteristics of the respiratory related activities of imaging data. **b**: Artificial signals of imaging data for C4VR output.

A

	Error ratio (R)	Gain (K)	Dead Time (L)	Threshold Value (a)	Time Constant (T)
Type1	4.7	0.99	14.0	3.66	4.08
Type2	6.5	0.91	3.0	6.52	3.73
Type3	14.3	0.78	0.0	12.67	2.16
Type4	25.5	0.46	0.0	-1.72	1.00
Type5	8.0	5.27	0.0	4.66	8.76

B
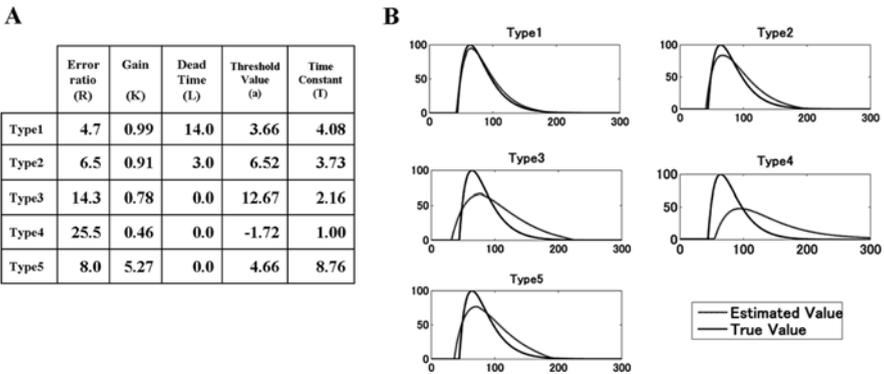
Fig. 12.4: Model parameters and estimation results for artificial data. **a**: Table of the estimation error standard deviation and the model parameters. **b**: Comparison between the estimation values and the true values.

We then applied STF model to artificially composed time series data that exemplifies each activation patterns. Figure 12.4a shows parameter values and the estimation error ratio for each category datum and Fig. 12.4b indicates the comparison between the estimated and the actual values for each category data. The estimation error is small when we estimate the respiratory motor activity using Type-1, Type-2, or Type-5 pixels, whereas it becomes bigger when we estimate it using Type-3 or Type-4 pixels. The dead time L of Type-1 pixel is large, and the gain K of Type-5 pixel is large. The results suggest that respiratory related pixels can be characterized by applying the STF model to optical time series data in each pixel. The conventional cross-correlation technique can only discriminate Type-1 from other activity patterns based on the maximum lag. Therefore, the present model provides a more sophisticated method to characterize dynamics of respiratory related optical signals.

12.3 Results

12.3.1 Estimation of STF Model Parameters

We applied the STF model to actual optical signals that were preprocessed by the method described in the previous section. Figure 12.5 shows the spatial distribution of the model parameters and the estimation error ratio.
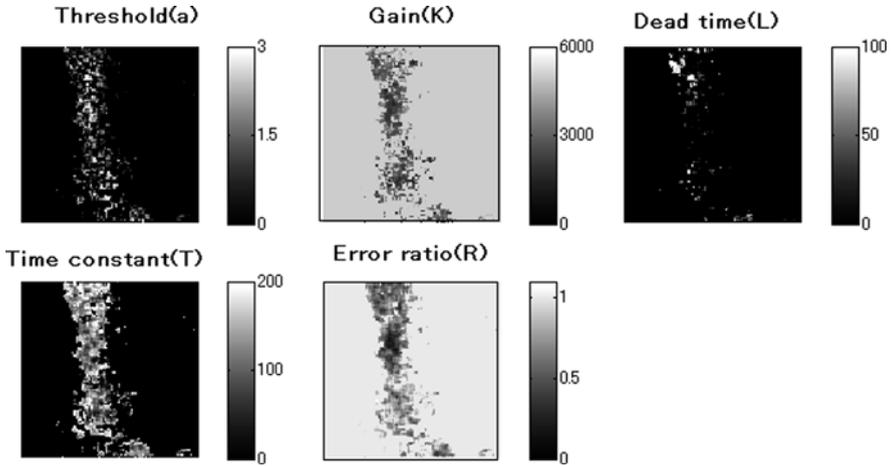
Fig. 12.5: Spatial distribution of the model parameters and the estimation error ratio.

To evaluate the variability of the dynamic characteristics of each pixel among breaths, we applied the STF model to optical signals of each pixel in each breathing epoch within consecutive 17 breaths and calculated the mean and the standard deviation of estimation error ratio in each pixel (Fig. 12.6a, b). We found that both the mean and the variance of estimation error ratio were small, and variances of model parameter were also small in the pre-BötC region, suggesting that the dynamics of neuronal activities in this region are robust and stable. In contrast in the pFRG region, although the mean of estimation error ratio was small, the variance was large. These results suggest that the dynamics of neuronal activities in the pFRG are more variable than those in the pre-BötC, which might be a reflection of loose synchronization within the preinspiratory neuron network in the pFRG [10]. Pixels on the outskirts of the pre-BötC area had larger mean and variance, suggesting that neurons in this area do not directly contribute to the respiratory motor pattern formation.
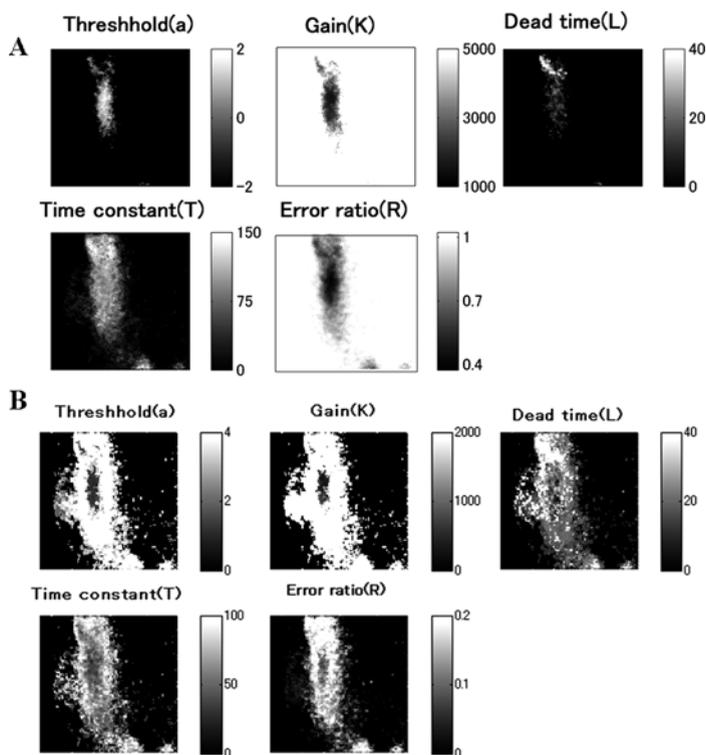
Fig. 12.6: Distribution of the mean values and the standard deviation values of the model parameters and the estimation error ratio. **a**: Mean value **b**: Standard deviation values.

12.3.2 Classification of Pixel Activity Patterns

We then categorized respiratory related pixels according to the criteria described in the previous section. The values of the criteria are shown in Fig. 12.7a. We next compared the estimated and the actual values for a representative pixel in each category. As shown in Fig. 12.7b, we obtained a good estimate using Type-1 or Type-2 pixels, even in the case of the SISO model. The dead time of Type-1 pixel was large, whereas the dead time and the gain were both small in Type-2 pixels. We did not obtain a good estimate using Type-3 pixels. The estimation precision became even worse when we use a Type-4 pixel. Type-5 pixels gave a good estimate, but the gain was high.

Figure 12.8 shows the spatial distribution of activity patterns of respiratory related pixels. Type-1 pixels were mostly distributed in the pFRG region, and Type-2 pixels were mainly distributed in the pre-BötC. These results are consistent with

A

	Dead time (L)		Error ratio (R)		Gain (K)	
	min	max	min	max	min	max
Type1	60	200	0	0.6	500	4000
Type2	0	40	0	0.4	500	4000
Type3	0	40	0.4	0.6	500	4000
Type4	0	0	0.6	0.8	500	4000
Type5	0	200	0	0.8	4000	10000

B

Fig. 12.7: Model parameters and estimation results of C4VR output. **a**: Model parameters and the estimation error ratio criteria for categorization. **b**: Comparison between the measured values (*true line*) and the estimated values (*dotted line*) of C4VR output.
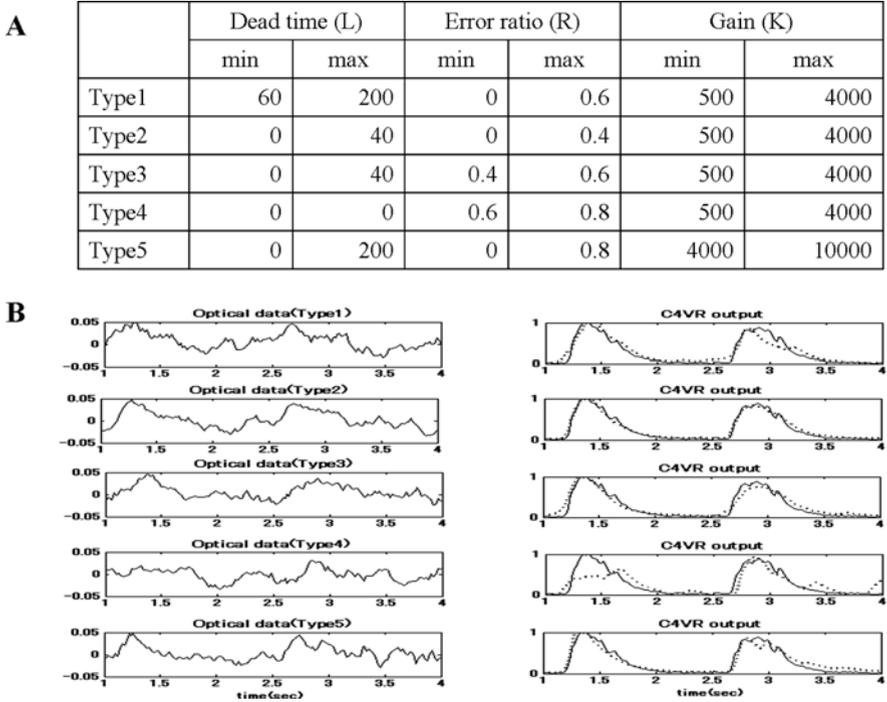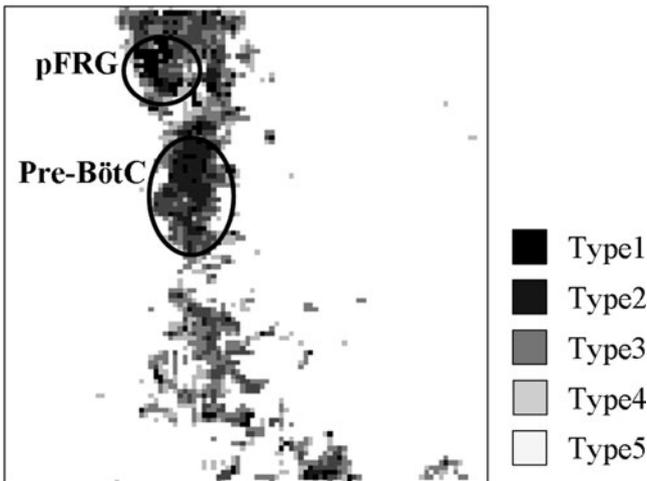

Fig. 12.8: Spatial distribution of activities pattern of imaging data.

the conventional cross-correlation analysis [22]. In addition, Type-3 pixels were observed in the vicinity of Type-2 pixel aggregates, and Type-4 pixels were found caudal to the pre-BötC.

12.4 Discussion

In this study, we developed a novel parametric modeling approach to objectively evaluate voltage-imaging signals recorded from the brain. We applied our model to voltage-imaging signals of the rat brainstem within a single breath without performing cycle triggered averaging. The union set of the detected area, which is predominated by each type in the model, corresponded to the extracted respiratory related areas reported by cross-correlation analysis [22]. In fact, our parametric model could decompose the known respiratory related area into substructures, each of which has a distinct functional property. In other words, by classifying each pixel into five types based on our model parameter values and the estimation error ratio, we could obtain more detailed categorization of neuronal activity patterns than by cross-correlation analysis. Although our STF model was simple, it could precisely estimate the respiratory motor output at least when the output pattern was unimodal. Further study is needed to test whether the model can estimate more complicated one such as a bimodal or a trimodal output pattern. We conclude that the parametric modeling approach can be effectively employed for the objective evaluation of voltage-imaging data of the brain and is expected to be universally applied to analyses of other types of imaging data.

References

1. Baker, B., Kosmidis, E., Vucinic, D., Falk, C., Cohen, L., Djurisic, M., Zecevic, D. Imaging brain activity with voltage-and calcium-sensitive dyes. Cell Mol Neurobiol **25**, 245–282 (2005)
2. Buzàki, G. Large-scale recording of neuronal ensembles. Nat Neurosci **7**, 446–451 (2004)
3. Cohen, L., Keynes, R., Hille, B. Light scattering and birefringence changes during nerve activity. Nature **218**, 438–441 (1968)
4. Cohen, L., Salzberg, B. Optical measurement of membrane potential. Rev Physiol Biochem Pharmacol **83**, 35–88 (1978)
5. Ebner, T., Chen, G. Use of voltage-sensitive dyes and optical recordings in the central nervous system. Prog Neurobiol **46**, 463–506 (1995)
6. Ezure, K. Reflections on respiratory rhythm generation. Prog Brain Res **143**, 67–74 (2004)
7. Feldman, J., Mitchell, G., Nattie, E. Breathing: Rhythmicity, plasticity, chemosensitivity. Annu Rev Neurosci **26**, 239–266 (2003)
8. Feldman, J., Negro, C.D. Looking for inspiration: New perspectives on respiratory rhythm. Nat Rev Neurosci **7**, 232–242 (2006)

 9. Fisher, J., Marchenko, V., Yodh, A., Rogers, R. Spatiotemporal activity patterns during respiratory rhythmogenesis in the rat ventrolateral medulla. J Neurophysiol **95**, 1982–1991 (2006)
10. Fujii, M., Umezawa, K., Arata, A. Dopamine desynchronizes the pace-making neuronal activity of rat respiratory rhythm generation. Eur J Neurosci **23**, 1015–1027 (2006)
11. Fukuda, K., Okada, Y., Yoshida, H., Aoyama, R., Nakamura, M., Chiba, K., Toyama, Y. Ischemia-induced disturbance of neural network function in the rat spinal cord analyzed by voltage-imaging. Neuroscience **140**, 1453–1465 (2006)
12. Fukunishi, K., Murai, N. Temporal coding in the guinea-pig auditory cortex as revealed by optical imaging and its pattern-time-series analysis. Biol Cybern **72**, 463–473 (1995)
13. Gill, P., Murray, W., Saunders, M., Wright, M. Procedures for optimization problemswith a mixture of bounds and general linear constraints. ACM Trans Math Software **10**, 282–298 (1984)
14. Gill, P., Murray, W., Wright, M. Practical Optimization. Academic Press, London (1981)
15. Kamino, K. Optical studies of early developing cardiac and neural activities using voltage-sensitive dyes. Jpn J Physiol **40**, 443–461 (1990)
16. Lindsey, B., Morris, K., Segers, L., Shannon, R. Respiratory neuronal assemblies. Respir Physiol **122**, 183–196 (2000)
17. Okada, Y., Chen, Z., Yoshida, H., Kuwana, S., Jiang, W., Maruiwa, H. Optical recording of the neuronal activity in the brainstem-spinal cord: Application of a voltage-sensitive dye. Adv Exp Med Biol **499**, 113–118 (2001)
18. Okada, Y., Kuwana, S., Masumiya, H., Kimura, N., Chen, Z., Oku, Y. Chemosensitive neuronal network organization in the ventral medulla analyzed by dynamic voltage-imaging. Adv Exp Med Biol **605**, 353–358 (2007)
19. Okada, Y., Masumiya, H., Tamura, Y., Oku, Y. Respiratory and metabolic acidosis differentially affect the respiratory neuronal network in the ventral medulla of neonatal rats. Eur J Neurosci **26**, 2834–2843 (2007)
20. Okada, Y., Mückenhoff, K., Holtermann, G., Acker, H., Scheid, P. Depth profiles of ph and p02 in the isolated brainstem-spinal cord of the neonatal rat. Respir Physiol **93**, 315–326 (1993)
21. Oku, Y., Kimura, N., Masumiya, H., Okada, Y. Spatiotemporal organization of frog respiratory neurons visualized on the ventral medullary surface. Resp Physiol Neurobiol **161**, 281–290 (2010)
22. Oku, Y., Masumiya, H., Okada, Y. Postnatal developmental changes in activation profiles of the respiratory neuronal network in the rat ventral medulla. J Physiol **585**, 175–186 (2007)
23. Onimaru, H., Homma, I. A novel functional neuron group for respiratory rhythm generation in the ventral medulla. J Neurosci **23**, 1478–1486 (2003)
24. Ramirez, J., Richter, D. The neuronal mechanisms of respiratory rhythm generation. Curr Opin Neurobiol **6**, 817–825 (1996)
25. Segers, L., Shannon, R., Saporta, S., Lindsey, B. Functional associations among simultaneously monitored lateral medullary respiratory neurons in the cat. I. Evidence for excitatory and inhibitory actions of inspiratory neurons. J Neurophysiol **57**, 1078–1100 (1987)
26. Smith, J., Ellenberger, H., Ballanyi, K., Richter, D., Feldman, J. Pre-Bötzinger complex: A brainstem region that may generate respiratory rhythm in mammals. Science **254**, 726–729 (1991)
27. Suzue, T. Respiratory rhythm generation in the in vitro brainstem-spinal cord preparation of the neonatal rat. J Physiol **354**, 173–183 (1984)
28. Tasaki, I., Watanabe, A., Sandlin, R., Carnay, L. Changes in fluorescence, turbidity, and birefringence associated with nerve excitation. Proc Natl Acad Sci **61**, 883–888 (1968)
29. Tominaga, T., Tominaga, Y., Yamada, H., Matsumoto, G., Ichikawa, M. Quantification of optical signalswith electrophysiological signals in neural activities of di-4-anepps stained rat hippocampal slices. J Neurosci Methods **102**, 11–23 (2000)
30. Yoshida, T., Sakagami, M., Katura, T., Yamazaki, K., Tanaka, S., Iwamoto, M., Tanaka, N. Anisotropic spatial coherence of ongoing and spontaneous activities in auditory cortex. Neurosci Res **61**, 49–55 (2010)

Chapter 13
Advances Toward Closed-Loop Deep Brain Stimulation

Stathis S. Leondopulos and Evangelia Micheli-Tzanakou

Abstract A common treatment for advanced stage Parkinsonism is the application of a periodic pulse stimulus to specific regions in the brain, also known as *deep brain stimulation* (or DBS). Almost immediately following this discovery, the idea of dynamically controlling the apparatus in a "closed-loop" or neuromodulatory capacity using neural activity patterns obtained in "real-time" became a fascination for many researchers in the field. However, the problems associated with the reliability of signal detection criteria, robustness across particular cases, as well as computational aspects, have delayed the practical realization of such a system. This review seeks to present many of the advances made toward closed-loop deep brain stimulation and hopefully provides some insight to further avenues of study toward this end.

13.1 Introduction

The uses of electrical stimulation and recording in medicine have a history dating back to the first century AD [95, 139, 121, 153, 76, 85, 21, 97, 138, 37, 30, 47]. However, since the first advances in microelectronics began to appear [7], medical electro-stimulation and recording equipment became portable and even implantable [23]. Soon after that, with the invention of the integrated circuit [84, 115], an ever-increasing number of components became available on a silicon chip of millimeter or even micron dimensions [107]. As a consequence, the availability and sophistication of electronic bio-implants began to greatly increase starting with the work of House [68] on the cochlear implant in 1969, the work of Humayun and de Juan [69] on the retinal implant in 1996, and the cortical implant reported by Donoghue [35] and Nicolelis [111] in 2002 and 2003.

S.S. Leondopulos
Rutgers University, Piscataway, USA, e-mail: `stathis@ece.rutgers.edu`

Evangelia Micheli-Tzanakou
Rutgers University, Piscataway, USA, e-mail: `etzanako@rci.rutgers.edu`

Electrical stimulation of nuclei in the basal ganglia of the brain as a treatment for Parkinson's disease, also known as *deep brain stimulation* (or DBS), was approved by the US Food and Drug Administration and became commercially available in 1997 [151]. The apparatus consists of a stimulus generator implanted under the collar bone and a subcutaneous lead connecting the stimulator to an electrode fixed at the cranium and reaching the basal ganglia in the center of the human brain. Following implantation, a wireless link facilitates communication with the implant for the routine adjustment of the stimulus waveform by medical staff. In this manner, the treatment can be tuned or optimized over time while avoiding side effects. The neural signals emanating from the basal ganglia during DBS have been recorded and analyzed by Dostrovsky et al. [36], Wu et al. [162], Wingeier et al. [158], and Rossi et al. [130]. Moreover, there have been studies regarding the use of information contained in the neural activity of the basal ganglia as a control signal or regulator of the stimulus apparatus [106, 146, 134, 78, 39, 90, 12].

13.2 Nerve Stimulation

The simplest model of electrical nerve stimulation was introduced by Arvanitaki and uses the passive membrane model with membrane resistance R_m and capacitance C_m [4, 95]. In this scenario, assuming the stimulus current applied across the cell membrane is a constant I_s, then the change in transmembrane voltage becomes

$$V_m(t) = I_s R_m \left(1 - e^{-t/R_m C_m}\right). \tag{13.1}$$

Moreover, given a threshold voltage ΔV_{th}, then the minimum stimulus current needed for the transmembrane voltage to reach ΔV_{th} is found for $t = \infty$ and is called the *rheobase current*:

$$I_{rh} = \frac{\Delta V_{th}}{R_m}. \tag{13.2}$$

Also, another useful measure of stimuli is the time required to reach ΔV_{th} when $I_s = 2I_{rh}$. This is called *chronaxy* or *chronaxie* [95, 154] and is calculated as

$$t_c = R_m C_m \ln 2. \tag{13.3}$$

As an example, Fig. 13.1 illustrates the decay of the minimum amplitude needed for stimulating a neuron as pulse width increases [99].

More sophisticated distributed models such as the core conductor model incorporate the shape of the neuron axon and conductivity of external media [24, 95]. Moreover, the shape and timing of stimuli are also influential as shown in detailed studies by Warman, McIntyre, Grill, and others [154, 99, 100, 54]. However, the passive membrane model with appropriate effective values for R_m and C_m remains a useful approximation for many applications [125, 74].
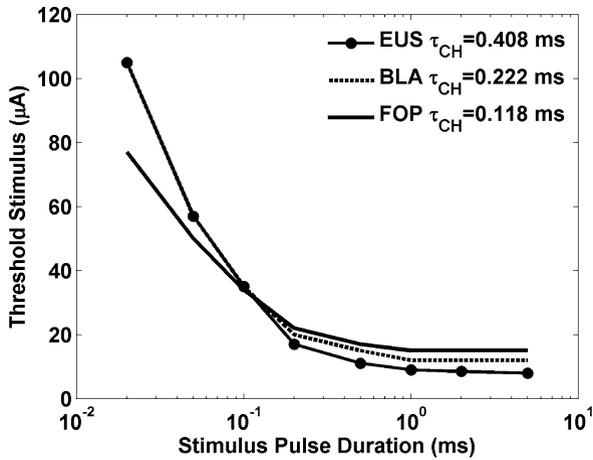
Fig. 13.1: Firing threshold of the external urethral sphincter motoneuron (EUS), the neuron innervating the bladder (BLA), and the fiber of passage in the white matter (FOP) stimulated with bipolar stimulation as predicted by simulation techniques and reported by McIntyre and Grill [99]. τ_{CH} represents the calculated chronaxie of the particular neuron).

13.3 Local Field Potentials

Measurable electrical phenomena that occur in the human body are due primarily to the transport of charged ions across the membrane of neurons as they relay and process information governing movement and perception. In particular, rapid changes in membrane permeability occurring on a millisecond scale produce current spikes or "action potentials" [9,65]. At the same time, thousands of synaptic junctions contribute to the "postsynaptic potential" or subthreshold changes in the transmembrane potential. Furthermore, random processes within the neuron membrane may cause spontaneous events to occur in addition to synaptic stimuli [81].

The *local field potential* (LFP) is related to the aggregate of the electric fields produced by individual neurons in the vicinity of the electrode within the dielectric medium of brain tissue. Furthermore, it is known that the recorded signal is influenced by a frequency filtering characteristic, so that only low-frequency elements of neural activity such as postsynaptic potentials propagate beyond the immediate cellular environment to produce measurable signals [11,10]. Also, characteristics of the analog front-end recording apparatus performing DC bias stability and prefiltering further modify the frequency band of the signal.

Bedard et al. [11, 10] have shown that the frequency-dependent attenuation with distance can be explained by using a nonhomogeneous model of extracellular dielectric properties that take into consideration the properties of neighboring neuron membranes. Also, at the macroscopic level, a comprehensive study of dielectric properties of tissues in the range of 10 Hz–20 GHz was prepared by

Gabriel et al. [45], including an empirical parametric model that fits well to the experimental data.

A more practical model for describing the dielectric properties at the neuroelectrode interface was developed by Johnson et al. [79]. In that study, an equivalent circuit model is used for explaining voltage-biasing effects of the recorded signal.

13.4 Parkinson's Disease

Parkinson's disease is due to the death or alteration of cells that produce the neurotransmitter dopamine in a region of the brain called *substantia nigra pars compacta* (SNc). In turn, the lack of dopamine weakens synaptic pathways between the SNc and the region called the striatum resulting in a general imbalance of activity within a group of brain nuclei collectively known as the basal ganglia [31]. As a result, the spike patterns of neurons in the *external globus pallidus* (GPe) become sparse, while the neurons in the *subthalamic nucleus* (STN) and *internal globus pallidus* (GPi) exhibit pronounced activity that is often in the form of synchronized oscillatory bursting [16, 92, 156, 71, 126]. Figures 13.2 and 13.3 show neural pathways of the basal ganglia as well as activity of key nuclei under normal physiological conditions and Parkinsonism, respectively. Moreover, dark arrows represent inhibitory synaptic pathways, gray arrows excitatory, and perforated arrows are pathways associated with dopamine. Externally, these processes are manifested as the Parkinsonian symptoms of essential tremor, muscle rigidity, bradykinesia (slowness of movement), and postural imbalance.
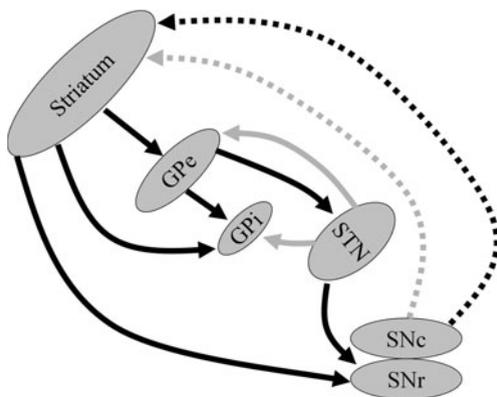
Fig. 13.2: Basal ganglia under normal conditions. This figure shows the nuclei in the basal ganglia and their synaptic paths including excitatory (*gray line*), inhibitory (*dark line*), and dopaminergic paths (*gray perforated line, dark perforated line*). A feedback loop between the STN and the GPe can be seen. This figure is modified from the figures reported by Gurney et al. [56] to emphasize changes due to dopamine depletion as described by Delong [31].
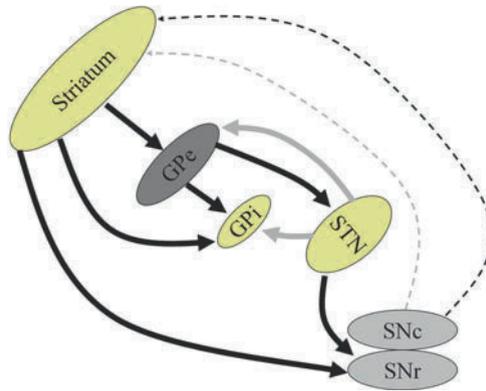
Fig. 13.3: Basal ganglia during a lack of dopamine (Parkinson's disease). Key nuclei and their synaptic paths including excitatory (*gray line*), inhibitory (*dark line*), and dopaminergic (*gray perforated line, dark perforated line*) paths are shown. *Dark-colored* nuclei signify diminished activity while *bright-colored* regions signify heightened activity. This figure is modified from the figures reported by Gurney et al. [56] to emphasize changes due to dopamine depletion as described by Delong [31].

13.4.1 Treatments

The treatment for early stage Parkinson's disease typically consists of the administration of levodopa (L-DOPA) orally. L-DOPA crosses the blood–brain barrier where it is converted into dopamine, thus restoring some of the movement capabilities to the patient. However, side effects that may emerge are dyskinesia (difficulty performing voluntary movements), depression, and psychotic episodes in some patients [28, 110].

Surgical procedures that have been used in the past as a treatment for advanced stage Parkinson's disease include pallidotomy, thalamotomy, and subthalamotomy [55]. In these procedures, functional MRI imaging techniques detect the location of specific nuclei in the brain of the patient. Following this, stereotactic surgical techniques are employed for the placement of electrodes at the target location. Next, electrode recordings are analyzed to achieve a more precise placement [59]. Finally, high temperatures (80°C) or electric currents are applied to cause destruction of cells (known as lesioning) in the STN or GPi.

The success of pallidotomies is hypothesized to be due to a reduction of activity in the GPi that is caused by the administered (or artificially placed) lesions [84]. Furthermore, lesioning the STN with a subthalamotomy has a similar effect in the GPi because of the excitatory neuronal paths from the STN to the GPi [3]. Thus, lesions in the GPi simulate the inhibitory input to the STN and GPi that would otherwise be present under physiological conditions (see Figs. 13.2 and 13.3).

13.5 Deep Brain Stimulation

Electrical stimulation of the brain as a treatment for Parkinson's disease was first reported by Benabid et al. [13] in 1987. In particular, during stereotactic neurosurgery it was observed that stimulating the *ventral intermediate nucleus* (VIM) of the brain with a sequence of 1–2 V 0.5 ms pulses at 100 Hz blocked symptoms of the disease. Eventually, the lesioning procedures mentioned previously were replaced by the implantation of electrodes connected to a pulse generator. Moreover, the physician could tune the signal generator through a wireless link, thus adjusting the stimulus parameters.

13.5.1 DBS Mechanism

A primary contributing factor to the inhibitory effect of DBS on the STN and GPi is likely the release of adenosine by astrocytes as they are electrically stimulated [12]. Also, the same study reports how the inhibition is likely a combination of adenosine-related and "axonal" effects. That is, there are a number of hypotheses that attempt to explain the inhibitory effect of DBS on the STN and GPi. In particular, these are: (1) the blocking of action potentials by affecting properties of ion conductance in the neuron membrane, (2) the preferential stimulation of axons that terminate at inhibitory synapses rather than neurons themselves, and (3) the desynchronization of mechanisms occurring in the network as a whole. Out of these hypotheses, desynchronization seems to be the least refuted and least understood [101].

In practice, the effect of DBS on neural activity can be seen in recordings using extracellular electrodes that have been taken from patients during surgical implantation of DBS systems, as shown in Fig. 13.4. In particular, the work of Dostrovsky et al. [36] shows how the activity of pallidal neurons displays a period of quiescence after each stimulating pulse of DBS. Furthermore, the quiescent period increases with respect to the DBS pulse amplitude as can be seen in Fig. 13.5. Also, as the pulses become more dense at higher frequency stimulation, the quiescent periods seem to overlap, thus causing the inhibitory effect. A more macroscopic view of the effect of pulse amplitude is provided in Fig. 13.6 [162].

Figure 13.7 shows the neuron activity rate following a stimulus pulse measured as a percentage of the activity preceding the pulse (baseline activity). As can be seen in Fig. 13.7, neural activity is nearly 0 after the DBS pulse, but returns to normal firing after some time (between 50 and 100 ms).

13.5.2 Apparatus

All commercially available DBS systems are currently designed and manufactured by the Medtronic corporation. By name, the neurostimulators commonly used for
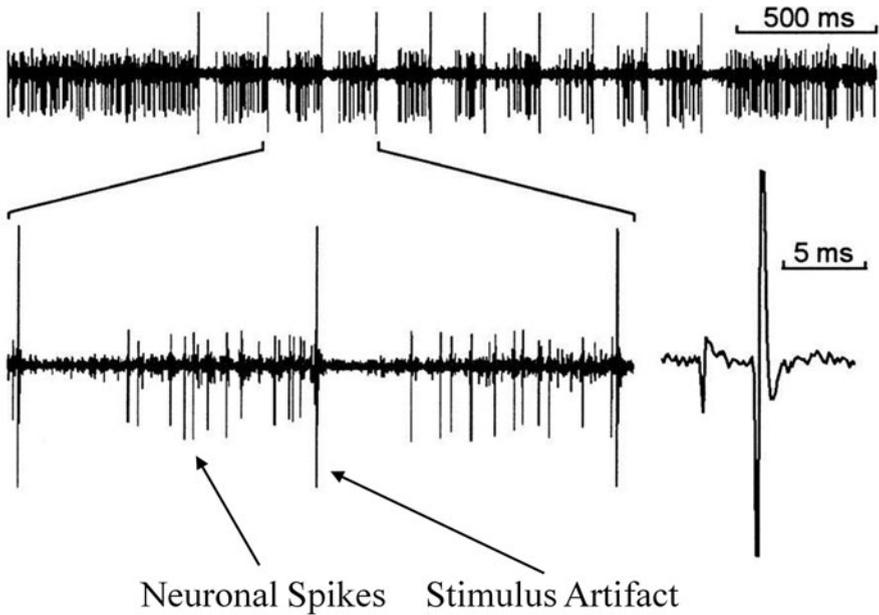
Fig. 13.4: Effects of DBS pulses on neural activity in the GPi as observed experimentally and reported by Dostrovsky et al. [36]. The larger *vertical line* segments are stimulus artifacts while the *shorter line* segments can be attributed to neuronal spike activity. A quiescent or inhibitory period during which there is no neuronal activity can be observed after each stimulus.

DBS are the "Itrel II Soletra," "Kinetra," and "Extrel" units (with Extrel used less frequently than the former two). Moreover, the specifications of the apparatus have been described in a number of publications [59, 101, 5, 89, 152]. Specifically, a 1.27 mm diameter probe with four 1.5 mm long contacts spaced 0.5 mm or 1.5 mm apart (depending on the version) is in contact with the target area of the brain and secured to the cranium at its base. Furthermore, a subcutaneous lead connects the base of the probe to a $53 \times 60 \times 10\,\text{mm}^3$ neurostimulator implanted in the chest area under the collarbone of the patient [101].

   The Extrel unit differs from the Soletra and Kinetra units in that an external stimulus generator communicates with the implant. In particular, the external apparatus generates the pulse waveform and then modulates it using a carrier frequency in the RF range. In turn, an implanted receiver demodulates the signal using passive circuit components including a capacitor [89, 137, 102].

13.5.3 Stimulus Specifications

The DBS units are capable of applying stimulus waveforms that consist of a train of pulses with the following specifications [152, 101]:
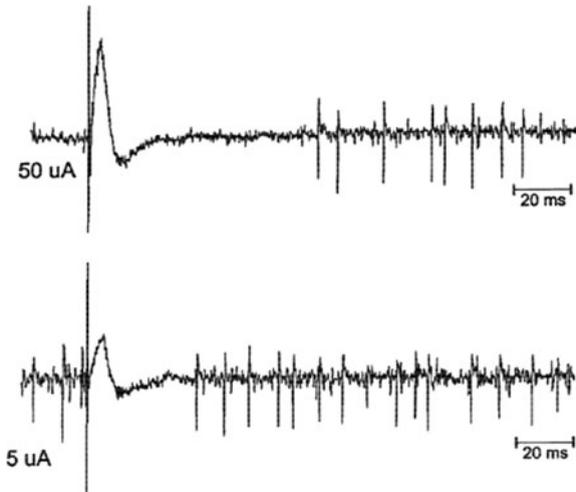
Fig. 13.5: Detail of the effects of a 50 and 5 µA DBS pulse of duration 150 µs on a single GPi neuron of a Parkinson's patient as observed experimentally and reported by Wu et al. [162]. The *tallest thin vertical line* segments are the stimulus artifacts, while the *shorter line* segments can be attributed to neuronal spike activity. A large pulse immediately followed by an inhibitory period is observed following the stimulus. Moreover, the smaller stimulus (5 µA) is followed by a short inhibitory period (roughly 30 ms), while the larger stimulus is followed by a longer inhibitory period (roughly 60 ms).

Pulse amplitude: 0–10.5 V (in steps of 0.1 V), and assuming a 1 kΩ load as re-
    ported, this means a 0–10.5 mA stimulation current.[1]
Pulse duration: 60–450 µs (1,000 µs maximum in the case of Extrel).
Pulse frequency: 2–185 Hz in the Soletra, 2–250 Hz in the Kinetra, and 2–1,000 Hz
    in the Extrel.
Pulse polarity: both monopolar and bipolar modes are available (only bipolar in
    the Extrel).

---

[1] The amplitude used in commercial DBS units (0–10.5 mA) is obviously much larger than what is reported in the experiments of Dostrovsky et al. [36], Hamilton et al. [60], and Lehman et al. [91], namely 5–100 µA. However, the current density turns out to be similar because of the differences in electrode diameter. In particular, the experimental work sited uses 25 µm (length) by 25–100 µm (diameter) electrodes, while commercial devices use a 1.5-mm (length) by 1.27-mm (diameter) electrodes.
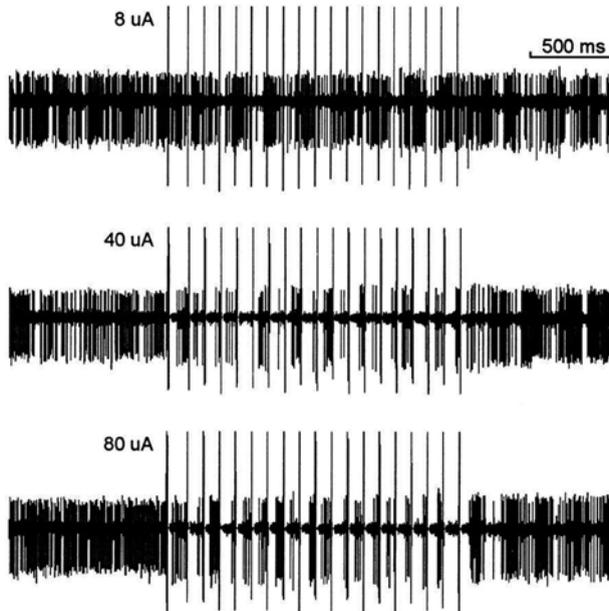
Fig. 13.6: Effects of DBS pulses (at 10 Hz) on a single GPi neuron in the GPi as observed experimentally and reported by Dostrovsky et al. [36]. The larger vertical line segments are stimulus artifacts, while the shorter line segments can be attributed to neuronal spike activity. It can be seen that as stimulus energy increases from 8 to 80 μA, the neural activity becomes more sparse.

13.5.4 DBS Programming

The typical procedure for programming DBS apparatus postoperatively begins with the determination of the "therapeutic window" of stimulation for each electrode [5, 152]. That is, using monopolar stimulus, keeping the pulse width at 60 μs and the frequency at 130 Hz, the pulse amplitude is increased from 0 V at increments of 0.2–0.5 V. Furthermore, the therapeutic window or range for a particular electrode is the set of amplitude values between the smallest therapeutic amplitude and the onset of undesirable side effects such as rigidity and dystonia (sustained muscle contractions). Next, the electrode with the largest therapeutic range is selected as the stimulus electrode [152].

Over the months following implantation, DBS parameters are modified according to the side effects and therapeutic results observed. Typically, the amplitude or frequency is increased as the patient develops a tolerance to the stimulus effect. Moreover, it is believed that a higher impedance or displacement of the electrodes due to glial tissue scarring is responsible for the diminishing effectiveness of DBS over the first postoperative months [40, 108]. In addition, long-term physiological processes
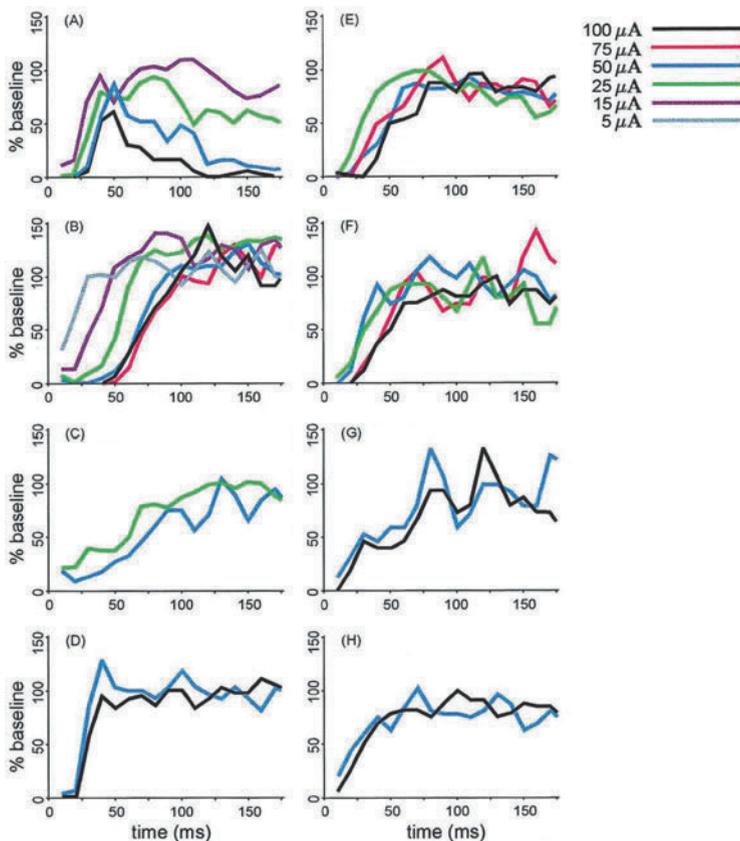
Fig. 13.7: Spike-rate in 10 ms bins, smoothed with a 20 ms sliding window, as percentage of baseline (no stimulus) and a function of time (stimulus at time 0) as observed experimentally and reported by Dostrovsky et al. [36]. A period of quiescence or inhibition can be seen immediately following a stimulus. Then, normal neural firing rates gradually resume.

influenced by neural activity cause the modification of synapses, thus strengthening or weakening the influence of one neuron on the behavior of another [140].

Increasing the pulse width is avoided due to the recruitment of and possible damage to adjacent brain centers and the resulting side effects such as dysarthria (a speech disorder) and ataxia (loss of movement coordination) [152, 99, 100]. For example, Fig. 13.8 shows curves of the minimum pulse width–amplitude combinations that cause tremor suppression and onset of adverse side effects as found through experimentation on human subjects. Moreover, this is a verification of the response of the theoretical lumped parameter model shown previously in Fig. 13.1.

In DBS, bipolar stimulation is avoided due to the higher power dissipation that it requires. Only if side effects persist, the bipolar mode turned on because of the
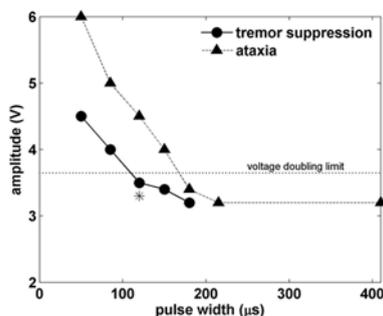
Fig. 13.8: Minimum pulse width–amplitude combinations causing tremor suppression and onset of adverse side effects as found experimentally and reported by Volkmann et al. [152]. The *asterisk* shows the pulse width suggested by Volkmann, while the voltage-doubling limit is a property of the Itrel II and Soletra stimulus generators reported by Volkmann.

more localized stimulation that it provides [5, 14]. At 6 months postoperatively, the stimulation parameters require only minor adjustments, as reported by Ashkan [5].

13.5.5 Side Effects

The undesirable side effects of DBS are primarily due to excess current leakage into adjacent brain centers and include cognitive degradation and severe emotional disturbances. However, other ill side effects may occur when DBS therapy is administered in conjunction with unrelated methods of diagnosis or treatment. For example, electrodes may be displaced by intense electromagnetic fields during MRI sessions, thus causing damage to brain tissue and displacing the location of the applied stimulus. Also, temperatures may become dangerously high during the administration of therapeutic diathermy (tissue heating), thus resulting in massive trauma or death [115, 131].

13.6 Biosignal Processing

All biological processes associated with perception and limb movement involve measurable electrical phenomena. Moreover, depending on where and how a measurement is taken, the recorded signal will exhibit particular characteristics [65, 144]. Typically, biosignal processing involves the analysis and classification of

recorded biosignals using any combination of signal processing techniques that are suitable for the particular application at hand [25]. In particular, the signal processing reduces the dimensionality of the data space by extracting useful information or "features" of the signal [29]. Thus, the high-dimensional recorded data is mapped to a lower dimensional "feature space." Moreover, the feature space is divided into regions or "classes" in order to categorize each measured signal.

13.6.1 Features

Biosignals can be analyzed using a large set of signal processing methods. However, some features are relatively simple to calculate while others are computationally demanding. Moreover, the issue of computational complexity becomes particularly important for integrated circuit implementations. Accordingly, Table 13.1 shows the computational complexities of various useful features in terms of signal sample size N, filter order n, decomposition levels L (for wavelets), number of signals m (PCA), lag q in terms of clock cycles, and the number of ALOPEX iterations c [29] (a blank "–" where present indicates that no studies were found).

Table 13.1: Feature extraction methods

Method	Complexity	Parallel and/or pipelined
Mean	$O(N)$	$O(\log(N))$
Variance	$O(2N)$	$O(2\log(N))$
FFT [124, 26]	$O(N\log(N))$	$O(\log(N))$
LPC (Levinson) [33, 87]	$O(nN + n^2)$	169 cycles/iteration
Wavelets (lifting) [93]	$O\left(4 + 2N\left(1 - 1/2^L\right)\right)$	–
Karhunen–Loeve with ALOPEX [29]	$O(2cN)$	$O(2c\log N)$
PCA – SGA [32]	Onm	$O(n^2)$
Third-order cumulant (skewness) [1]	$O(Nq^2 + 3qN)$	$O(N + q)$
Fourth-order cumulant (kurtosis) [96]	$O(N^6)$	–

[a] The 169 clock cycles (actually 3,378 per 20 iterations) for a pipelined multiplier implementation of the Levinson algorithm are reported in [136], however, there is no explicit mention of complexity in that paper. It seems evident, however, that for p multipliers in parallel, a pipelined implementation of the Levinson algorithm would be $O\left(\frac{N}{p} + n^2\right)$. Also, $O(L^4)$ is mentioned in [141] for fourth-order moments.

13.6.2 Classifiers

When some features of measured neural activity contain useful information that can be applied in regulating a stimulus generator, a method for automated classification may be in order. To this end, there are various methods that can be employed

broadly categorized as probability density estimation, nearest neighbor search, and neural networks [88, 66, 2]. In particular, probability density estimation or Bayes estimation categorizes the measurement in order to minimize the probability of error, nearest neighbor search finds the class that is associated with the nearest neighbors of the measurement, while neural networks consist of simple interconnected computational elements that have the end result of dividing the feature space into specific regions [59, 60, 58].

Among these classifiers, neural networks seem to be the most widely used methods in biomedical applications. However, choosing the best classifier as well as a feature set for a particular case is often an empirical task. Thus, a set or "ensemble" of different classifiers is often used for a single classification task [116].

13.6.3 Feature Selection

Selecting the features that minimize a cost function, such as the probability of misclassification, can be done exhaustively by examining each subset. However, this process is of complexity $\binom{N}{n}$ and may become intractable for large feature sets. Alternatively, there are a number of methods that reduce the complexity of the task, including "branch and bound," "sequential forward and backward selection," "Plus-l-take-away-r algorithm," and "max–min feature selection" [122, 19, 118].

13.7 Closed-Loop DBS

Following the discovery of the effects of electrical brain stimulation on the symptoms of Parkinson's disease [13] in 1987, investigations were initiated to explain how the stimulus achieved the desired result [101, 54]. Also, methods for administrating the newfound treatment as an implantable "brain pacemaker" were being explored [106, 146, 134, 54, 127, 78, 39]. In particular, the first disclosure of such an apparatus was the original patent on DBS filed by Rise and King [127] of the Medtronic corporation in 1996, where a system consisting of an electrode sensor, a microprocessor, stimulus generator, and additional peripheral circuitry was proposed for the purpose of measuring tremor-related symptoms in the arm and adjusting stimulus parameters based on the measurements. Subsequently, another patent was filed by John [78] in 2000, elaborating on the original proposal by including provisions for multiple sensors such as electrodes implanted in the brain and/or surface electrodes on the scalp and limbs. In addition, John proposed particular signal processing methods for assessing the measured data including the computation of signal variance, correlation, discrete Fourier transform, peak detection, and Mahalanobis distance or Z-scores. Also, provisions for wireless data telemetry to an external PC or handheld processor were included in that patent.

In the scientific literature, improvements to DBS have been suggested by a number of authors [106, 146, 134, 39]. In particular, Montgomery and Baker [106] suggested that a future direction of DBS would be to incorporate the ability of acquiring and decoding neurophysiological information "to compute the desired action." Also, using results from a mathematical model of interconnected phase oscillators, Tass [146] proposes a method of demand-controlled double-pulse stimulation that would hypothetically enhance the effectiveness of DBS while reducing the power consumption of a stimulator in the long term. In addition, Sanghavi [134] and Feng et al. [39] propose methods for adaptively modifying stimulus parameters while seeking to minimize measures of brain activity in the vicinity of the implant.

13.7.1 Demand-Controlled DBS

From a theoretical perspective, Tass established a stimulus methodology based on a model of Parkinsonian brain activity [146, 147]. In particular, Tass simulated the synchronized oscillatory behavior of the basal ganglia using a network of phase oscillators. This method is as follows: given N oscillators with global coupling strength $K > 0$ where the phase, stimulus intensity, and noise associated with the jth oscillator are Ψ_j, I_j, and $F_j(t)$, respectively, the behavior of the jth oscillator and its relation to other oscillators as well as the stimulus is shown in Equations (13.4), (13.5), and (13.6). In particular, defining factors $S_j(\Psi_j)$ and $X_j(t)$ as

$$S_j(\Psi_j) = I_j\cos(\Psi_j) \text{ and} \tag{13.4}$$

$$X_j(t) = \begin{pmatrix} 1: \text{neuron}_j \text{ is stimulated} \\ 0: \text{otherwise} \end{pmatrix}, \tag{13.5}$$

the rate of change of the jth phase oscillator is given by

$$\dot{\psi} = \Omega - \frac{K}{N}\sum_{k=1}^{N}\sin(\psi_j - \psi_k) + X_j(t)S_j(\psi_j) + F_j(t). \tag{13.6}$$

Tass showed that the model in Equations (13.4), (13.5), and (13.6) is able to generate patterns of both synchronized oscillatory firing and random nonoscillatory behavior. Moreover, the network tends to remain in a synchronized oscillation until a global stimulus is applied at time t_0 so that $X_j(t_0) = 1$ for all j.

Effective stimulation methods for suppression of abnormal burst activity in this model, as reported by Tass, include low-amplitude high-frequency stimulation (20 times the burst frequency), low-frequency stimulation (equal to the burst frequency), or a single high-amplitude pulse, with the high-amplitude pulse being the most effective when it is applied at the appropriate phase of each neuron. Furthermore, Tass proposes a demand-controlled stimulation technique whereby the synchronicity

among individual oscillators is measured, and when passing a predefined threshold, it activates a stimulation pulse.

In order to detect synchronicity among neurons, Tass proposes the calculation of cluster variables – the center of gravity in phase space of all oscillators. Specifically, if $R_m(t)$ and $\phi_m(t)$ are the magnitude and phase respectively of the center of gravity of m clusters, and Ψ_j is the phase of the jth oscillator, then the cluster variable is

$$Z_m(t) = R_m(t)e^{i\phi_m(t)} = \frac{1}{N}\sum_{j=1}^{N} e^{im\Psi_j(t)}. \tag{13.7}$$

Thus, if the magnitude of the cluster variable is close to 0, there is very little synchronicity, but when it is close to unity, there is high synchronicity.

13.7.2 ALOPEX and DBS

Sanghavi [134] proposed an integrated circuit (IC) design of an adaptive DBS system where power estimation of recorded neural activity is used as a global "error measure" that drives the modification of stimulus pulse width, amplitude, and frequency of multiple signal generators. Furthermore, the modification is accomplished in simulation with minimal power requirements (roughly 0.8 mW) using an analog design of the stochastic optimization algorithm ALOPEX.

Since its application to BCI [150, 62, 105, 38], the ALOPEX algorithm was applied to numerous studies involving image pattern recognition and artificial neural networks [29]. The algorithm itself is based on the principle of Hebbian learning wherein the synaptic strength between two neurons increases in proportion to the correlation between the activities of those neurons [140]. Similarly, given a set of modifiable variables at iteration k, $b_k = \{b_{1,k}, b_{2,k}, , b_{N,k}\}$, and a global response estimate R_k, ALOPEX recursively modifies each $b_{j,k}$ by using correlation measures between previous changes in $b_{j,k}$ and changes in R_k. Moreover, to keep the algorithm from falling into an infinite loop, stochastic noise $r_{j,k}$ is included. Finally, given stochastic and deterministic step sizes $\sigma_{j,k}$ and $\sigma_{j,k}$, a reformulation of the algorithm in its most simplified "parity" form, as it is described in [62], is

$$d_{j,k} = \frac{(R_{k-1} - R_{k-2})}{|R_{k-1} - R_{k-2}|} \cdot \frac{(b_{j,k-1} - b_{j,k-2})}{|b_{j,k-1} - b_{j,k-2}|}, \tag{13.8}$$

$$b_{j,k} = b_{j,k-1} + \gamma_{j,k} \cdot d_{j,k} + \sigma_{j,k} \cdot r_{j,k}. \tag{13.9}$$

Subsequently, new versions were developed including the 2T-ALOPEX algorithm contributed by Sastry et al. [135] and the ALOPEX-B algorithm contributed by Bia [18]. In particular, 2T-ALOPEX incorporates explicit probability distributions into the calculation of each iteration, while ALOPEX-B is a similar but

simplified version of 2T-ALOPEX. Finally, Haykin et al. [63] improved convergence by combining the original formulation with that of Bia. Moreover, Haykin et al. provide a good contextual introduction and derivation of ALOPEX, while Sastry et al. prove that 2T-ALOPEX behaves asymptotically as a gradient-descent method. Also, Meissimilly et al. [103] introduced parallel and pipelined implementations of ALOPEX applied to template matching with corresponding computational and temporal complexities of calculating the global response function R_k.

13.7.3 Genetic Algorithms and DBS

Feng et al. [39] use a model by Terman et al. [149] to test a method of stimulus administration where each stimulus parameter is obtained from a distribution of such measures, thus incorporating a degree of randomness in the stimulus waveform. Moreover, in this method, the shape of each distribution curve is a piecewise linear model where the model parameters are modified by a genetic algorithm that seeks to reduce the cross-correlation and/or autocorrelation of measurements taken from multiple sensors. Figure 13.9 shows a diagram of the method proposed by Feng et al.
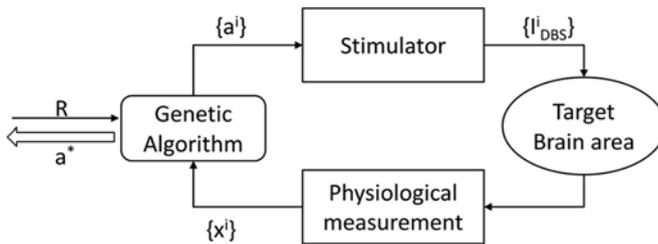
Fig. 13.9: The method proposed by Feng et al. [39] to draw deep brain stimulation parameters (I_{DBS}^i) from distributions whose shape descriptors (a^i) are selected by a genetic algorithm that seeks to minimize correlations in measured data (x^i). Constraints (R) on the genetic algorithm may be imposed externally.

13.7.4 Hardware Implementations

Various components of a closed-loop system have been implemented as a microelectronic design, including power and telemetry components [159], and stimulus/recording circuits interfacing with an external computing platform [90]. A typical setup for the real-time transmission of biosignals from a neural implant is shown in Fig. 13.10 [159].
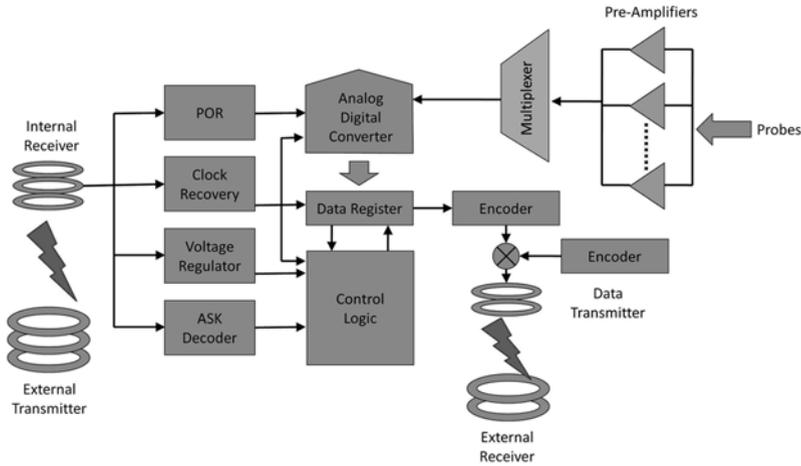
Fig. 13.10: A system for recording and decoding neuron activity. Power and data are transmitted through wireless telemetry [159].

13.8 Related Advances in Other Neuroprosthetic Research

Real-time biosignal processing has also advanced in other applications of neural prostheses in addition to DBS, such as cardiac pacemakers [133], retinal and cochlear implants [123, 69, 144], and brain-to-computer interfaces (BCI) [150, 62, 48, 91, 132, 161, 155, 49, 46]. In particular, pattern recognition systems for detecting abnormal heart activity have been proposed for cardiac pacemaker technology [133, 86]. Also, the decoding of neural activity in the premotor cortex of the brain to control robotic limbs has been successfully implemented in experiments with primates [111, 35]. Moreover, wireless telemetry and power transfer to implanted circuitry have been successful for cochlear and retinal implants [109]. There has also been research on detecting epileptic seizures and building an artificial hypocampus [72, 15].

Retinal and cochlear implants are relevant to DBS because of their wireless power transfer and data telemetry capabilities [123, 69, 144], while real-time signal processing of biosignals seems to have advanced more in cardiac pacemaking [6, 103, 128, 42] and especially BCI systems [150, 62, 48, 91, 132, 161, 155, 49, 46].

A typical setup for the real-time transmission of biosignals from a neural implant includes sensors (chemical or electrode) for detecting neural activity, signal processing for coding the activity, and communications circuitry for transmitting the information as shown in Fig. 13.10. In addition, the need for analog amplifiers, filters, and stimulus generators is ubiquitous among these designs [159]. Thus, methods included in the preprocessing and stimulus pulse generation stages have

also been proposed including amplifier designs [50, 117, 52], analog-to-digital conversion (A/D) [51], and voltage multiplier designs [113].

13.8.1 Closed-Loop Cardiac Pacemaker Technology

Some research in cardiac pacemaker technology has sought to modify stimulus parameters in response to measured neural activity. Moreover, this notion of autonomous regulation is similar in principal to adaptive, autonomous, or closed-loop *deep brain stimulation* (DBS).

The current standard for signal processing in cardiac pacemaking still consists of a simple band-pass filter with adaptive threshold detection [6, 103, 128]. However, new methods have been proposed that also include nonlinear filtering, wavelet analysis, and linear regression as well as threshold detection [86, 128, 42]. For example, Rodrigues et al. [128] implement filter banks (wavelets) with linear regression and threshold techniques in an IC design for detecting "R-waves" in cardiograms. In particular, given an input waveform $x(n)$ and wavelet filter H, the output of the wavelet decomposition is

$$y(n) = x(n)^T H. \tag{13.10}$$

Next, the "decision signal" is computed as

$$T(n) = x(n)^T H (H^T H)^{-1} H^T x(n). \tag{13.11}$$

Finally, the detection of the R-wave is considered positive if for some $\beta > 0$ and maximum decision signal $T_{\max}$, $T(n) \geq \beta T_{\max}$. Furthermore, complexity of the algorithm is $O(N)$, while the circuit design reported in [128] requires 6 multiplications and 45 summations per iteration and achieves a performance of roughly 99% correct detection and less than 1% false alarm.

13.8.2 Brain-to-Computer Interface

The first reported brain-to-computer interface (BCI) employing an adaptive algorithm and feedback was reported by Tzanakou et al. [150, 105, 38] where pixels on a screen were modified by the ALOPEX algorithm [62] to excite particular neurons (or receptive fields) in the visual pathway of a frog brain. Recently, BCI methods have been reported for detecting intended movements of primates. These include linear methods such as the "population vector" algorithm [48], finite impulse response (FIR) filters [132], Kalman filtering [161], nonlinear methods such as neural networks (NN) including time-delay NN's (TDNN) [155], gamma models [49] and recurrent NN's [132], and probabilistic approaches such as Bayesian inference [46]. Moreover, the nonlinear methods tend to achieve more accurate results at the expense of computational complexity.

In the case of linear methods, a typical formulation consists of sampling neuron spike-counts at intervals of 50 ms from multiple (15) recording sites. Moreover, the training stage consists of sampling roughly 1 s of data (20 intervals) and storing this information into a matrix $R_{(20 \times 15)}$ while storing the resulting hand position in terms of $x - y$ coordinates into a vector k. Next, the filter is constructed as $f = (R^T R)^{-1} R^T k$ and the reconstruction of movement for a history of neural activity R is obtained as $u = R \times f$.[2] In addition, there are more sophisticated formulations that take into account the velocity and acceleration of the movement as well as prior information about the behavior of neurons in the cortex [82].

Almost all reported BCI methods utilize the same preprocessing stage that consists of spike detection, sorting, and counting over an interval typically in the range of 50–100 ms. Moreover, correlation methods and principal component analysis (PCA) with threshold detection are reported as methodologies for the spike detection [22, 80]. However, Wessberg et al. [155] report using straight linear regression with no spike detection.

13.9 Neural Network Modeling and the Basal Ganglia

The neurocomputational aspects of Parkinson's disease and DBS have been examined using neural network models. Aside from their usefulness as classifiers [129, 98, 67], static neural networks have been used to model the basal ganglia and the outcome of pallidotomies [60, 58, 104]. In addition, the temporal characteristics of neurons in these areas and the effects of DBS on neural activity have been investigated using dynamic, pulsed, or spiking neural networks [56, 57, 53, 17, 43, 44, 70, 149, 54, 8]. The models employed typically include Hodgkin–Huxley formulations as well as larger networks of simpler integrate-and-fire units [70]. However, there is a plethora of models that range in complexity and accuracy that may be used to this end, such as the Noble [112] and Fitzhugh-Nagumo [41] models, as well as many others [136, 61, 64, 160, 157, 143, 141, 27, 73, 75].

Three general methods of modeling nuclei of the basal ganglia can be found in the scientific literature. These can be broadly categorized into "functional" models that are designed to provide insight into the computational function of the basal ganglia [56, 57, 53, 17, 43, 44, 142, 8], "physiological" models that incorporate more details of ion transport [70, 149, 54], and "conceptual" models [20, 77, 145, 34, 148] that provide a description of the synaptic connectivity. Moreover, the physiological models have been used in simulations of applied deep brain simulation (DBS). In particular, Grill et al. [54] show that extrinsic high frequency stimulation "masks" or

---

[2] The formulation is included here as it appears in the literature. However, there are some unresolved questions. In particular, it would seem that a separate filter would be required for each movement element so that given a history of 20 positions, there are corresponding x and y-coordinate vectors x and y of 20 elements each. In that case, two filters would be derived as $f_x = (R^T R)^{-1} R^T x$ and $f_y = (R^T R)^{-1} R^T y$. Then, given a set of new data S in the testing phase, the corresponding hand positions would be given as $x_{\text{new}} = S \times f_x$ and $y_{\text{new}} = S \times f_y$.

prevents internal activity of single neurons from being expressed at the output, thus causing an "informational lesion," while Feng et al. [39] use a model by Terman et al. [149] to test a novel method of stimulus administration. Also, in response to in vitro studies of the rat GPe and STN [120], Humphries and Gurney [70] design models that reproduce the oscillatory and bursting modality of the neural circuits. In addition, an analog CMOS model of Parkinsonian activity has been investigated by Sridhar [142].

13.10 Summary

Overall, various methods for implementing a closed-loop neuromodulator have been presented including conceptual schemes in simulation as well as hardware designs facilitating the goal. Also, both experimental and simulation studies have provided some insight into the neural mechanisms involved in the success of DBS. However, there remains a need for some performance criteria in deciding which method of closed-loop DBS will be the most successful. To this end, some preliminary comparisons of computational complexity are merely a starting point. What is needed is a rigorous test on animal and human subjects including quantitative measures of success in reducing symptoms while avoiding side effects. Ultimately, the progress will depend on what is (or is not) approved by organizations such as the United States (US) Food and Drug Administration (FDA) [119].

References

1. Ahmed, R.E., Al-Turaig, M.A., Alshebeili, S.A. VLSI architecture for computing third-order cumulants, Int J Electron **77**(1), 95–104 (1994)
2. Alippi, C., Braione, P. Classification methods and inductive learning rules: What we may learn from theory. IEEE Trans Syst, Man, Cybern C Appl Rev **31**(4), 364–378 (2006)
3. Alvarez, L., Macias, R., Guridi, J., Lopez, G., Alvarez, E., Maragoto, C., Teijeiro, J., Torres, A., Pavon, N., Rodriguez-Oroz, M.C., Ochoa, L., Hetherington, H., Juncos, J., De Long, M.R., Obeso, J.A. Dorsal subthalamotomy for Parkinsons disease. Mov Disord **16**(1), 72–78 (2001)
4. Arvanitaki, A. Les variations gradues de la polarisation des systmes excitables, Thesis, University Lyons, Hermann et cie, Paris, 1938.
5. Ashkan, K., Wallace, B., Bell, B.A., Benabid, A.L. Deep brain stimulation of the subthalamic nucleus in Parkinsons disease 1993 2003: Where are we 10 years on? Br J Neurosurg **8**(1), 19–34 (Feb 2004)
6. Bai, J., Lin, J. A pacemaker working status telemonitoring algorithm. IEEE Trans Inform Technol Biomed **3**(3) (Sep 1999)
7. Bardeen, J., Brattain, W.H. The transistor, a semiconductor triode. Phys Rev **74**(2), 230 (1948)
8. Barto, A.G. Adaptive critic and the basal ganglia. In Houk, J.C., Davis, J.L., Beiser, D.G. (eds.) Models of Information Processing in the Basal Ganglia, pp. 215–232. MIT Press, Cambridge (1995)

9. Bear, M.F., Connors, B.W., Pardiso, M.A. Neuroscience: Exploring the Brain. LippincottWilliams & Wilkins, Philadelphia (2001)

10. Bedard, C., Kroger, H., Destexhe, A. Model of low-pass filtering of local field potentials in brain tissue. Phys Rev E – Stat Nonlin Soft Matter Phys **73**(5), 051911 (2006)

11. Bedard, C., Kroger, H., Destexhe, A. Modeling extracellular field potentials and the frequency-filtering properties of extracellular space. Biophys J **86**, 1829–1842 (March 2004)

12. Bekar, L., Libionka, W., Tian, G., et al. Adenosine is crucial for deep brain stimulation mediated attenuation of tremor. Nat Med 14(1), 7580.s (2008)

13. Benabid, A.L., Pollak, P., Louveau, A., Henry, S., de Rougemont, J. Combined (thalamotomy and stimulation) stereotactic surgery of the VIM thalamic nucleus for bilateral Parkinson disease. Appl Neurophys **50**(16), 34–46 (1987)

14. Benabid, A.L. Deep brain stimulation for Parkinson's disease, Curr Opin Neurobiol, **13**, 696–706 (2003)

15. Berger, T.W. Implantable biomimetic microelectronics for the replacement of hippocampal memory function lost due to damage or disease. IEEE International Joint Conference on Neural Networks, Vol. 3, pt.3. p. 1659 (2004)

16. Bergman, H., Wichmann, T., Karmon, B., DeLong, M.R. The primate subthalamic nucleus. II. Neuronal activity in the MPTP model of Parkinsonism. J Neurophysiol **72**(2), 507–520 (Aug 1994)

17. Berns, G.S., Sejnowski, T.J. A computational model of how the basal ganglia produce sequences. J Cogn Neurosci **10**(1), 108–121 (1998)

18. Bia, A. ALOPEX-B: A new, simple but yet faster version of the ALOPEX training algorithm, Int J Neural Syst **11**(6), 497–507 (2001)

19. Bluma, A.L., Langley, P. Selection of relevant features and examples in machine learning. Artif Intell **97**, 245–271 (1997)

20. Brown, J.W., Bullock, D., Grossberg, S. How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. Neural Netw, **17**, 471–510 (2004)

21. Cavallo, T. An Essay on the Theory and Practice of Medical Electricity. Printed for the author, London (1780)

22. Chapin, J.K., Moxon, K.A., Markowitz, R.S., Nicolelis, M.A.L. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. Nat Neurosci, **2**(7) (July 1999)

23. Chardack, W., Gage, A., Greatbatch, W. A transistorized, self-contained, implantable pacemaker for the long-term correction of complete heart block. Surgen **48**, 543 (1960)

24. Clark, J., Plonsey, R. A mathematical evaluation of the core conductor model. Biophys J **6**, 95 (1966)

25. Coatrieux, J.L. Integrative science: Biosignal processing and modeling. IEEE Eng Med Biol Mag **23**(3), 9–12 (May–June 2004)

26. Cooley, J.W., Tukey, J.W. An algorithm for the machine calculation of complex Fourier series, Math Comput **19**, 297–301 (1965)

27. Coop, A.D., Reeke, G.N., Jr. The composite neuron: A realistic one-compartment Purkinje cell model suitable for large-scale neuronal network simulations. J Comput Neurosci **10**(2), 173–186 (2001)

28. Cotzias, G.C., VanWoert, M.H., Schiffer, L.M. Aromatic amino acids and modification of parkinsonism. N Engl J Med **276**, 374–379 (1967)

29. Dasey, T.J., Micheli-Tzanakou, E. Fuzzy neural networks. In: Micheli-Tzanakou, E. (ed.) Supervised and Unsupervised Pattern Recognition Feature Extraction and Computational Intelligence, pp. 135–162. CRC Press, LLC (2000)

30. De Forest, L. Device for amplifying feeble electrical currents, US Patent #841387 (1907)

31. DeLong, M.R. Primate models of movement disorders of basal ganglia origin, Trends Neurosci **13**, 281–285 (1990)

32. Dehaene, J., Moonen, M., Vandewalle, J. An improved stochastic gradient algorithm for principal component analysis and subspace tracking. IEEE Trans Signal Process **45**(10) (Oct 1997)

33. Delsarte, P., Genin, Y. On the splitting of the classical algorithms in linear prediction theory. IEEE Trans Acoust **ASSP-35**(5) (May 1987)

34. Djurfeldt, M., Ekeberg, O., Graybiel, A.M. Cortex-basal ganglia interaction and attractor states. Neurocomputing **3840**, 573–579 (2001)

35. Donoghue, J.P. Connecting cortex to machines: Recent advances in brain interfaces, Nat Neurosci Suppl **5** (Nov 2002)

36. Dostrovsky, J.O., Levy, R., Wu, J.P., Hutchison, W.D., Tasker, R.R., Lozano, A.M. Microstimulation-induced inhibition of neuronal firing in human globus pallidus. J Neurophys **84**, 570–574 (Jul 2000)

37. Du Bois-Reymond, E. Untersuchungen ber thierische elektricitt. G. Reimer, Berlin (1848)

38. Tzanakou, E. Principles and Design of the ALOPEX Device: A Novelmethod of Mapping Visual Receptive Fields, Ph.D. dissertation, Syracuse University, Department of Physics (1977)

39. Feng, X., Greenwald, B., Rabitz, H., Shea-Brown, E., Kosut, R. Toward closed-loop optimization of deep brain stimulation for Parkinson's disease: Concepts and lessons from a computational model. J Neural Eng **4**, L14–L21 (2007)

40. Fitch, M.T., Doller, C., Combs, C.K., Landreth, G.E., Silver, J. Cellular and molecular mechanisms of glial scarring and progressive cavitation: In vivo and in vitro analysis of inflammation-induced secondary injury after CNS trauma. J Neurosci **19**(19), 8182–8198, Oct 1 (1999)

41. Fitz Hugh, R. Impulses and physiological states in theoretical models of nerve membrane. Biophys J **1**(6), 445–466 (1961)

42. Friesen, G., Jannett, T., JadallahM., Yates, S., Quint, S., Nagle, H. A comparison of the noise sensitivity of nine QRS detection algorithms. IEEE Trans Biomed Eng **37**(1), 85–98 (Jan 1990)

43. Fukai, T. Modeling the interplay of short-term memory and the basal ganglia in sequence processing. Neurocomputing **26–27**, 687–692 (1999)

44. Fukai, T. Sequence generation in arbitrary temporal patterns from theta-nested gamma oscillations: A model of the basal ganglia-thalamo-cortical loops. Neural Netw **12**(7–8), 975–987 (1999)

45. Gabriel, S., Lau, R.W., Gabriel, C. The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues. Phys Med Biol **41**, 2271–2293 (1996)

46. Gao, Y., Blacky, M.J., Bienenstock, E., Shoham, S., Donoghue, J.P. Probabilistic inference of hand motion from neural activity in motor cortex. Adv Neural Inf Process Syst **14**, 213–220 (2002)

47. Gasser, H.S., Erlanger, J. A study of the action currents of the nerve with the cathode ray oscillograph. Am J Physiol **62**, 496–524 (1922)

48. Georgopoulos, A., Schwartz, A., Kettner, R. Neural population coding of movement direction. Science **233**, 1416–1419 (1986)

49. Georgopoulos, A.P., Lurito, J.T., Petrides, M., Schwartz, A.B., Massey, J.T. Mental rotation of the neuronal population vector. Science, **243**, 234–236 (1989)

50. Gerosa, A., Maniero, A., Neviani, A. A fully integrated dual-channel logdomain programmable preamplifier and filter for an implantable cardiac pacemaker, IEEE Transactions on Circuits and Systems. I: Regular Papers, **51**(10) (Oct 2004)

51. Gerosa, A., Maniero, A., Neviani, A. A fully integrated two-channel a/d interface for the acquisition of cardiac signals in implantable pacemakers. IEEE J Solid-State Circuits **39**(7), July (2004)

52. Ghovanloo, M., Najafi, K. A Modular 32-site wireless neural stimulation microsystem. IEEE J Solid-State Circuits **39**(12), 2457–2466 (2004)

53. Gillies, A., Arbuthnott, G. Computational models of the basal ganglia, Mov Dis **15**(5), 762–770 (2000)

54. Grill, W.M., Snyder, A.N., Miocinovic, S. Deep brain stimulation creates an informational lesion of the stimulated nucleus. Neuroreport **15**(7), 1137–1140, May 19 (2004)

55. Guridi, J., Lozano, A.M. A brief history of pallidotomy. Neurosurgery **41**(5), 1169–1180 (1997)

56. Gurney, K., Prescott, T.J., Redgrave, P. A computational model of action selection in the basal ganglia. I. A new functional anatomy. Biol Cybern **84**(6), 401–410 (2001)

57. Gurney, K., Prescott, T.J., Redgrave, P. A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. Biol Cybern **84**(6), 411–423 (2001)

58. Hamilton, J. Analysis of physiological function in the globus pallidus with neural networks, (PhD-MD) Jan (2000)

59. Hamilton, J.L., Micheli-Tzanakou, E., Lehman, R. Analysis of electrophysiological data in surgical treatment for Parkinson's disease. Proceedings of the 24th IEEE Northeast Conference on Bioengineering, pp. 5–6 (1998)

60. Hamilton, J.L., Micheli-Tzanakou, E., Lehman, R.M. Neural networks trained with simulation data for outcome prediction in pallidotomy for Parkinson's disease. IEEE Eng Med Biol Soc Conf **1**, 1–4 (2000)

61. Hanson, F.B., Tuckwell, H.C. Diffusion approximation for neuronal activity including reversal potentials. J Theor Neurobiol **2**, 127–153 (1983)

62. Harth, E., Tzanakou, E. ALOPEX: A stochastic method for determining visual receptive fields, Vision Research, Vol. 14, pp.1475–1482, (1974)

63. Haykin, S., Chen, Z., Becker, S. Stochastic correlative learning algorithms. IEEE Trans Signal Process, **52**(8) (Aug 2004)

64. Hindmarsh, J.L., Rose, R.M. A model of neuronal bursting using three coupled first order differential equations. Proc R Soc Lond B: Biol Sci **221**(1222), 87–102 March (1984)

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

66. Holmstrom, L., Koistinen, P., Laaksonen, J., Oja, E. Neural and statistical classifiers-taxonomy and two case studies. IEEE Trans Neural Netw **8**(1) Jan (1997)

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

68. House, W.F. Cochlear implants: My perspective, Dr. William F. House, 1995.

69. Humayun, M.S., de Juan E., Jr., Dagnelie, G., Greenberg, R.J., Propst, R.H., Phillips, D.H. Visual perception elicited by electrical stimulation of retina in blind humans. Arch Ophthalmol **114**(1) (1996)

70. Humphries, M.D., Gurney, K.N. A pulsed neural network model of bursting in the basal ganglia. Neural Netw **14**(6–7), 845–863 (2001)

71. Hurtado, J.M., Gray, C.M., Tamas, L.B., Sigvardt, K.A. Dynamics of tremor-related oscillations in the human globus pallidus: A single case study. Proc Natl Acad Sci USA **96**, 1674–1679 Feb (1999)

72. Iasemidis LD, Shiau DS, Pardalos PM, Chaovalitwongse, W., Narayanan, K., Prasad, A., Tsakalis, K., Carney, P.R., Sackellares, J.C. Long-term prospective online real-time seizure prediction. Clin Neurophys **116**(3), 532–544 (2005)

73. Izhikevich, E.M. Simple model of spiking neurons. IEEE Trans Neural Netw **14**, 1569–1572 Nov (2003)

74. Izhikevich, E.M. Which model to use for cortical spiking neurons? IEEE Trans Neural Netw **15**, 1063–1070 (2004)

75. Izhikevich, E.M. Which model to use for cortical spiking neurons. IEEE Transactions on Neural Netw **15**(5) (2004)

76. Jallabert, J. Experiences sur lectricit, Geneve, Barrillot & Fils (1748)

77. Joel, D., Niv, Y., Ruppin, E. Actor-critic models of the basal ganglia: New anatomical and computational perspectives. Neural Netw **15**, 535–547 (2002)

78. John, M.S. Adaptive brain stimulation method and system, US Patent#6463328, (2002)

79. Johnson, M.D., Otto, K.J., Kipke, D.R. Repeated voltage biasing improves unit recordings by reducing resistive tissue impedances. IEEE Trans Neural Syst Rehabil Eng **13**(2), 160–165 (2005)

80. Kaneko, H., Suzuki, S.S., Okada, J., Akamatsu, M. Multineuronal spike classification based on multisite electrode recording, whole-waveform analysis, and hierarchical clustering. IEEE Trans Biomed Eng **46**(3), March (1999)

81. Katz, B., Miledi, R. The measurement of synaptic delay, and the time course of acetylcholine release at the neuromuscular junction. Proc R Soc Lond, B Biol Sci **161**(985), 483–495 (Feb 16, 1965)

82. Kemere, C., Shenoy, K.V., Meng, T.H. Model-based neural decoding of reaching movements: A maximum likelihood approach. IEEE Trans Biomed Eng **51**(6) (Jun 2004)

83. Kilby, J.S. Miniaturized electronic circuits, US Patent #3138743, 1964.

84. Kimber, T.E., Tsai, C.S., Semmler, J., Brophy, B.P., Thompson, P.D. Voluntary movement after pallidotomy in severe Parkinson's disease, Brain **122**, 895–906 (1999)

85. Kite, C. An Essay on the Recovery of the Apparently Dead. C. Dilly, London (1788)

86. Kohler, B.U., Hennig, C., Orglmeister, R. The principles of QRS detection. IEEE Eng Med Biol Mag **21**(1), 42–57, Jan–Feb (2002)

87. Konstandinides, K., Tyree, V.C., Yao, K. Single chip implementation of the Levinson algorithm. IEEE J Solid-State Circuits **SC-20**(5) (Oct 1985)

88. Kulkarni, S.R., Lugosi, G., Venkatesh, S.S. Learning pattern classification—a survey. IEEE Trans Inform Theory **44**(6) (1998)

89. Kumar, R. Methods for programming and patient management with deep brain stimulation of the globus pallidus for the treatment of advanced Parkinson's disease and dystonia. Mov Dis **17**(3), S198–S207 (2002)

90. Lee, J., Rhew, H., Kipke, D., Flynn, M. A 64 channel programmable closed-loop deep brain stimulator with 8 channel neural amplifier and logarithmic ADC, 2008 Symposium on VLSI Circuits Digest of Technical Papers, pp.76–77 (2008)

91. Lehman, R.M., Micheli-Tzanakou, E., Medl, A., Hamilton, J.L. Quantitative online analysis of physiological data for lesion placement in pallidotomy. Stereotact Funct Neurosurg **75**(1), 1–15 (2000)

92. Lenz, F.A., Kwan, H.C., Martin, R.L., Tasker, R.R., Dostrovsky, J.O., Lenz, Y.E. Single unit analysis of the human ventral thalamic nuclear group. Tremor related activity in functionally identified cells. Brain **117**(3), 531–543 (1994)

93. Liao, H., Mandal, M.K., Cockburn, B.F. Efficient architectures for 1D and 2D lifting-based wavelet transforms. IEEE Trans Signal Process **52**(5) (May 2004)

94. Licht, S. Therapeutic Electricity and Ultraviolet Radiation. New Haven, E. Licht (1967)

95. Malmivuo, J., Plonsey, R. Bioelectromagnetism, Principles and Applications of Bioelectric and Biomagnetic Fields. Oxford University Press, New York (1995)

96. Manolakos, E.S., Stellakis, H.M. Systematic synthesis of parallel architectures for the computation of higher order cumulants. Parall Comput **26**, 655–676 (2000)

97. Matteuci, C. Sur un phenomene physiologique produit par les muscles en contracion, Annales de Chimie et de Physique, 6(339) (1842)

98. McCulloch, W.S., Pitts, W.H. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys **5**, 115–133 (1943)

99. McIntyre, C.C., Grill, W.M. Extracellular stimulation of central neurons: Influence of stimulus waveform and frequency on neuronal output. J Neurophysiol **88**, 1592–1604 (2002)

100. McIntyre, C.C., Grill, W.M. Excitation of central nervous system neurons by nonuniform electric fields. Biophys J **76**, 878–888 (Feb 1999)

101. McIntyre, C.C., Thakor, N.V. Uncovering the mechanisms of deep brain stimulation for Parkinson's disease through functional imaging, neural recording and neural modeling. Crit Rev Biomed Eng **30**(4–6), 249–281 (2002)

102. Medtronic Corporation, Extension kit for deep brain stimulation, spinal cord stimulation, or peripheral nerve stimulation, implant manual, Medtronic, Inc. (2002)

103. Meissimilly, G., Rodriguez, J., Rodriguez, G., Gonzalez, R., Canizares, M. Microcontroller-based real-time QRS detector for ambulatory monitoring. Proc IEEE Eng Med Biol Soc **3**, 17–21 (2003)

104. Micheli-Tzanakou, E., Hamilton, J., Zheng, J., Lehman, R. Computational intelligence for target assessment in Parkinson's disease. In: Bosacchi, B., Fogel, D.B., Bezdek, J.C. (eds.) Proceedings of the SPIE, Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation IV, Vol. 4479, pp. 54–69. SPIE-Medical Imaging (2001)

105. Micheli-Tzanakou, E., Michalak, R., Harth, E. The Alopex process: Visual receptive fields with response feedback. Biol Cybern **35**, 161–174 (1979)
106. Montgomery, E.B., Jr., Baker, K.B. Mechanisms of deep brain stimulation and future technical developments. Neurol Res **22**, 259–266 (2000)
107. Moore, G.E. Cramming more components onto integrated circuits. Electronics **38**(8) (1965)
108. Moxon, K.A., Kalkhoran, N.M., Markert, M., Sambito, M.A., McKenzie, J.L., Webster, J.T. Nanostructured surface modification of ceramic-based microelectrodes to enhance biocompatibility for a direct brain-machine interface. IEEE Trans Biomed Eng **51**(6) (June 2004)
109. Mraz, S.J. Rewiring the retina. Machine Design **75**(13), 60–64 (Jul 10, 2003)
110. Muenter, M.D., Tyce, G.M. l-dopa therapy of Parkinson's disease: Plasma l-dopa concentration, therapeutic response, and side effects. Mayo Clin Proc **46**, 231–239 (1971)
111. Nicolelis, M.A.L. Brain machine interfaces to restore motor function and probe neural circuits. Nat Rev Neurosci **4**(5), 417–422 (2003)
112. Noble, D. A modification of the Hodgkin-Huxley equations applicable to purkinje fibre action and pacemaker potentials. J Physiol **160**, 317–352 (1962)
113. Novo, A., Gerosa, A., Neviani, A. A sub-micron CMOS programmable charge pump for implantable pacemaker. Analog Integrated Circuits Signal Process **27**, 211–217 (2001)
114. Noyce, R.N. Semiconductor device-and-lead structure, US Patent # 2981877, 1961.
115. Nutt, J., Anderson, V.C., Peacock, J.H., Hammerstad, J.P., Burchiel, K.J. DBS and diathermy interaction induces severe CNS damage. Neurology **56**, 1384–1386 (2001)
116. Pardo, M., Sberveglieri, G. Learning from data: A tutorial with emphasis on modern pattern recognition methods. IEEE Sens J **2**(3), 203–217 (2002)
117. Patterson, W.R., Song, Y., Bull, C.W., Ozden, I., Deangellis, A.P., Lay, C., McKay, J.L., Nurmikko, A.V., Donoghue, J.D., Connors, B.W. A microelectrode/ microelectronic hybrid device for brain implantable neuroprosthesis applications. IEEE Trans Biomed Eng **51**(10) (Oct 2004)
118. Peng, H., Long, F., Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell **27**(8), 1226 (2005)
119. Pea, C., Bowsher, K., Costello, A., De Luca, R., Doll, S., Li, K., Schroeder, M., Stevens, T. An overview of FDA medical device regulation as it relates to deep brain stimulation devices. IEEE Trans Neural Syst Rehabil Eng **15**(3), 421–424 (2007)
120. Plenz, D., Kitai, S. A basal ganglia pacemaker formed by the subthalamic nucleus and external globus pallidus. Nature **400**, 677–682 (1999)
121. Priestley, J. The history and present state of electricity: With original experiments. Printed for C. Bathurst, and T. Lowndes . . . J. Rivington, and J. Johnson . . . S. Crowder, G. Robinson, and R. Baldwin . . . T. Becket, and T. Cadell . . ., London, MDCCLXXV (1775)
122. Pudil, P., Novovicova, J., Somol, P. Feature selection toolbox software package. Pattern Recognit Lett **23**, 487–492 (2002)
123. Eckmiller, R., Eckhorn, R. Final report of the feasibility study for a neurotechnology program, NeuroTechnology Report, BMBF, Bonn, Germany (1994)
124. Rajasekaran, S. Efficient parallel algorithms for template matching. Parallel Process Lett **12**(3–4), 359–364 (2002)
125. Rall, W., Burke, R.E., Holmes, W.R., Jack, J.J., Redman, S.J., Segev, I. Matching dendritic neuron models to experimental data. Physiol Rev **72**(4) (Suppl), S159–S86 (1992)
126. Raz, A., Vaadia, E., Bergman, H. Firing patterns and correlations of spontaneous discharge of pallidal neurons in the normal and the tremulous 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine vervet model of parkinsonism. J Neurosci **20**(22), 8559–8571 (Nov 15, 2000)
127. Rise, M.T., King, G.W. Method of treating movement disorders by brain stimulation, US Patent #5716377 (1998)
128. Rodrigues, J.N., Owall, V., Sornmo, L. A wavelet based R-wave detector for cardiac pacemakers in 0.35 CMOS technology, IEEE Circuits Syst Proc (ISCAS) **4**, 23–26 (2004)
129. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev **65**(6), 386–408 (1958)

130. Rossi, L., Marceglia, S., Foffani, G., Cogiamanian, F., Tamma, F., Rampini, P., Barbieri, S., Bracchi, F., Priori, A. Subthalamic local field potential oscillations during ongoing deep brain stimulation in Parkinson's disease. Brain Res Bull **76**(5), 512–521 (2008)

131. Ruggera, P.S., Witters, D.M., Maltzahn, G., Bassen, H.I. In vitro assessment of tissue heating near metallic medical implants by exposure to pulsed radio frequency diathermy. Phys Med Biol **48**, 2919–2928 (2003)

132. Sanchez, J.C., Sung-Phil, K., Erdogmus, D., Rao, Y.N., Principe, J.C., Wessberg, J., Nicolelis, M. Input-output mapping performance of linear and nonlinear models for estimating hand trajectories from cortical neuronal firing patterns. Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing, 4–6, 139–148 (Sep. 2002)

133. Sanders, R.S., Lee, M.T. Implantable pacemakers. Proc IEEE **84**(3), 480–486 (March 1996)

134. Sanghavi, T. Design of an alopex architecture and its application to adaptive deep brain stimulation (ADBS), Rutgers theses. Graduate Program in Electrical and Computer Engineering (2005)

135. Sastry, P.S., Magesh, M., Unnikrishnan, K.P. Two timescale analysis of the ALOPEX algorithm for optimization. Neural Comput, 14, 2729–2750 (2002)

136. Scharstein, H. Input–output relationship of the leaky-integrator neuronmodel. J Math Biol **8**(4), 403–420 (1979)

137. Schueler, B.A., Parrish, T.B., Lin, J., Hammer, B.E., Pangrle, B.J., Ritenour, E.R., Kucharczyk, J., Truwit, C.L. MRI compatibility and visibility assessment of implantable medical devices. J Mag Res Imaging **9**, 596–603 (1999)

138. Schweigger, J.S.C. Zustze zu ersteds elektromagnetischen versuchen, vorgelesen in der naturforschenden. Journal fr Chemie und Physik, Schweigger Ed. **31**(1–17), 431 (1821)

139. Scribonius, L. Compositiones. In: Sergio Sconocchia (ed.) Scribonii Largi Compositiones. Teubner, Leipzig (1983)

140. Sejnowski, T.J. The book of Hebb. Neuron **24**, 773–776 (1999)

141. Shinomoto, S., Kuramoto, Y. Phase transitions in active rotator systems. Prog Theor Phys **75**, 1105–1110 (1986)

142. Sridhar, A. Analog CMOS Model of Parkinson's Disease, Thesis (M.S.), Rutgers University (2005)

143. Stein, R.B. A theoretical analysis of neuronal variability. Biophys J **5**, 173–194 (1965)

144. Struijk, J.J. Thomsen, M., Larsen, J.O. Sinkjaer, T. Cuff electrodes for longterm recording of natural sensory information. IEEE Eng Med Biol Mag **18**(3), pp. 91–98 (1999)

145. Suri, R.E. Albani, C., Glattfelder, A.H. A dynamic model of motor basal ganglia functions. Biol Cybern **76**(6), 451–458 (1997)

146. Tass, P.A. Amodel of desynchronizing deep brain stimulation with a demand controlled coordinated reset of neural subpopulations. Biol Cybern **89**(2), 81–88 (August 2003)

147. Tass, P.A. Effective desynchronization with bipolar double-pulse stimulation. Phys Rev E **66**, 036226 (2002)

148. Taylor JG, Taylor NR. Analysis of recurrent cortico-basal ganglia-thalamic loops for working memory. Biol Cybern **82**(5), 415–432 (2000)

149. Terman, D., Rubin, J.E., Yew, A.C., Wilson, C.J. Activity patterns in a model for the Subthalamopallidal Network of the Basal Ganglia. J Neurosci **22**, 2963–2976 (2002)

150. Tzanakou, E., Harth, E. Determination of visual receptive fields by stochastic methods. Biophys J **15**, 42a (1973)

151. U.S. Department of Health and Human Services. FDA approves implanted brain stimulator to control tremor, Press Release P97–24 (August 4, 1997)

152. Volkmann, J., Herzog, J., Kopper, F., Deuschl, G. Introduction to the programming of deep brain stimulators. Mov Dis **17**(Suppl. 3), S181–S187 (2002)

153. Volta, A. On the electricity excited by the mere contact of conducting substances of different kinds. Philosophical Trans **90**, part 2, 403–431, with one folding engraved plate, numbered XVII, (1800)

154. Warman, E.N., Grill, W.M., Durand, D. Modeling the effects of electric fields on nerve fibers: Determination of excitation thresholds. IEEE Trans Biomed Eng, **39**(12) (Dec. 1992)

155. Wessberg, J., Stambaugh, C.R., Kralik, J.D., Beck, P.D., Laubach, M., Chapin, J.K., Kim, J., Biggs, S.J., Srinivasan, M.A., Nicolelis, M.A.L. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. Nature **408**, 361–365 (November 16, 2000)
156. Wichman, T., DeLong, M.R. Pathophysiology of Parkinsons disease: The MPTP primate model of the human disorder. Ann NY Acad Sci **991**, 199–213 (2003)
157. Wilson, H.R. Simplified dynamics of human and mammalian neocortical neurons. J Theor Biol **200**, 375–388 (1999)
158. Wingeier, B., Tcheng, T., Koop, M.M., Hill, B.C., Heit, G., Bronte-Stewart, H.M. Intra-operative STN DBS attenuates the prominent beta rhythm in the STN in Parkinson's disease. Exp Neurol **197**, 244–251 (2006)
159. Wise, K.D., Anderson, D.J., Hetke, J.F., Kipke, D.R., Najafi, N. Wireless implantable microsystems: High-density electronic interfaces to the nervous system. Proc IEEE **92**(1) (January 2004)
160. Wolpert, S., Micheli-Tzanakou, E. A neuromime in VLSI. IEEE Trans Neural Netw **7**(2) (March 1996)
161. Wu, W., Black, M.J., Gao, Y., Bienenstock, E., Serruya, M., Shaikhouni, A., Donoghue, J.P. Neural decoding of cursor motion using Kalman filtering. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems, Vol. 15, pp.117–124. MIT Press, Cambridge (2003)
162. Wu, Y.R., Levy, R., Ashby, P., Tasker, R.R., Dostrovsky, J.O. Does stimulation of the GPI control dyskinesia by activating inhibitory axons? Mov Dis **16**(2), 208–216 (2001)

Chapter 14
Molecule-Inspired Methods for Coarse-Grain Multi-System Optimization

Max H. Garzon and Andrew J. Neel

Abstract A major goal in multi-objective optimization is to strike a compromise among various objective functions subject to diverse sets of conflicting constraints. It is a reality, however, that we must face optimization of entire systems in which multiple objective sets make it practically impossible to even formulate objective functions and constraints in the standard closed form. We present a new approach techniques inspired by biomolecular interactions such as embodied in DNA. The advantages are more comprehensive and integrated understanding of complex chains of local interactions that affect an entire system, such as the chemical interaction of biomolecules in vitro, a living cell, or a mammalian brain, even if done in simulation. We briefly describe a system of this type, EdnaCo (a high-fidelity simulation in silico of chemical reactions in a test tube in vitro), that can be used to understand systems such as living cells and large neuronal assemblies. With large-scale applications of this prototype in sight, we propose three basic optimization principles critical to the successful development of robust synthetic models of these complex systems: physical–chemical, computational, and biological optimization. We conclude with evidence for and discussion of the emerging hypothesis that multi-system optimization problems can indeed be solved, at least approximately, by so-called coarsely optimal models of the type discussed above, in the context of a biomolecule-based asynchronous model of the human brain.

M.H. Garzon
Department of Computer Science, The University of Memphis, Memphis, TN 38152-3240, USA,
e-mail: mgarzon@memphis.edu

A.J. Neel
Department of Computer Science, The University of Memphis, Memphis, TN 38152-3240, USA,
e-mail: aneel@memphis.edu

14.1 Introduction

Optimization is an important branch in many scientific domains, particularly in mathematics, physics, and biology. Several branches of these sciences have developed a great variety of methods to solve optimization problems, including original calculus and mathematical analysis (for numerical functions), calculus of variations (for functionals in a function space), and energy minimization (for physical problems.) These developments have added a great deal of understanding and power to the wide range of optimization techniques available now. More recently, the computational complexity (NP-hardness) of important and difficult problems, such as the many variations of TSP (traveling salesperson problems) [10], have given rise to search-based techniques such as genetic algorithms [20] and genetic programming [22]. The recent success stories of these techniques, even at the level of multi-objective optimization [9], flesh out the optimizing power of randomization.

At the larger scale of entire systems, however, the picture is radically different. Despite the fact that physical and biological systems exhibit function and operation at optimal tuning of their components and through minimal usage of energy and time, optimization techniques have been underdeveloped at best. Even physical problems such as the n-body problem (i.e., finding the stable equilibrium of a number of masses endowed with initial position and momenta and subject to Newtonian gravitation) remains a difficult and unresolved problem [37]. In biology, the principle of natural selection can be construed as an optimization technique, both in terms of feedback communication from environments to organic genomes and as an overall optimization of the adaptation of a species to a given environment. The underlying mechanisms of this principle, in the form of crossover and mutation operations, have inspired the computational methods of genetic algorithms mentioned above. Few other general principles that afford deep understanding of the operation of complex physical or biological systems, such as brains, or even the optimization of some man-made systems, have been unveiled. Of these, most make assumptions about the system, equilibrium, for example, which usually fail to hold in most systems of practical interest, especially in biology, society, and economics, where optimization becomes perennial fitness changes, or "optimal energy flow" [2, p. 534], or even intelligence [25].

Nonetheless, system optimization is not only meaningful, but also required in many contexts, varying from biological entities (cells, organs, organisms) through all degrees of complexity in larger systems. The main stumbling block with complex system optimization is twofold: (a) there does not seem to exist a natural definition of optimality, let alone a formal specification in the standard terms of mathematical programming or standard methods in operations research; and (b) there is no universally accepted scientific metric even to quantify the efficiency of a complex system, even in business/political contexts where profit or military power might be strong candidates. For example, Daly [6] discusses optimality of nations as systems. The result is that system optimization remains an art and its methods are generally developed for or applicable only to specialized niches.

In this chapter, we describe several recent results that illustrate the power of simulation as the basis of an optimization technique for design of complex systems. In Section 14.2, we describe the prototype phenomenon being used, DNA and their molecular interactions, the theme of study in the new field of biomolecular computing (BMC), also known as DNA computing [17, 1]. In Section 14.3, we describe a high-fidelity simulation of this phenomenon that has produced comparable results, using in a fundamental manner once again the resource of massive randomization, here provided by RNGs (random number generators.) In Section 14.4, we show how recent developments in BMC provide the basic ingredients to implement large neuronal systems with complex interactions among them. Finally, in Section 14.5 we discuss some of the challenges in the implementation of the design and discuss some of the possible applications of the model. We assume that the reader is familiar with basic facts about molecular biology – see one of several surveys of the field (for example, [36]) for background details. We also assume that the reader is familiar with artificial neural networks [19].

14.2 Biomolecular Computing In Vitro

For several millions of years, DNA has demonstrated itself capable of reliably storing instructions for the makeup of living organisms. Only recently, however, did a more formal inquiry of DNA begin for its own sake, as an entity of its own, pioneered by [1], where the first successful demonstration of DNA's potential use for nonbiological purposes, more specifically, a solution to the Hamiltonian path problem (HPP), was demonstrated. HPP is a computational equivalent of the well-known traveling salesman problem (TSP). An instance of HPP is a directed graph (vertices and edges) with two singled out as source and destination vertices. The problem calls for a Boolean decision whether there exists a Hamiltonian path joining the source to the destination (i.e., a path passing through every vertex exactly once). Adleman [1] reduces the problem to 10^{12} recently available biotechnology by mapping vertices and edges to DNA oligonucleotides with vertices designed to partially stick to edges so that molecules representing paths in the graph would form by ligation/ concatenation of smaller edges or chains into longer DNA oligonucleotides. One such chain of the appropriate length and composition would witness a positive answer to a Hamiltonian graph. Since DNA oligonucleotides can be synthesized in lengths up to 200 base pairs at low cost for picomoles of the same species (about copies of a given molecule), edges and vertices are present in several millions of copies so that the protocol would fully explore all possible solutions. Extracting the nanoscopic answer has been made possible by the extraordinary advances witnessed in biotechnology in the course of the last two decades. The seemingly unlimited scalability of this approach to solve these difficult NP-hard problems of large sizes gave rise to the field of DNA computing, also called BMC (biomolecular computing). In the last decade, researchers in this field have emerged with DNA computers capable of solving simple games like TIC-TAC-TOE [24] or

implementing programmable and autonomous computing machines with biomolecules [39, 3, 32, 38].

14.3 Biomolecular Computing In Silico

Inspired by Adleman's approach and the relatively high cost of molecular protocols, we have developed a computational environment to reproduce in simulation essentially equivalent chemistry in silico [13, 12]. Software of this type, called virtual test tube (VTT) EdnaCo, was developed to understand biochemical reactions with DNA oligonucleotides for computational purposes [15]. We next provide a high-level description of the software involved in the simulation. Further details can be found in [13]

EdnaCo follows the complex systems paradigm [2] of entities (objects) and interactions, i.e., instead of programming their entire behavior over time; only entities (originally DNA molecules) and individual interactions between pairs of them are programmed by the user. Conceptually, the VTT is spatially arranged as a 3D coordinate system in which molecules can move about. The tube moves molecules by simulating three different types of motion: Brownian, according to a predetermined schedule, or no motion at all. The entities are allowed to interact freely in a predetermined manner specified by the experimenter. Entities could be homogenous (all of the same type) or heterogenous (different types) and may represent any complex biomolecules, herein referred to as DNA complexes. Each molecule is located at a unique spatial coordinate at any given time. The VTT can also code for physical–chemical properties such as temperature, pressure, salinity, and pH that may vary across space and time and affect the way structures interact therein. Interactions between entities are programmed by the experimenter depending on the nature of the molecules being simulated. Multiple instances of an entity behave in the same manner as a function of environmental factors such as temperature, pH, and the like.

All entities are capable of sensing the position of other entities up to a specified distance defined by a radius of interaction, a parameter in the simulation common to all entities. If two or more entities come within the interaction distance, an encounter is triggered between them. An encounter is resolved by appropriate software that may not affect the molecules at all (e.g., DNA molecules may not form a duplex if the temperature is too high) or may reflect an appropriate chemical reaction (e.g., formation of a DNA duplex and disappearance of the encountering single strands.) An interaction between two entities may be viewed as a chemical or mechanical reaction between them. As a result of an interaction, existing entities may get consumed, their status may get changed, and/or new entities may, or may not, get created. Moreover, the concentration of entities may be manipulated externally by adding or removing entities to or from the VTT at any point of time. The running time of a simulation is divided into discrete time steps or iterations. At every iteration, the state of the objects and the tube may change recursively, based on the

current state resulting from previous changes, in order to reflect the interaction rules of the objects themselves and/or with their environment.

In the actual implementation of this conceptual model, EdnaCo is implemented by dividing the computer's memory into a number of discrete segments, each running on a different processor. This allows multiple interactions to take place at once. When entities move, they either change positions within a segment, or they may migrate across processor boundaries. Thus, the container of the VTT is a discrete 3D space residing inside a digital computer, the entities are instantiated as objects in classes in a programming language (C++), and the interactions are appropriate functions and methods associated with these objects. These methods either leave the objects unperturbed or make the appropriate deletions and insertions to reflect the reactions they simulate. The communication between processors is implemented using a message-passing interface, such as MPI, on a cluster of personal computers or a high-performance cluster (in our case, an IBM cluster containing 112 dual processors). The architecture of EdnaCo allows it to be scaled to an arbitrarily large number of processors and to be portable to any other cluster supporting C++ and MPI. The results of these simulations are thus guaranteed to be reproducible on any other systems running the same software as the original simulation. Further details of this simulation environment can be found in [11].

The striking property of the VTT is that the programming stops at the level of local interactions. Nothing else is programmed, every other observable is an emergent property of the simulation. In particular, if the equivalent molecules in Adleman's original experiment are placed in the tube, they will seem to be moved about randomly by Brownian motion. Nevertheless, encounter between vertex and edge molecules may create longer and longer paths, and eventually a Hamiltonian one if one is possible. This is indeed the case with a very high probability (reliability of 99.6% with no false negatives has been reported [12]). In other words, the solution to an optimization problem has been captured in silico by a simple simulation of the most relevant properties of the natural phenomena occurring inside a test tube containing DNA oligonucleotides with the appropriate characteristics. The scalability of this solution is clear in two different directions. First, the parallelism of EdnaCo provides parallelism of the type inherent in chemistry. Second, the size of problems solvable by this method is only limited by our ability to find sets of oligonucleotides large enough to code for the vertices and edges without causing any undesirable interactions between them. Although it was initially suggested that random encodings would provide sufficient stock of oligonucleotides [1], further work had determined that care must be exercised in selecting oligonucleotides that are inert to cross-hybridization due to the uncertain and thermodynamic nature of DNA hybridization. Although this problem has proven to be in itself NP-complete [29], similar methods can be used to provide nearly optimal solutions, both in vivo by the PCR selection protocol of [5, 4, 7], based on an extension of the polymerase chain reaction, and in silico by its simulation [16]. So-called DNA code sets, i.e., noncross-hybridizing (nxh) molecules are now readily available to solve large problems systematically, of the order of tens to hundreds of thousands of noncrosshybriding 20-mers, for example, as seen in Fig. 14.1 [16, 14].

Size of Code Sets
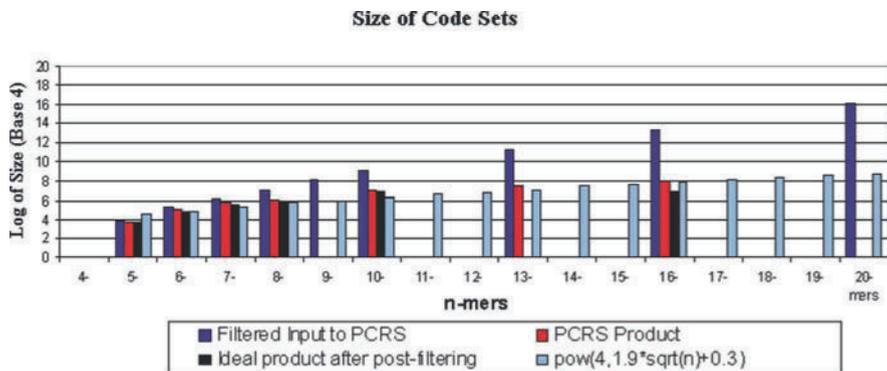
Fig. 14.1: Size of the PCR selection product obtained from the filtered input sets (*left bars*) based on the polymerase chain reaction. The size of the resulting set (*middle bars*) can be estimated by the subpower law $(1.9n + 0.6)^4$ (*light bars on the right*) [16].

14.4 Neural Nets in Biomolecules

In this section we use the results described in the previous sections to show how to implement large sets of neuronal ensembles in DNA molecules. Although we will use artificial neural nets as the prototype of such ensembles, we note that similar techniques can be used to implement higher order structures of the type exemplified by more complex ensembles known as brains (more in the discussion in the next section). For a review of neural networks, the reader is referred to [19]. Hopfield networks [21] will be used as a prototype to describe the design, but the techniques readily extend to other types of neural nets and associative memories [19, 18]. We will also omit technical details of the biotechnical implementation in a wet lab and concentrate only on the simulation in silico.

Turberfield et al. [35] describe the first attempt to implement a specific type of neural network in DNA, specifically, recurrent Hopfield nets. Their approach showed the experimental possibility of using DNA as an almost direct replacement of neurons with the hybridization affinity as the critical component to capture synaptic connections for associative retrieval. The model required the use of restriction enzymes and high-quality nxh DNA codewords (unavailable at the time). The experimental implementation proved to have low reliability and speed in activation updates of the neuronal units due to the use of restriction enzymes in many cycles of PCR amplification [27] and the uncertainty of hybridization reactions, which we can a posteriori recognize as due to cross-hybridization. Moreover, their work remained untested for scalability since implementing large networks required the use of large sets of nxh codewords, which were not readily available at the time. However, this attempt provides a very good description and test of the experimental requirements for their implementation, and much of it can be used in an appropriately updated form in this implementation.

Our model addresses most of these problems because of the modifications described next and the inherent high quality of the code sets now in our possession. A neuronal unit i is represented by a codeword Ci from a DNA code; these codewords only cross-hybridize to their own complements Di, not to any other codewords or their complements. These single strands can be permanently affixed to a solid medium (such as those used for DNA microarrays) so as to form a DNA chip A, by what is now standard biotechnology [8,26]. The chip A contains two spots for each neuron i (one spot for positive activation and one spot for negative activation, located far apart enough to avoid cross-hybridization), each with as many single strands (Ci or Di) as necessary for a given resolution on the activation levels (for example, three decimal digits will require $M = 1,000$ copies of the corresponding oligonucleotide attached at each spot). A positive activation level of the unit i is given at any time t by the concentration of the particular double stranded DNA species Ci–Di, whereas a negative activation is given by the concentration of its complementary labeled word Di–Ci. An optical census of double-stranded DNA (or their Watson–Crick complements) can be taken on this chip by using fluorescent tags (e.g., SYBR green attached to the double-stranded pair Ci–Di) [8, 36] that will reveal the current activation levels of the various units $x_i(t)$ at a given time t, if normalized to M. We will denote by m the length of the m-mer oligonucleotides Ci in the code set. Note that a complementary copy of the activation vector $x(t)$ at a given time t can be made by simply heating the chip to an appropriate temperature exceeding the maximum melting temperature of all pairs Ci–Di, then washing the complementary single-stranded representation x' into a temporary tube T.

The transition from an activation state $x(t)$ to another state $x(t+1)$ stipulated by the Hopfield model requires longer strands representing the synaptic weights Wij from neuron j into neuron i to be attached to the oligonucleotide Ci that represents it. For this design, we will assume that the weights are integer valued for simplicity (similar designs could be used for rational values to approximate any other values). The weights are themselves copies of the complementary DNA oligomers Di for neuron i extended by $-Wi$– copies of Di separated by a restriction site r, so they can be eventually separated into their original pieces. For example, a weight $Wij = 3$ would be expressed in DNA as $Wij = circjrcj\ rcj$. Zero weights are not represented, i.e., the absence of any molecular representation containing Ci and Cj means that $Wij = 0$. These weights will likewise be permanently affixed to another chip W in a similar manner to chip A. (We will use the same symbol Wij to denote the weight and its molecular representation for simplicity. The context makes clear which one is being referred to.)

A Hopfield network transitions from total activation vector $x(t)$ to activation vector $x(t+1) = s(Wx(t))$, where s is a saturation function (typically a sigmoid, but here assumed to be just a linear approximation as the identity map squashed to 0 for negative values and 1 otherwise) and $Wx(t)$ is the product of matrix $W = [Wij]$ and activation vector $x(t)$, i.e., each unit i computes its weighted net input by the scalar product of row i and column $x(t)$ and saturates it using s (i.e., activation values outside a unit's range are squashed down within the activation range). The matrix product is here realized by a two-step process, first make a complementary

copy of current activation x' into a tube T, pour it over the weight chip W and allow enough time and reaction conditions for hybridization of Di to Ci to occur; next, PCR extension is used on the weight template, now primed by Di, to obtain a complementary product $Vij = di\ r'dj\ r'dj\ r'dj$ attached to Wij in a double strand. This copy Vij is first detached by heating, then primed with copies of all Ci's, extended, and digested separately by the restriction enzyme in a clean separate tube T, which now contains a molecular representation of the net input (i.e., the product $Wx(t)$, but in complementary form). In order to produce the next activation vector, the saturation function is now computed as in [35], by pouring the net input back on recently washed chip A, allowing time for hybridization, and flushing the chip in order to eliminate any excess strands beyond saturation levels and to preserve the accuracy of the process in successive reuse/iterations. The concentration of double-stranded remaining oligonucleotides is the new activation $x(t+1)$ of the neural network at the next step (at time $t+1$). Figure 14.2 illustrates the saturation function. Step 1 creates the DNA of the net input. In Step 2, the DNA is poured over the chip A. The remaining DNA is then passed over chip W for saturation in Step 3. The entire procedure is then repeated as many times as desired to iterate the Hopfield network until a stable activation state is reached. Several variants of this design are possible, taking care of preserving the basic ideas presented above.

In order to verify the reliability of this design experimentally, the discrete Hopfield net example 14.4.2 from [19, p. 690], with three units and three memories was seeded with three sets of inputs and allowed to run for 10 rounds or 10 transitions beginning at state $x(0)$ and ending at state $x(10)$. The total state of the Hopfield memory was recorded at the end of each round. The experiment was performed several times for each of the three inputs. The first input was ideal and should cause the network to instantly recognize the memory and converge to the same stable state immediately (in one round). The second input contained one mismatch and should converge toward the same stable as the first input after several rounds. The last input contained nothing but errors and should converge to the complementary fixed point of the ideal input.

Figure 14.3 shows the Hopfield memory with ideal input. This memory converges within one round to the fixed point of $(-1, 1, -1)$, as expected for this memory. Figure 14.4 further shows that the same memory with one mismatch converges to the ideal output in the fourth round. Again, this behavior is entirely consistent with the behavior of Hopfield memories implemented in silico. Figure 14.5 shows that the Hopfield memory converges away from the correct output when the input is totally corrupted. This behavior is again consistent with Hopfield memories implemented in silico.

The noncross-hybridizing property of the code set of neuron guarantees that a similar behavior will be observed with much larger set of neuronal ensembles in a parallel computing environment, either in vitro or in silico. Note that only a modest amount of hardware (two DNA chips) are required, they are reusable, and the updates can be automated easily, even in microscales using microfluidics [28], in a parallel fashion that is to a large extent independent of the number of neurons. This is a particularly interesting property for potential applications of these systems.

Step 1:
Create DNA Chip for Inputs

Step 2:
Pass DNA over CHIP to activate sites

Step 3:
Remove Excess from Chip

Product:
Activation of each neuron = Signal from each spot
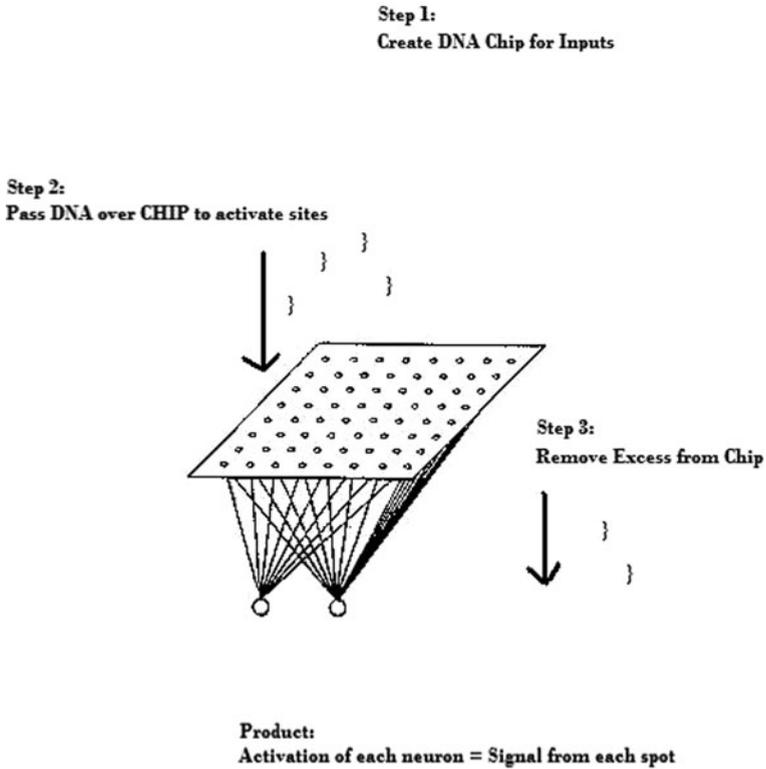
Fig. 14.2: The saturation function is implemented by pouring the Hopfield network's net input over a DNA chip with enough DNA at each spot to express as much as the maximum possible activation. Excess DNA is washed from the chip to obtain the exact next activation.

Hopfield Memory
(Ideal Input)
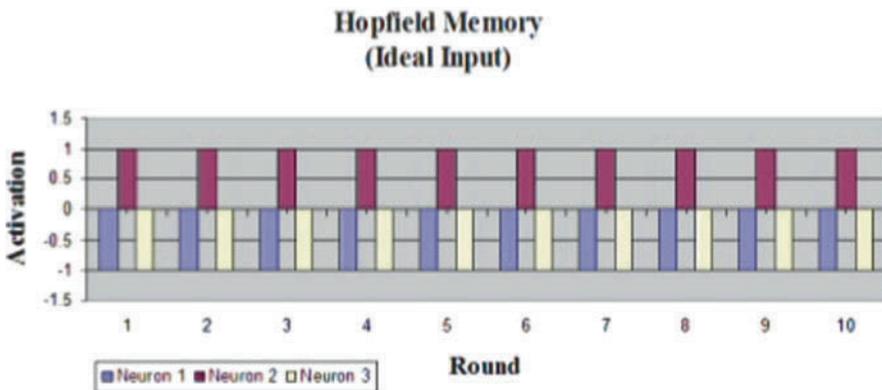
Fig. 14.3: The DNA Hopfield memory was seeded with ideal input. The Hopfield memory converged immediately to the same expected fixed point.
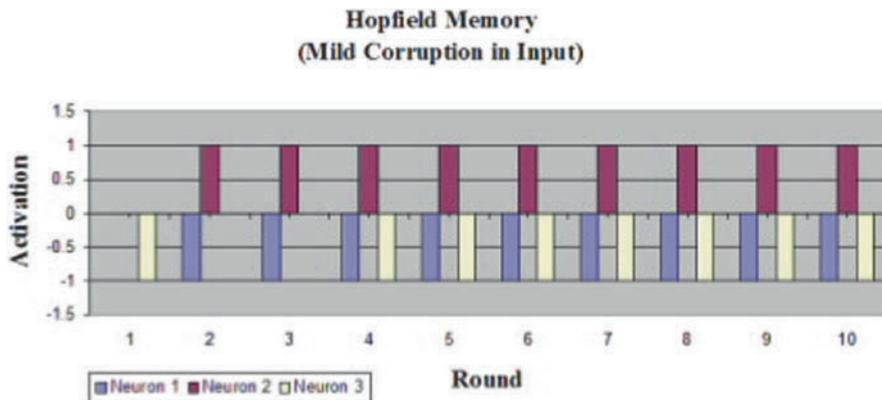
Fig. 14.4: The DNA Hopfield memory was seeded with nearly ideal input containing one error (essentially mild corruption). The Hopfield memory converged to the correct fixed point after some four iterations.
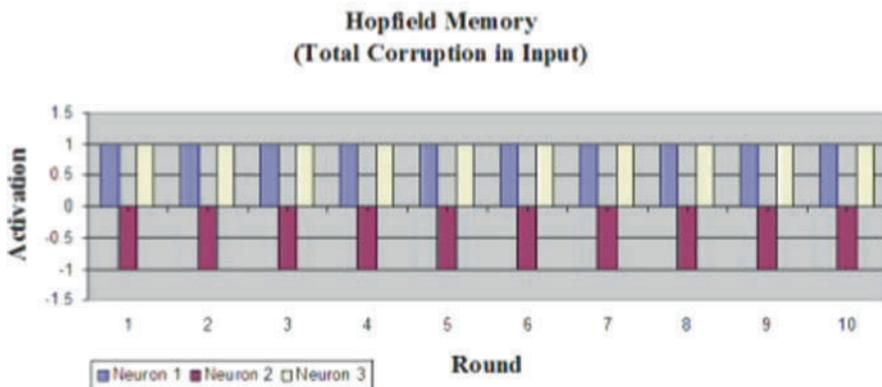
Fig. 14.5: Totally corrupted input results in local minimum of the neural network, as expected, i.e., the complement of the ideal input retrieves the complement result of the ideal input. This behavior corresponds exactly to that of the Hopfield memories implemented on conventional media (in silico).

14.5 Conclusions and Future Work

We have shown that DNA oligonucleotides can be effectively manipulated by currently available biotechnology in order to produce good (perhaps nearly optimal) models of complex systems such as neuronal ensembles, in such a way that the optimality criteria embedded in the phenomenon inherited by the corresponding system. These criteria may include fault tolerance, both in experimental

contamination as well as input data, as illustrated by Hopfield networks. Furthermore, we have shown how simulation of such phenomena in silico can also exhibit similar characteristics while affording a digital implementation, when desired. In summary, a Hopfield net with n neuronal units can be implemented as a DNA Hopfield memory with up to $4n2 + 2n = 2n(2n+1)$ noncross-hybridizing (nxh) oligonucleotides. The requirements on codewords increase linearly with the number of neurons n (not of weights, which increase quadratically). The overall conclusion is that Hopfield nets can be implemented in DNA in ways that preserve their fault tolerance by mapping their energy landscapes to Gibbs landscapes using nxh sets [16], while their implementation is very feasible and becoming easier with currently available biotechnology.

The resulting neuronal ensembles can also be implemented on high-performance digital clusters in silico, or in test tubes in vitro, in scales and densities fairly large in comparison to other implementations. For example, we are in possession of the code sets necessary to implement networks of order up to 100 K neurons on a standard DNA chip of small dimensions (the order of a square inch) with current biotechnology. The rapid progress in microarray design and manufacturing will make feasible much larger arrays at relatively small prices. Other network architectures can be implemented by similar methods. Furthermore, the technique can be easily generalized to other recurrent networks and it can be automated easily, even in microscales using microfluidics [28]. This is a particularly interesting property for potential applications of the networks.

This type of models presented here can be called coarse grained because they only capture critical aspects of the target phenomenon while ignoring most others in order to preserve physical simulations of DNA chemistry. They offer a sharp contrast to systems developed using more traditional methods that aim at physical realism in the simulation of natural phenomena [31]. Thus, the advantages of feasible implementation (both in vitro and in silico, as desired) on the one hand, and the robustness and fault tolerance of neural networks and the optimality inherent in DNA reactions, on the other, can be guaranteed in the design of these systems, at least to a close degree of approximation. The universality properties of neural nets as function approximators ([19], Chapter 4), as well as their ability to even learn certain dynamical systems [30], make them a promising tool in the design of robust and nearly optimal complex systems. Similar efforts are underway to model other complex systems such as biological cells, see, for example, [33, 34, 23].

References

1. Adleman, L. Molecular computation of solutions to combinatorial problems. Science **266**, 1021 (1994)
2. Bar-Yam, Y. Dynamics of Complex Systems. Addison-Wesley, Reading, MA (1997)
3. Benenson, Y., Paz-Elizur, T., Adar, R., Keinan, E., Liben, Z., Shapiro, E. Programmable and autonomous computing machine made of biomolecules. Nature **414**, 430–434 (2001)
4. Bi, H., Chen, J., Deaton, R., Garzon, M., Rubin, H., Wood, D. A PCR based protocol for in vitro selection of non-crosshybridizing oligonucleotides. J Nat Comput **2**(4), 461–477 (2003)
5. Chen, J., Deaton, R., Garzon, M., Kim, J., Wood, D., Bi, H., Carpenter, D., Wang, Y. Characterization of noncrosshybridizing DNA oligonucleotidesmanufactured in vitro. J Nat Comput **1567–7818**, 165–181 (2006)
6. Daly, H. Beyond Growth. Beacon Press, Boston (1996)
7. Deaton, R., Chen, J., Bi, H., Garzon, M., Rubin, H., Wood, D. A PCR-based protocol for in-vitro selection of noncrosshybridzing oligonucleotides. Proceedings of 9th International Meeting on DNA Computing, *LNCS*, Vol. 2568, pp. 196–204. Springer-Verlag, New York (2002)
8. Draghici, S. Data Analysis for DNA Microarrays. Chapman and Hall/CRC, Boca Raton (2003)
9. Ehrgott, M. Multicriteria Optimization. Springer-Verlag, New York (2005)
10. Garey, M., Johnson, D. Computers and Intractability. Freeman, New York (1979)
11. Garzon, M. Biomolecular computing in silico. Selected Collection of EATCS Papers 2000–2003. World Scientific, pp. 505–528 (2004)
12. Garzon, M., Blain, D., Bobba, K., Neel, A., West, M. Self-assembly of DNA-like structures in silico. J Genetic Programming and Evolvable Machines **4**, 185–200 (2003)
13. Garzon, M., Blain, D., Neel, A. Virtual test tubes for biomolecular computing. J Nat Comput **3**(4), 461–477 (2004)
14. Garzon, M., Bobba, K., Hyde, B. Digital Information Encoding on DNA, *LNCS*, Vol. 2950, pp. 152–166. Springer-Verlag, New York (2004)
15. Garzon, M., Deaton, R. Codeword design and information encoding in DNA ensembles. J. of Natural Computing **3**(33), 253–292 (2004)
16. Garzon, M., Phan, V., Roy, S., Neel, A. In search of optimal codes for DNA computing. Proceedings of DNA Computing, 12th International Meeting on DNA Computing, *LNCS*, Vol. 4287, pp. 143–156. Springer-Verlag, New York (2006)
17. Garzon, M., Yao, H. DNA Computing. Proceedings of 13th InternationalMeeting. Proceeding of 9th International Meeting on DNA Computing, *LNCS*, Vol. 4848. Springer-Verlag, New York (2008)
18. Hassoun, M. Associative Neural Networks: Theory and Implementation. Oxford University Press, New York (1993)
19. Haykin, S. Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice-Hall, New Jersey (1999)
20. Holland, J. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor (1975)
21. Hopfield, J. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA **79**(8), 2554–2558 (1982)
22. Koza, J. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Boston (1992)
23. Loew, L., Schaff, J. The virtual cell: A software environment for computational cell biology. Trends Biotechnol **19**(10), 401–406 (2001)
24. Minkle, J. DNA computer plays complete game of tic-tac-toe. Scientific American (2006). Http://www.sciam.com/article.cfm?id=dna-computerplays- comple&ref=rss. Accessed 18 October 2008
25. Minsky, M. The Society of Mind. Simon & Schuster, New York (1985)
26. Mount, D. Bioinformatics: Sequence and Genome Analysis. Spring Harbor, Lab Press (2001)
27. Mullis, K. The unusual origin of the polymerase chain reaction. Sci Am **262**(4), 56–61 (2001)

28. Noort, D.V. A poor man's microfluidic DNA computer. Proceedings of 11th International Meeting on DNA Computing, *LNCS*, Vol. 3892, pp. 380–386. Springer-Verlag, New York (2005)
29. Phan, V., Garzon, M.H. On codeword design in metric DNA spaces. J Nat Comput **8**(3), 571–588 (2009)
30. Rodriguez, J., M, M.G. Learning dynamical systems using neural networks. In: Botelho, J.J., Hagen, M.F. (eds.) Proceedings of the Conference on Fluids and Flows: Recent Trends in Applied Analysis, Vol. 440, pp. 197–206. Contemporary Mathematics, American Mathematical Society (2007)
31. Schlick, T. Molecular Modeling and Simulation. Springer-Verlag, New York (2002)
32. Seeman, N. DNA engineering and its application to nanotechnology. Trends Biotechnol **17**, 437–443 (1999)
33. Sundararaj, S., Guo, A., Habibi-Nazhad, B., Rouani, P., Stothard, M., Ellison, M., Wishar, D. The CyberCell Database (CCDB): A comprehen-sive, self-updating, relational database to co-ordinate and facilitate in silico modeling of Escherichia coli. Nucleic Acids Res **32**(Database is-sue), D293–D295 (2004)
34. Takahashi, K., Ishikawa, N., Sadamoto, Y., et al. E-cell2: Multi-platform e-cell simulation system. Bioinformatics **19**(13), 1727–1729 (2003)
35. Turberfield, A.M.J.M., Turberfield, A., Yurke, B., Platzman, P. Experimental aspects of DNA neural network computation. In: Soft Computing: A Fusion of Foundations, Methodologies, and Applications, Vol. 5(1), pp. 10–18. Springer-Verlag, New York (2001)
36. Watson, J., Baker, T., Bell, S., Gann, A., Levine, M., Losick, R. Molecular Biology of the Gene, 5th edn. Benjamin Cummings, New York (2003)
37. Wikipedia: http://en.wikipedia.org/wiki/n-body problem. Accessed on 12 April 2008
38. Winfree, E., Liu, F., Wenzler, L., Seeman, N. Design and self-assembly of two dimensional DNA crystals. Nature **394**, 539–544 (1998)
39. Yurke, B., Mills, A. Using DNA to power nanostructures. Genet Prog Evolvable Mach **4**, 111–112 (2003)

Part III
Brain Dynamics/Synchronization

Chapter 15
A Robust Estimation of Information Flow in Coupled Nonlinear Systems

Shivkumar Sabesan, Konstantinos Tsakalis, Andreas Spanias, and Leon Iasemidis

Abstract Transfer entropy (TE) is a recently proposed measure of the information flow between coupled linear or nonlinear systems. In this study, we first suggest improvements in the selection of parameters for the estimation of TE that significantly enhance its accuracy and robustness in identifying the direction and the level of information flow between observed data series generated by coupled complex systems. Second, a new measure, the net transfer of entropy (NTE), is defined based on TE. Third, we employ surrogate analysis to show the statistical significance of the measures. Fourth, the effect of measurement noise on the measures' performance is investigated up to $S/N = 3$ dB. We demonstrate the usefulness of the improved method by analyzing data series from coupled nonlinear chaotic oscillators. Our findings suggest that TE and NTE may play a critical role in elucidating the functional connectivity of complex networks of nonlinear systems.

15.1 Introduction

Recent advances in information theory and nonlinear dynamics have facilitated novel approaches for the study of the functional interactions between coupled

Shivkumar Sabesan
The Harrington Department of Bioengineering, Arizona State University, Tempe, AZ 85287, USA; Barrow Neurological Institute, Phoenix, AZ 85013, USA, e-mail: `ssabesa@asu.edu`

Konstantinos Tsakalis
Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA, e-mail: `tsakalis@asu.edu`

Andreas Spanias
Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA, e-mail: `spanias@asu.edu`

Leon Iasemidis
The Harrington Department of Bioengineering, Arizona State University, Tempe, AZ 85287, USA; Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA; Mayo Clinic, Phoenix, AZ 85054, USA, e-mail: `leon.iasemidis@asu.edu`

linear and nonlinear systems. The estimation of these interactions, especially when the systems' structure is unknown, holds promise for the understanding of the mechanisms of their interactions and for a subsequent design and implementation of appropriate schemes to control their behavior. Traditionally, cross-correlation and coherence measures have been the mainstay of assessing statistical interdependence among coupled systems. These measures, however, do not provide reliable information about directional interdependence, i.e., if one system drives the other.

To study the directional aspect of interactions, many other approaches have been employed [24, 22, 11, 18, 19]. One of these approaches is based on the improvement of the prediction of a series' future values by incorporating information from another time series. Such an approach was originally proposed by Wiener [24] and later formalized by Granger in the context of linear regression models of stochastic processes. Granger causality was initially formulated for linear models, and it was then extended to nonlinear systems by (a) applying to local linear models in reduced neighborhoods, estimating the resulting statistical quantity and then averaging it over the entire dataset [20] or (b) considering an error reduction that is triggered by added variables in global nonlinear models [2].

Despite the relative success of the above approaches in detecting the direction of interactions, they essentially are model-based (parametric) methods (linear or nonlinear), i.e., these approaches either make assumptions about the structure of the interacting systems or the nature of their interactions, and as such they may suffer from the shortcomings of modeling systems/signals of unknown structure. For a detailed review of parametric and nonparametric (linear and nonlinear) measures of causality, we refer the reader to [9, 15]. To overcome this problem, an information theoretic approach that identifies the direction of information flow and quantifies the strength of coupling between complex systems/signals has recently been suggested [22]. This method was based on the study of transitional probabilities of the states of systems under consideration. The resulted measure was termed transfer entropy (TE).

We have shown [18, 19] that the direct application of the method as proposed in [22] may not always give the expected results. We show that tuning of certain parameters involved in the TE estimation plays a critical role in detecting the correct direction of the information flow between time series. We propose a methodology to also test the significance of the TE values using surrogate data analysis and we demonstrate its robustness to measurement noise. We then employ the improved TE method to define a new measure, the net transfer entropy (NTE). Results from the application of the improved TE and NTE show that these measures are robust in detecting the direction and strength of coupling under noisy conditions.

The organization of the rest of this chapter is as follows. The measure of TE and the estimation problems we identified, as well as the improvements and practical adjustments that we introduced, are described in Section 15.2. In Section 15.3, results from the application of this method to a system of coupled *Rössler* oscillators are shown. These results are discussed and conclusions are drawn in Section 15.4.

15.2 Methodology

15.2.1 Transfer Entropy (TE)

Consider a kth order Markov process [10] described by

$$P(x_{n+1}|x_n,x_{n-1},\cdots,x_{n-k+1}) = \\ P(x_{n+1}|x_n,x_{n-1},\cdots,x_{n-k}), \tag{15.1}$$

where P represents the conditional probability of state x_{n+1} of a random process X at time $n+1$. Equation (15.1) implies that the probability of occurrence of a particular state x_{n+1} depends only on the past k states $[x_n,\cdots,x_{n-k+1}] \equiv x_n^{(k)}$ of the system. The definition given in Equation (15.1) can be extended to the case of Markov interdependence of two random processes X and Y as

$$P(x_{n+1}|x_n^{(k)}) = P(x_{n+1}|(x_n^{(k)},y_n^{(l)})), \tag{15.2}$$

where $x_n^{(k)}$ are the past k states of the first random process X and $y_n^{(l)}$ are the past l states of the second random process Y. This generalized Markov property implies that the state x_{n+1} of the process X depends only on the past k states of the process X and not on the past l states of the process Y. However, if the process X also depends on the past states (values) of process Y, the divergence of the hypothesized transition probability $P(x_{n+1}|x_n^{(k)})$ (L.H.S. of Equation (15.2)), from the true underlying transition probability of the system $P(x_{n+1}|(x_n^{(k)},y_n^{(l)}))$ (R.H.S of Equation (15.2)), can be quantified using the Kullback–Leibler measure [11]. Then, the Kullback–Leibler measure quantifies the transfer of entropy from the driving process Y to the driven process X, and if it is denoted by TE(Y→X), we have

$$\text{TE}(Y \rightarrow X) = \sum_{n=1}^{N} P(x_{n+1},x_n^{(k)},y_n^{(l)}) \log_2 \frac{P(x_{n+1}|x_n^{(k)},y_n^{(l)})}{P(x_{n+1}|x_n^{(k)})}. \tag{15.3}$$

The values of the parameters k and l are the orders of the Markov process for the two coupled processes X and Y, respectively. The value of N denotes the total number of the available points per process in the state space.

In search of optimal k, it would generally be desirable to choose the parameter k as large as possible in order to find an invariant value (e.g., for conditional entropies to converge as k increases), but in practice the finite size of any real data set imposes the need to find a reasonable compromise between finite sample effects and approximation of the actual value of probabilities. Therefore, the selection of k and l plays a critical role in obtaining reliable values for the transfer of entropy from real data. The estimation of TE as suggested in [22] also depends on the neighborhood size (radius r) used in the state space for the calculation of the involved joint and conditional probabilities. The value of radius r in the state space defines the maximum norm distance in the search for neighboring state space points. Intuitively,

different radius values in the estimation of the multidimensional probabilities in the state space correspond to different probability bins. The values of radius for which the probabilities are not accurately estimated (typically large r values) may eventually lead to an erroneous estimate of TE.

15.2.2 Improved Computation of Transfer Entropy

15.2.2.1 Selection of k

The value of k (order of the driven process) used in the calculation of TE $(Y \rightarrow X)$ (see Equation (15.3)) represents the dependence of the state x_{n+1} of the system on its past k states. A classical linear approach to autoregressive (AR) model order selection, namely the Akaike information criterion (AIC), has been applied to the selection of the order of Markov processes. Evidently, AIC suffers from substantial overestimation of the order of the Markov process order in nonlinear systems and, therefore, is not a consistent estimator [12]. Arguably, a method to estimate this parameter is the delayed mutual information [13]. The delay d at which the mutual information of X reaches its first minimum can be taken as the estimate of the interval within which two states of X are dynamically correlated with each other. In essence, this value of d minimizes the Kullback–Leibler divergence between the dth and higher order corresponding probabilities of the driven process X (see Equation (15.1)), i.e., there is minimum information gain about the future state of X by using its values that are more than d steps in the past. Thus, in units of the sampling period, d would be equal to the order k of the Markov process.

If the value of k is severely underestimated, the information gained about x_{n+1} will erroneously increase due to the presence of y_n and would result to an incorrect estimation of TE. A straightforward extension of this method for estimation of k from real-world data may not be possible, especially when the selected value of k is large (i.e., the embedding dimension of state space would be too large for finite duration data in the time domain). This may thus lead to an erroneous calculation of TE. From a practical point of view, a statistic that may be used is the correlation time constant t_e, which is defined as the time required for the autocorrelation function (AF) to decrease to $1/e$ of its maximum value (maximum value of AF is 1) (see Fig. 15.1d) [13]. AF is an easy metric to compute over time, has been found to be robust in many simulations, but detects only linear dependencies in the data. As we show below and elsewhere [18, 19], the derived results from the detection of the direction and level of interactions justify such a compromise in the estimation of k.

15.2.2.2 Selection of l

The value of l (order of the driving system) was chosen to be equal to 1. The justification for the selection of this value of l is the assumption that the current state
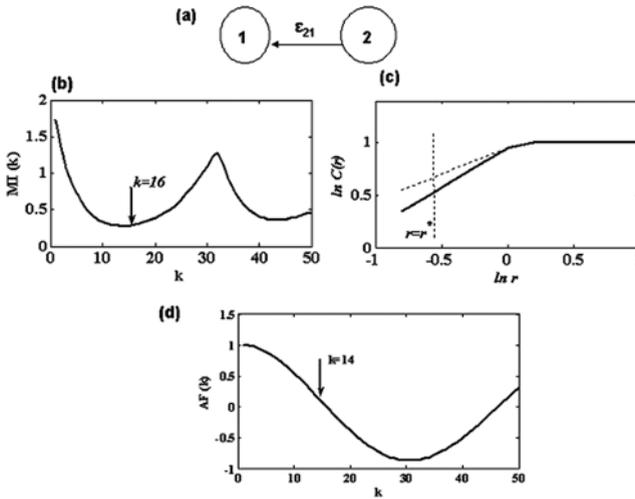
Fig. 15.1: (**a**) Unidirectionally coupled oscillators with $\varepsilon_{21} = 0.05$. (**b**) Mutual information MI vs. k (the first minimum of the mutual information between the X variables of the oscillators is denoted by a downward arrow at $k = 16$). (**c**) $\ln C$ vs. $\ln r$ with $k = 16$, $l = 1$, where C and r denote average joint probability and radius, respectively (*dotted line* for direction of flow $1 \rightarrow 2$; *solid line* for direction of flow $2 \rightarrow 1$). (**d**) Autocorrelation function (AF) vs. k (the delay at which the AF decreases to $1/e$ of its maximum value is denoted by a *downward arrow* at $k = 14$).

of the driving system is sufficient to produce a considerable change in the dynamics of the driven system within one time step (and hence only immediate interactions between X and Y are assumed to be detected in the analysis herein). When larger values for l were employed (i.e., a delayed influence of Y on X), detection of information flow from Y to X was also possible. These results are not presented in this chapter.

15.2.2.3 Selection of Radius r

The multi-dimensional transitional probabilities involved in the definition of transfer entropy (Equation (15.3)) are calculated by joint probabilities using the conditional probability formula $P(A|B) = P(A,B)/P(B)$. One can then reformulate the transfer entropy as

$$
\begin{aligned}
\text{TE}(Y \rightarrow X) = \\
\sum_{n=1}^{N} P(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{P(x_{n+1}, x_n^{(k)}, y_n^{(l)}) P(x_n^{(k)})}{P(x_{n+1}, x_n^{(k)}) P(x_n^{(k)}, y_n^{(l)})}
\end{aligned} \tag{15.4}
$$

From the above formulation, it is clear that probabilities of a vector in the state space at the nth time step are compared with ones of vectors in the state space at the $(n+1)$th time step, and, therefore, the units of TE are in bits/time step, where time step in simulation studies is the algorithm's (e.g., Runge–Kutta) iteration step (or a multiple of it if one downsamples the generated raw data before the calculation of TE). In real life applications (like in electroencephalographic (EEG) data), the time step corresponds to the sampling period of the sampled (digital) data. In this sense, the units of TE denote that TE actually estimates the rate of the flow of information. The multidimensional joint probabilities in Equation (15.4) are estimated through the generalized correlation integrals $C_n(r)$ in the state space of embedding dimension $p = k+l+1$ [14] as

$$
P_r(x_{n+1}, x_n^{(k)}, y_n^{(l)}) =
$$
$$
\frac{1}{N} \sum_{m=0}^{N-1} \Theta \left(r - \left| \begin{matrix} x_{n+1} - x_{m+1} \\ x_n^{(k)} - x_m^{(k)} \\ y_n^{(l)} - y_m^{(l)} \end{matrix} \right| \right) \tag{15.5}
$$
$$
= C_{n+1}(r),
$$

where $\Theta(x > 0) = 1$; $\Theta(x = 0) = 0$, $|\cdot|$ is the maximum distance norm, and the subscript $(n+1)$ is included in C to signify the dependence of C on the time index n (note that averaging over n is performed in the estimation of TE, using Equation (15.5) into Equation (15.3)). In the rest of the chapter we use the notation $C_n(r)$ or $C_{n+1}(r)$ interchangeably. Equation (15.5) is in fact a simple form of a kernel density estimator, where the kernel is the Heaviside function Θ. It has been shown that this approach may present some practical advantages over the box-counting methods for estimating probabilities in a higher dimensional space. We also found that the use of a more elaborate kernel (e.g., a Gaussian or one which takes into account the local density of the states in the state space) than the Heaviside function does not necessarily improve the ability of the measure to detect direction and strength of coupling. Distance metrics other than the maximum norm, such as the Euclidean norm, may also be considered, however, at the cost of increased computation time. In order to avoid a bias in the estimation of the multidimensional probabilities, temporally correlated pairs of points are excluded from the computation of $C_n(r)$ by means of the Theiler correction and a window of $(p-1) * l = k$ points in duration [23].

The estimation of joint probabilities between two different time series requires concurrent calculation of distances in both state spaces (see Equation (15.4)). Therefore, in the computation of $C_n(r)$, the use of a common value of radius r in both state spaces is desirable. In order to establish a common radius r in the state space of X and Y, the data are first normalized to zero mean ($\mu = 0$) and unit variance ($\sigma = 1$). In previous publications [18, 19], using simulation examples (unidirectional as well as bidirectional coupling in two and three coupled oscillator model configurations), we have found that the TE values obtained for only a certain range of r accurately detect the direction and strength of coupling. In general, when any of the joint probabilities ($C_n(r)$) in log scale is plotted against the corresponding radius r in log scale,

it initially increases with increase in the value of the radius (linear increase for small values of r) and then saturates (for large values of r) [3]. It was found that using a value of r^* within the quasilinear region of the $\ln C_n(r)$ vs. $\ln r$ curve produces consistent changes in TE with changes in directional coupling.

Although such an estimation of r^* is possible in noiseless simulation data, for physiological data sets that are always noisy, and the underlying functional description is unknown, it is difficult to estimate an optimal value r^* simply because a linear region of $\ln C_n(r)$ vs. $\ln r$ may not be apparent or even exist. It is known that the presence of noise in the data will be predominant for small r values [10, 8] and over the entire space (high dimensional). This causes the distance between neighborhood points to increase. Consequently, the number of neighbors available to estimate the multidimensional probabilities at the smaller scales may decrease and it would lead to a severely biased estimate of TE. On the other hand, at large values of r, a flat region in $\ln C_n(r)$ may be observed (saturation). In order to avoid the above shortcomings in the practical application of this method (e.g., in simulation models with added noise or in the EEG), we approximated TE as the average of TEs estimated over an intermediate range of r values (from $\sigma/5$ to $2\sigma/5$). The decision to use this range for r was made on the practical basis that r less than $\sigma/2$ typically (well-behaved data) avoids saturation and r larger than $\sigma/10$ typically filters a large portion of A/D-generated noise (simulation examples offer corroborative evidence for such a claim). Even though these criteria are soft for r (no exhaustive testing of the influence of the range of r on the final results), it appears that the proposed range constitutes a very good compromise (sensitivity and specificity-wise) for the subsequent detection of the direction and magnitude of flow of entropy (see Section 15.3). Finally, to either a larger or lesser degree, all existing measures of causality suffer from the finite sample effect. Therefore, it is important to always test their statistical significance using surrogate techniques (see next subsection).

15.2.3 Statistical Significance of Transfer Entropy

Since TE calculates the direction of information transfer between systems by quantifying their conditional statistical dependence, a random shuffling applied to the original driver data series Y destroys the temporal correlation and significantly reduces the information flow $\text{TE}(Y \rightarrow X)$. Thus, in order to estimate the statistically significant values of $\text{TE}(Y \rightarrow X)$, the null hypothesis that the current state of the driver process Y does not contain any additional information about the future state of the driven process X was tested against the alternate hypothesis of a significant time dependence between the future state of X and the current state of Y. One way to achieve this is to compare the estimated values of $\text{TE}(Y \rightarrow X)$ (i.e., the $\text{TE}(x_{n+1}|x_n^{(k)}, y_n^{(l)})$), thereafter denoted by TE_o, with the TE values estimated by studying the dependence of future state of X on the values of Y at randomly shuffled time instants (i.e., $\text{TE}(x_{n+1}|x_n^{(k)}, y_p^{(l)})$), thereafter denoted by TE_s, where $p \in 1, \ldots, N$ is selected from the shuffled time instants of Y. The above described surrogate

analysis is valid when $l = 1$; for $l > 1$, tuples from original Y, each of length l, should be shuffled instead.

The shuffling was based on generation of white Gaussian noise and reordering of the original data samples of the driver data series according to the order indicated by the generated noise values (i.e., random permutation of all indices $1, \ldots, N$ and reordering of the Y time series accordingly). Transfer entropy TE_s values of the shuffled datasets were calculated at the optimal radius r^* from the original data. If the TE values obtained from the original time series (TE_o) were greater than Tth standard deviations from the mean of the TE_s values, the null hypothesis was rejected at the $\alpha = 0.01$ level (The value of Tth depends on the desired level of confidence $1-\alpha$ and the number of the shuffled data segments generated, i.e., the degrees of freedom of the test). Similar surrogate methods have been employed to assess uncertainty in other empirical distributions [4, 21, 16].

15.2.4 Detecting Causality Using Transfer Entropy

Since it is difficult to expect a truly unidirectional flow of information in real-world data (where flow is typically bidirectional), we have defined the causality measure net transfer entropy (NTE) that quantifies the driving of X by Y as

$$\text{NTE}(Y \to X) = \text{TE}(Y \to X) - \text{TE}(X \to Y). \tag{15.6}$$

Positive values of $\text{NTE}(Y \to X)$ denote that Y drives (causes) X, while negative values denote the reverse case. Values of NTE close to 0 may imply either equal bidirectional flow or no flow of information (then, the values of TE will help decide between these two plausible scenarios). Since NTE is based on the difference between the TEs per direction, we expect this metric to generally be less biased than TE in the detection of the driver. In the next section, we test the ability of TE and NTE to detect direction and causality in coupled nonlinear systems and also test their performance against measurement (observation) noise.

15.3 Simulation Example

In this section, we show the application of the method of TE to nonlinear data generated from two coupled, nonidentical, *Rössler*-type oscillators i and j [7], each governed by the following general differential equations:

$$
\begin{aligned}
\dot{x}_i &= -\omega_i y_i - z_i + \sum_{j=1, i \neq j}^{2} \varepsilon_{ji} x_j - \varepsilon_{ii} x_i, \\
\dot{y}_i &= \omega_i x_i + \alpha_i y_i, \\
\dot{z}_i &= \beta_i x_i + z_i (x_i - \gamma_i),
\end{aligned}
\tag{15.7}
$$

where $i, j = 1, 2$, and $\alpha_i = 0.38$, $\beta_i = 0.3$, $\gamma_i = 4.5$ are the standard parameters used for the oscillators to be in the chaotic regime, while we introduce a mismatch in their parameter ω (i.e., $\omega_1 = 1$ and $\omega_2 = 0.9$) to make them nonidentical, ε_{ji} denotes the strength of the diffusive coupling from oscillator j to oscillator i; ε_{ii} denotes self-coupling in the ith oscillator (it is taken to be 0 in this example). Also, in this example, $\varepsilon_{12} = 0$ (unidirectional coupling) so that the direction of information flow is from oscillator 2→1 (see Fig. 15.1a for the coupling configuration). The data were generated using an integration step of 0.01 and a fourth-order Runge–Kutta integration method. The coupling strength ε_{21} is progressively increased in steps of 0.01 from a value of 0 (where the two systems are uncoupled) to a value of 0.25 (where the systems become highly synchronized). Per value of ε_{21}, a total of 10,000 points from the x time series of each oscillator were considered for the estimation of each value of the TE after downsampling the data produced by Runge–Kutta by a factor of 10 (common practice to speed up calculations after making sure the integration of the differential equations involved is made at a high enough precision). The last data point generated at one value of ε_{21} was used as the initial condition to generate data at a higher value of ε_{21}. Results from the application of the TE method to this system, with and without our improvements, are shown next.

Figure 15.1b shows the time-delayed mutual information MI of oscillator 1 (driven oscillator) at one value of ε_{21} ($\varepsilon_{21} = 0.05$) and for different values of k. The first minimum of MI occurs at $k=16$ (see the downward arrow in Fig. 15.1b). The state spaces were reconstructed from the x time series of each oscillator with embedding dimension $p = k + l + 1$. Figure 15.1c shows the $\ln C_{2 \to 1}(r)$ vs. $\ln r$ (dotted line), and $\ln C_{1 \to 2}(r)$ vs. $\ln r$ (solid line), estimated according to Equation (15.5). TE was then estimated according to Equation (15.4) at this value of ε_{21}. The same procedure was followed for the estimation of TE at the other values of ε_{21} in the range $[0, 0.25]$. Figure 15.1d shows the lags of the autocorrelation function AF of oscillator 1 (driven oscillator – see Figure 15.1a) at one value of ε_{21} (that is, $\varepsilon_{21} = 0.05$) and for different values of k. The value of k at which AF drops to $1/e$ of its maximum value was found equal to 14 (see the downward arrow in Fig. 15.1b), that is close to 16 that MI provides us with. Thus, it appears that AF could be used instead of MI in the estimation of k, an approximation that can speed up calculations, as well as end up with an accurate estimate for the direction of information flow.

15.3.1 Statistical Significance of TE and NTE

A total of 50 surrogate data series for each original data series at each ε_{21} coupling value were produced. The null hypothesis that the obtained values of TE_o are not statistically significant was then tested at $\alpha = 0.005$ for each value of ε_{21}. For every ε_{21}, if the TE_o values were greater than 2.68 standard deviations from the mean of the TE_s values, the null hypothesis was rejected (one-tailed t-test; $\alpha = 0.005$). Figure 15.2a depicts the TE_o and the corresponding mean of 50 surrogate TE_s values along with 99% confidence interval error bars in the directions 1→2 and 2→1 (using
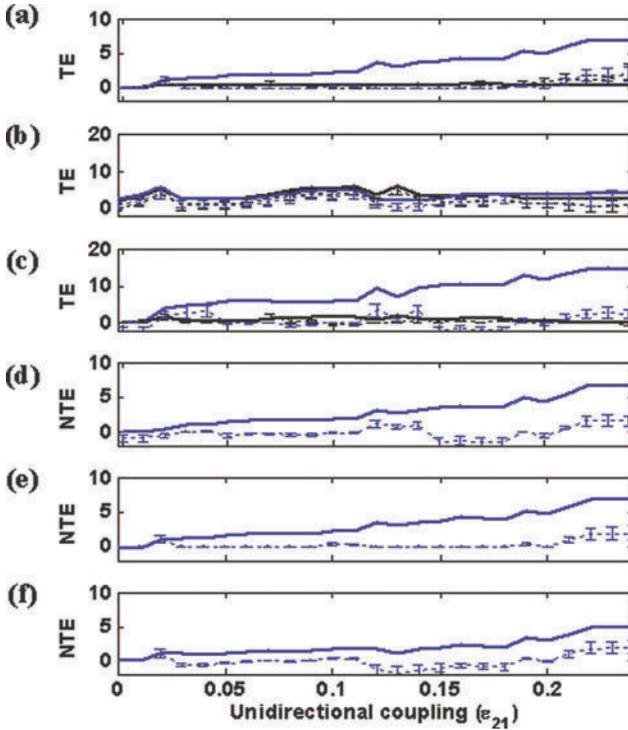
Fig. 15.2: Transfer entropy TE and net transfer of entropy NTE between coupled oscillators 1 and 2 (1→2 *black line*, 2→1 *blue line*) and mean ± 99.5% error bars of their corresponding 50 surrogate values as a function of the systems' underlying unidirectional coupling ε_{21} (from 0 to 0.25). Each TE value was estimated from $N = 10,000$ data points at each ε_{21}. The value of ε_{21} was increased by a step of 0.01. (**a**) TE_o (original data), mean, and 99.5% error bars from the distribution of TE_s (surrogate data). With $k = 16$, $l = 1$ (i.e., the suggested values by our methodology), TE is estimated at radius r^* within the linear region of $\ln C(r)$ vs. $\ln r$ from the original data (see Fig. 15.1c). $TE_o(2→1)$ (*solid blue line*) is statistically significant ($p <0.01$) and progressively increases in value with an increase in ε_{21}, whereas $TE_o(1→2)$ (*solid black line*) is only locally statistically significant and remains constant and very close to 0 despite the increase in ε_{21}. (**b**) TEs estimated with $k = 5$, $l = 5$ as an average of the TEs at intermediate values of the radius r $[\sigma/5 <\ln r < 2\sigma/5]$. Neither $TE_o(2→1)$ nor $TE_o(1→ 2)$ is statistically significant ($p >0.01$) and does not progressively increase in value with an increase in ε_{21}. (**c**) TE estimated with the optimal values $k = 16$, $l = 1$. The picture is very similar to the one in (a) above, suggesting that use of r^* is not critical in the estimation of TEs. (**d**) $NTE_o(2→1)$ and their corresponding $NTE_s(2→1)$ estimated from the TE values in (**c**). (**e**) As in (**d**) above with noise of SNR $= 10$ dB added to the data. (**f**) As in (**d**) above with more noise (SNR $= 3$ dB) added to the data. Detection of direction of information flow is possible at all ε_{21} values ($p <0.01$), except at very small ε_{21} values ($\varepsilon_{21} <0.02$). Units of the estimated measures TE and NTE are in bits per iteration (time step was 0.1, i.e., Runge's time step 0.01 *times* 10, because of the 10:1 decimation we applied on the generated data before analysis).

a *black* and *blue lines* respectively) estimated over ε_{21} at a value of r^* chosen in the linear region of the $\ln C_n(r)$ vs. $\ln r$ of the original data, using MI-suggested k values (k values decrease from 16 to 14 as ε_{21} increases) and with $l = 1$. (The corresponding values for k through the use of AF changed from 14 to 12 with the increase of ε_{21}.) From this figure, it is clear that the $\text{TE}_o(2 \to 1)$ is significantly greater than $\mu(\text{TE}_s) + 2.68 \times \sigma(\text{TE}_s)$ almost over the entire range of coupling, where $\mu(\text{TE}_s)$ is the mean of TE_s over 50 surrogate values and $2.68 \times \sigma(\text{TE}_s)$ is the error bar on the distribution of TE_s (49 degrees of freedom) at the $\alpha = 0.005$ level. [For very small values of coupling ($\varepsilon_{21} < 0.02$), detection of the direction of information flow is not possible ($p > 0.05$).] Also, TE_o shows a progressive increase in the direction $2 \to 1$, proportional to the increase of coupling in that direction, and no significant change in the direction $1 \to 2$. In Fig. 15.2b, the $\text{TE}_o(\varepsilon_{21})$ and the mean and 99.5% error bars on the distribution of $\text{TE}_s(\varepsilon_{21})$ are illustrated for a pair of arbitrary chosen values for k and l (e.g., $k = l = 5$). Neither a statistically significant preferential direction of information flow ($1 \to 2$ or $2 \to 1$) nor a statistically significant progressive increase in TE_o values with the increase of coupling ε_{21} were observed, due to erroneous selection of k and l for the estimation of TE.

In Fig. 15.2c, we present the same quantities as in Fig. 15.2a, but they now are estimated as averages of TEs over an intermediate range of values of r [$\sigma/5 < \ln r < 2\sigma/5$] (that is, not at r^*). We also observe that the TE values in Fig. 15.2c are larger than the ones in Fig. 15.2a with $r = r^*$ and that it is possible from Fig. 15.2c to detect the correct direction of flow and its significant changes with the strength of coupling. This result is very important for the estimation of TE in practical applications, where an optimal r^* is difficult to obtain. In Fig. 15.2d, we show the values of the measure of causality NTE and its statistical significance for the detection of direction and strength of coupling in the two coupled oscillator system over a range of ε_{21}. NTE was also estimated as an average of NTEs over intermediate values of r [$\sigma/5 < \ln r < 2\sigma/5$]. From the statistically significant values of NTE, it is clear that oscillator 2 drives oscillator 1 and the degree of driving increases proportional to the increase in their coupling.

15.3.2 Robustness to Noise

In order to assess the practical usefulness of this methodology for the detection of causality in noise-corrupted data, Gaussian noise with variance corresponding to a 10 or 3 dB signal-to-noise ratio (SNR) was added independently to the X series of the original data from each of the two coupled *Rössler* systems. The noisy data were then processed in the same way as the noise-free data, including testing against the null hypothesis that an obtained value of TE at each coupling value ε_{21} is not statistically significant. The corresponding TE_o and TE_s values are illustrated in Fig. 15.2e, f for each of the two SNR values, respectively. It is noteworthy that only at extremely low values of coupling ($\varepsilon_{21} < 0.02$) NTE cannot detect the direction of information flow. Thus, it appears that the NTE, along with the suggested

improvements and modifications for the estimation of TE, is a robust measure for detecting the direction and the rate of the net information flow in coupled nonlinear systems even under severe noise conditions (e.g., SNR = 3 dB).

15.4 Discussion and Conclusion

In this study, we suggested and implemented improvements for the estimation of transfer entropy (TE), a measure of the direction and the level of information flow between coupled subsystems, built upon it to introduce a new measure of information flow, and showed their application to a simulation example. The two innovations we introduced in the TE estimation were: (a) the distance in the state space at which the required probabilities should be estimated and (b) the use of surrogate data to evaluate the statistical significance of the estimated TE values. The new estimator for TE was shown to be consistent and reliable when applied to complex signals generated by systems in their chaotic regime. A more practical estimator of TE, that averages the values of TE produced in an intermediate range of distances r in the state space, was shown to be robust to additive noise up to $S/N=3$ dB, and could reliably and significantly detect the direction of information flow for a wide range of coupling strengths, even for coupling strengths close to 0. Our analysis in this chapter dealt with only pairwise (bivariate) interactions between subsystems and as such, it does not detect both direct and indirect interactions among multiple subsystems at the time resolution of the sampling period of the data involved. A multivariate extension of TE to detect information flow between more than two subsystems is straightforward. Such an extension could also be proven useful in distinguishing between direct and indirect interactions [6,5], and thus further enhance TE's capability to detect causal interactions from experimental data.

A new measure of causality, namely net transfer of entropy [NTE($i \rightarrow j$)], was then introduced for a system i driving a system j (see Equation (15.6)). NTE of the system i measures the outgoing net flow of information from the driving i to the driven j system, that is, it takes into consideration both incoming TE to and outgoing TE from the driving system i. Our simulation example herein also showed the importance of NTE for the identification of the driving system in a pair of coupled systems for a range of coupling strengths and noise levels. We believe that our approach to estimating information flow between coupled systems can have several potential applications to coupled complex systems in diverse scientific fields, from medicine and biology, to physics and engineering.

References

1. Bharucha-Reid, A. Elements of the Theory of Markov Processes and Their Applications. Courier Dover Publications, Chemsford, MA (1997)
2. Chen, Y., Rangarajan, G., Feng, J., Ding, M. Analyzing multiple nonlinear time series with extended Granger causality. Phys Lett A **324**(1), 26–35 (2004)
3. Eckmann, J., Ruelle, D. Ergodic theory of chaos and strange attractors. In: Ruelle, D. (ed.) Turbulence, Strange Attractors, and Chaos, pp. 365–404. World Scientific, Singapore (1995)
4. Efron, B., Tibshirani, R. An Introduction to the Bootstrap. CRC Press, Boca Raton (1993)
5. Franaszczuk, P., Bergey, G. Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures. Brain Topogr **11**(1), 13–21 (1998)
6. Friston, K. Brain function, nonlinear coupling, and neuronal transients. Neuroscientist **7**(5), 406–418 (2001)
7. Gaspard, P., Nicolis, G. What can we learn from homoclinic orbits in chaotic dynamics? J Stat Phys **31**(3), 499–518 (1983)
8. Grassberger, P. Finite sample corrections to entropy and dimension estimates. Phys Lett A **128**(6–7), 369–373 (1988)
9. Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M., Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. Phys Rep **441**(1), 1–46 (2007)
10. Iasemidis, L.D., Sackellares, J.C., Savit, R. Quantification of hidden time dependencies in the EEG within the framework of nonlinear dynamics. In: Jansen, B., Brandt, M. (eds.) Nonlinear Dynamical Analysis of the EEG, pp. 30–47. World Scientific, Singapore (1993)
11. Kaiser, A., Schreiber, T. Information transfer in continuous processes. Physica D **166**, 43–62 (2002)
12. Katz, R. On some criteria for estimating the order of a Markov chain. Technometrics **23**(3), 243–256 (1981)
13. Martinerie, J., Albano, A., Mees, A., Rapp, P. Mutual information, strange attractors, and the optimal estimation of dimension. Phys Rev A **45**(10), 7058–7064 (1992)
14. Pawelzik, K., Schuster, H. Generalized dimensions and entropies from a measured time series. Phys Rev A **35**(1), 481–484 (1987)
15. Pereda, E., Quiroga, R.Q., Bhattacharya, J. Nonlinear multivariate analysis of neurophysiological signals. Prog Neurobiol **77**(1–2), 1–37 (2005)
16. Politis, D., Romano, J., Wolf, M. Subsampling, Springer Series in Statistics, Springer Verlag, New York (1999)
17. Quiroga, R.Q., Arnhold, J., Lehnertz, K., Grassberger, P. Kulback-Leibler and renormalized entropies: Applications to electroencephalograms of epilepsy patients. Phys Rev E **62**(6), 8380–8386 (2000)
18. Sabesan, S., Narayanan, K., Prasad, A., Spanias, A. and Iasemidis, L. Improved measure of information flow in coupled nonlinear systems. In: Proceedings of International Association of Science and Technology for Development, pp. 24–26 (2003)
19. Sabesan, S., Narayanan, K., Prasad, A., Tsakalis, K., Spanias, A., Iasemidis, L. Information flow in coupled nonlinear systems: Application to the epileptic human brain. In: Pardalos, P., Boginski, V., Vazacopoulos, A. (eds.) Data Mining in Biomedicine, Springer Optimization and Its Applications Series, Springer, New York, pp. 483–504 (2007)
20. Schiff, S., So, P., Chang, T., Burke, R., Sauer, T. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. Phys Rev E **54**(6), 6708–6724 (1996)
21. Schreiber, T. Determination of the noise level of chaotic time series. Phys Rev E **48**(1), 13–16 (1993)
22. Schreiber, T. Measuring information transfer. Phys Rev Lett **85**(2), 461–464 (2000)
23. Theiler, J. Spurious dimension from correlation algorithms applied to limited time-series data. Phys Rev A **34**(3), 2427–2432 (1986)
24. Wiener, N. Modern Mathematics for the Engineers [Z]. Series 1. McGraw-Hill, New York (1956)

Chapter 16
An Optimization Approach for Finding a Spectrum of Lyapunov Exponents

Panos M. Pardalos, Vitaliy A. Yatsenko, Alexandre Messo, Altannar Chinchuluun, and Petros Xanthopoulos

Abstract In this chapter, we consider an optimization technique for estimating the Lyapunov exponents from nonlinear chaotic systems. We then describe an algorithm for solving the optimization model and discuss the computational aspects of the proposed algorithm. To show the efficiency of the algorithm, we apply it to some well-known data sets. Numerical tests show that the algorithm is robust and quite effective, and its performance is comparable with that of other well-known algorithms.

16.1 Introduction

Brain electrical activity can be recorded from electrodes placed on the scalp or intracranial. It is now possible to record from relatively large areas with macro-electrodes (EEG) or from more localized regions, using microelectrodes. Such recordings, performed in awake, moving animals, and humans have advanced our understanding of

Panos M. Pardalos

Department of Industrial and Systems Engineering, Center for Applied Optimization, University of Florida, Gainesville, FL, USA, e-mail: `pardalos@ufl.edu`

Vitaliy A. Yatsenko

Department of Industrial and Systems Engineering, Center for Applied Optimization, University of Florida, Gainesville, FL, USA, e-mail: `yatsenko@ufl.edu`

Alexandre Messo

Department of Optimization, Kungliga Tekniska Högskolan, Stockholm, Sweden, e-mail: `alex.messo@gmail.com`

Altannar Chinchuluun

Department of Industrial and Systems Engineering, Center for Applied Optimization, University of Florida, Gainesville, FL, USA, e-mail: `altannar@ufl.edu`

Petros Xanthopoulos

Department of Industrial and Systems Engineering, Center for Applied Optimization, University of Florida, Gainesville, FL, USA, e-mail: `petrosx@ufl.edu`

many normal physiological processes, such as the sleep–wake cycle and motor control, as well as pathological conditions such as epilepsy, Parkinson's disease and other movement disorders, and sleep disorders such as sleep apnea and narcolepsy. Traditionally, neurophysiologists analyze such signals using visual inspection or through statistical analysis of linear signal properties such as the spectrogram and coherence. More recently, investigators have begun to investigate the spatiotemporal dynamical features of neurological signals. However, many of these techniques have been applied to mathematical models or to dynamical systems that are much less complex than the brain. Therefore, there is need to develop and evaluate mathematical techniques that provide robust results in higher dimensional, nonstationary and noisy biological systems such as the brain. Although difficult or even impossible to prove, there has been little debate that brain activities should be modeled as nonlinear systems. As a result, during the last decade, a variety of nonlinear time series analysis techniques have been applied repeatedly to EEG recordings during physiologic and pathologic conditions. Among those, the algorithms based on the Lyapunov exponents appears promising for characterizing the spatiotemporal dynamics in electroencephalogram (EEGs) time series recorded from patients with temporal lobe epilepsy. Nevertheless, there are many improvements can be made in algorithms for finding Lyapunov exponents so that the estimation can be more robust, especially with respect to the presence of noise in the EEG. As the complexity of the algorithms for finding Lyapunov exponents with the noise and nonstationarity of EEG, this task requires development of novel techniques and numerous computational experiments.

This chapter is organized as follows. In the next section, we describe Lyapunov exponents and discuss some of the algorithms for finding them. In Section 16.3, an optimization model for estimating Lyapunov exponents is presented and a solution technique for the model is proposed. Brief descriptions of the models used for computational experiments and a comparison of performance, including sensitivity analysis, between the proposed algorithm and the two well-known algorithms are given in Sections 16.4 and 16.5, respectively. The details of the numerical computations are also given in Section 16.5.

16.2 Lyapunov Exponents

Chaos is one type of behavior exhibited by nonlinear dynamical systems, which are systems whose time evolution equations are nonlinear, that is, the dynamical variables describing the properties of the systems (for example, position, velocity, acceleration, pressure) appear in the equations in a nonlinear form. There are several techniques to measure chaos, depending on what one wants to characterize in the chaotic trajectory. Some of the techniques include: simple visual inspection of either the time series represented by a time plot of the trajectory, or the bounded strange attractor reconstructed from the time series; spectral analysis of the time series; Lyapunov exponents; and entropy analysis. Here, we focus

on *Lyapunov exponents*, as this metric has many important characteristics such as invariance to transformations and computability directly from data, without solving the differential or difference equations describing the corresponding dynamical system.

Let us consider any two nearby divergent trajectories originating from a 1D flow (i.e., trajectories in continuous time as described by differential equations). The growth of the difference δ_t between the two nearby trajectories over a time period $\Delta t = t_1 - t_0$ can be described by

$$\delta_t \sim \delta_0 e^{\lambda \Delta t}, \tag{16.1}$$

where λ denotes the systems Lyapunov exponent.

An important reason for using the Lyapunov exponent as a characteristic measure of a dynamical system is its invariance[1] to rescaling, shifts and other transformations of data such as the imprecise reconstruction of a strange attractor from a time series. The fact that trajectories diverge over the course of time would not in itself be very dramatic if it was only very slow, thus we speak of chaos only if this separation is exponentially fast. There are n different Lyapunov exponents for an n-dimensional system, defined as follows: Consider the evolution of an infinitesimal sphere of perturbed initial conditions. During its evolution along the reference trajectory, the sphere will become deformed into an infinitesimal ellipsoid. Let $\delta_k(t)$, $k = 1, 2, 3, \ldots, n$, denote the length of the kth principal axis of the ellipsoid. Thus, the deformation of the sphere corresponds to the stretching, contraction, and rotation of the principal directions. For large t, the diameter of the ellipsoid is controlled by the most positive λ_k. As we shall see later on, λ depends slightly on which trajectory we study, so we should average over many different points on the same trajectory to get the true value of λ (see Table 16.1).

In dissipative systems one can also find a negative maximal Lyapunov exponent which reflects the existence of a stable fixed point. Two trajectories which approach the fixed point also approach each other exponentially fast. If the motion settles down onto a limit cycle, two trajectories can only separate or approach each other slower than exponentially. In this case the maximal Lyapunov exponent is 0 and the motion is called marginally stable. If a deterministic system is perturbed by random noise, on the small scales it can be characterized by a diffusion process, with δ_t growing as $\sqrt{t}$. Thus the maximal Lyapunov exponent is infinite. According to the

Table 16.1: Possible types of motion and the corresponding Lyapunov exponents

Type of motion	Maximal Lyapunov exponent
Stable fixed point	$\lambda < 0$
Stable limit cycle	$\lambda = 0$
Chaos	$0 < \lambda < \infty$
Noise	$\lambda = \infty$

---

[1] See Oseledec's theorem in [12].

mathematical definition, this is true no matter how small the noise component is, however, we will show later on that Lyapunov exponents of an underlying deterministic system can in fact be measured.

When a system has a positive Lyapunov exponent, there is a time horizon beyond which prediction breaks down.

Wolf et al. [34] proposed the first algorithm for calculating the largest Lyapunov exponent. First, the phase space is reconstructed and the nearest neighbor is searched for one of the first embedding vectors. A restriction must be made when searching for the neighbor: it must be sufficiently separated in time in order not to compute as nearest neighbors successive vectors of the same trajectory. Without considering this correction, Lyapunov exponents could be spurious due to temporal correlation of the neighbors. Once the neighbor and the initial distance L is determined, the system is evolved forward some fixed time (evolution time) and the new distance L' is calculated. This evolution is repeated, calculating the successive distances, until the separation is greater than a certain threshold. Then a new vector (replacement vector) is searched as close as possible to the first one, having approximately the same orientation of the first neighbor. Finally, Lyapunov exponents can be estimated using the following formula:

$$\lambda_1 = \frac{1}{(t_M - t_0)} \sum_{k=1}^{M} \ln \frac{L'(t_k)}{L(t_k - 1)}, \qquad (16.2)$$

where k is the number of time propagation steps.

The Wolf algorithm only estimates the largest Lyapunov exponent and not the whole spectrum of exponents. It is said to be sensitive to the number of observations as well as to the degree of measurement or system noise in the observations. This discovery motivated a search for new algorithm designs with improved finite-sample properties. Sano and Sawada [28], Eckmann et al. [5], Abarbanel et al. [1], Rosenstein et al. [27], and Pardalos and Yatsenko [24], among others, came up with improved algorithms for calculating the Lyapunov exponents from observed data.

16.3 An Optimization Approach

In the previous section, we mentioned a number of algorithms that have been proposed for estimating the Lyapunov exponents from a scalar time series. The problem of calculating these exponents can be reformulated as an optimization problem (see Pardalos and Yatsenko [24]), and in these following sections we present an algorithm for its solution which is globally and quadratically convergent. Here, we use well-established techniques from numerical methods for dealing with the optimization problem. We also discuss the computational aspects of this method and the difficulties which inevitably arises when estimating the Lyapunov exponents based on the use of time-delay embedding. Using numerically generated data sets, we consider the influence of the system parameters and the optimization algorithm on the quality of the estimates.

16.3.1 Theory

Let us consider a vector x in the phase space $\mathbb{R}^n$ which can be considered as a solution of a certain dynamical system

$$\dot{x} = f(x,t), \quad x(t_0) = x_0, \tag{16.3}$$

where $f(x,t)$ is a smooth vector field on a manifold $\mathcal{M}$. The vector field f yields a flow $\phi = \{\phi(t)\}$ on the phase space, where $\phi(t)$ is a map,

$$x \mapsto \phi(x,t), \quad t \in \mathbb{R}, \quad x \in \mathbb{R}^n. \tag{16.4}$$

The observed trajectory, starting at x_0, is

$$\{\phi(x_0,t)|t \in \mathbb{R}^+\}. \tag{16.5}$$

To get an information about the time evolution of arbitrarily small perturbed initial conditions, consider the evolution of tangent vectors in the tangent space $\mathcal{TM}$. It is given by the linearization of Equation (16.3).

The Taylor expansion of $f(\phi(x_0,t))$ for small Δx is

$$f(\phi(x_0,t)) + Df(\phi(x_0,t))\Delta x + \cdots. \tag{16.6}$$

Here $Df(\phi(x_0,t))$ is the local Jacobian matrix of the vector field f at $\phi(x_0,t)$:

$$Df(\phi(x_0,t)) = J(x_0,t) = [(\partial f_i/\partial x_j)|_{\phi(x_0,t)}]. \tag{16.7}$$

For $\Delta x \to 0$, the following first-order approximation holds:

$$\dot{\delta x} = J(x_0,t)\delta x. \tag{16.8}$$

The solution of the linear nonautonomous variational equation (16.8) can be obtained as

$$\delta x(t) = D\phi(x_0,t)\delta x_0, \tag{16.9}$$

where $D\phi(x_0,t) = A(x_0,t) \in \mathbb{R}^{n \times n}$ is the linear operator which maps tangent vector δx_0 to $\delta x(t)$.

The spectrum of the Lyapunov exponents is the set of logarithms of the eigenvalues of the self-adjoint matrix

$$\Lambda_{x_0} = \lim_{t \to \infty}[A(x_0,t)^\top A(x_0,t)]^{1/2t}, \tag{16.10}$$

where $A(x_0,t)^\top$ is the transpose of the matrix $A(x_0,t)$. The existence of the limit in Equation (16.10) is proved by Oseledec's theorem in [12].

Let $E = (e_1,\ldots,e_n)$ be an $n \times n$ matrix, where the column vectors are a basis of the tangent space. If the following limit exists:

$$\lambda_i = \lim_{t \to \infty} \frac{1}{t} \log \|A(x_0,t)e_i\|, \tag{16.11}$$

then λ_i, $i = 1, \ldots, n$, are the Lyapunov exponents. They are ordered by their magnitudes $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and if they are independent of x_0, the system is called *ergodic*.

Therefore, one can write $A(x_0,t)$ as the product of $n \times n$ matrices $A(x_j, \Delta t)$, where each one maps $x_j = \phi(x_0, j\Delta t)$ to x_{j+1}:

$$A(x_0, k\Delta t) = \prod_{j=0}^{k-1} A(x_j, \Delta t), \tag{16.12}$$

with $k\Delta t = t$.

16.3.2 Implementation Details

We often have no knowledge of the nonlinear equations of the system which produce the observed time series. But there is a possibility of estimating the linearized flow map $A_{\Delta t} = D\phi(x_j, \Delta t)$ from a single trajectory by using the recurrent structure of strange attractors. Let $\{x_j\}$, $j = 1, 2, \ldots$, denote a time series of some physical quantity measured at the discrete time interval Δt, i.e., $x_j = x(t_0 + (j-1)\Delta t)$. Consider a small ball of radius ε centered at the orbital point x_j, and find a set of N difference vectors included in this ball, i.e.,

$$\{y_i\} = \{x_j - x_i | \ \|x_j - x_i\| \leq \varepsilon\}, \quad i = 1, 2, \ldots, N^2, \tag{16.13}$$

where y_i is the displacement vector between x_j and x_i. Here, $\|\cdots\|$ denotes a usual Euclidean norm defined as follows: $\|w\| = (w_1^2 + w_2^2 + \ldots + w_n^2)^{1/2}$ for some vector $w = (w_1, w_2, \ldots, w_n)$. After the evolution of a time interval $k\Delta t$, $y_i = x_j - x_i$ is mapped to the set

$$\{z_i\} = \{x_{j+k} - x_{i+k}\}, \quad i = 1, 2, \ldots, N. \tag{16.14}$$

If the radius ε and the evolution time Δt are small enough for the displacement vectors $\{y_i\}$ and $\{z_i\}$ to be regarded as a good approximation of tangent vectors in the tangent space $\mathcal{TM}$, the evolution of y_i to z_i can be represented by some matrix A_j as

$$z_i = A_j y_i. \tag{16.15}$$

The matrix A_j should be a good approximation of the matrix of linearized flow in Equation (16.9). A plausible procedure for optimal estimation is the least-square

---

[2] In the implementation, among the N displacement vectors found inside the sphere of radius ε, only five to seven vectors with the smallest norm are chosen. N is often chosen as $d_E \leq N \leq 20$ [28] and is kept at a low value to optimize the efficiency of the algorithm.

algorithm, which minimizes the average of the squared error norm between z_i and $A_j y_i$ with respect to all components of A_j as follows:

$$\min_{A_j} S = \frac{1}{N} \sum_{i=1}^{N} \|z_i - A_j y_i\|^2, \tag{16.16}$$

$$\text{subject to} \quad a_{kl}^j \in \mathbb{R}, \tag{16.17}$$

where a_{kl}^j denotes the (k,l) component of matrix A_j. The evolution times Δt in the renormalization and the approximation process do not necessarily have to be the same, but are chosen equal for convenience.

Each invertible $n \times n$ matrix can be split uniquely into the product of an upper triangular matrix R and an orthogonal matrix Q, such that

$$A_j E_j = Q_j R_j = E_{j+1} R_j, \tag{16.18}$$

with $E_j = (e_j^1, \ldots, e_j^n)$. The matrix Q_j serves as the new basis E_{j+1} and the logarithms of the diagonal elements of R_j are local expanding coefficients, whose time-averaged values are the Lyapunov exponents. Using

$$A(x_0, k\Delta t)E_0 = \prod_{j=0}^{k-1} A(x_j, \Delta t)E_0 = Q_{k-1} \prod_{j=0}^{k-1} R_j \tag{16.19}$$

in Equation (16.10), we obtain

$$\lambda_i = \lim_{k \to \infty} \frac{1}{k\Delta t} \sum_{j=0}^{k-1} \log r_{ii}^j, \tag{16.20}$$

where r_{ii}^j are the diagonal elements of the matrix R_j.

In the numerical procedure, we let A_j operate on an arbitrary chosen set $\{e_i^j\}$, and then renormalize $A_j e_i^j$ to have unit length. Mutual orthogonality of the basis is maintained by using the Gram–Schmidt renormalization procedure. This is repeated for n iterations where Equation (16.20) is computed each time [25].

16.3.2.1 Phase Space Reconstruction

One important reason for using the above approach to computing λ is that the stability of the dynamical system can be determined without actually knowing and solving the underlying differential equations explicitly. This occurs when we obtain a chaotic time series from a dynamical system, reconstruct its strange attractor in the corresponding phase space, and then compute the Lyapunov exponents from the reconstructed strange attractor directly, without its explicit mathematical model.

The most important phase space reconstruction technique is the *method of delays*. The basic idea is very simple. We use the time series data of a single variable

to create a multidimensional reconstruction (embedding) space. If the embedding space is generated properly, the behavior of trajectories in this embedding space will have the same geometric and dynamical properties that characterize the actual trajectories in the full multidimensional phase space of the system. The method of delays was suggested by Packard et al. [17] in 1980 and was put on a firm theoretical basis by Takens [32] in 1981.

From the set of observations $x(t_0 + n\Delta t) = x(n)$, multivariate vectors in d-dimensional space

$$y(n) = (x(n), x(n+\tau), \ldots, x(n+(d-1)\tau)) \tag{16.21}$$

are used to trace out the orbit of the system. The observations, $x(n)$, are a projection of the multivariate phase space of the system onto the 1D axis of the $x(n)$'s. The purpose of time-delay embedding technique is to unfold the projection back to a multivariate phase space that is representative of the original system. In practice, the natural questions of what time delay τ and what embedding dimension d to use in this reconstruction have had a variety of answers. The following sections present the methods used in this chapter for determining τ and d.

The time-delay parameter τ: The choice of time delay is not a straightforward problem. If it is taken too small, there is almost no difference between the different elements of the delay vectors. If on the other hand τ is very large, the different coordinates may be almost uncorrelated. In this case the reconstructed attractor may become very complicated, even if the true underlying attractor is simple. This is typical of chaotic systems, where the autocorrelation function decays fast. Unfortunately, since τ has no relevance in the mathematical framework, there exists no rigorous way of determining its optimal value. At least a dozen different methods have been suggested for the estimation of τ, and since all these methods yield values of similar magnitude, we should estimate τ just by a single preferred tool and work with this estimate [11]. Past studies have made use of the autocorrelation function, but a quite reasonable objection to this procedure is that it is based on linear statistics, not taking into account nonlinear dynamical correlations. Therefore, it is sometimes recommended that one look for the first minimum of the time-delayed *mutual information*. This is the information we already possess about the value of $x(t+\tau)$ if we know $x(t)$.

On the interval explored by the data, we create a histogram for the probability distribution of the data. We denote by p_i the probability that the signal assumes a value inside the ith bin of the histogram, and let $p_{ij}(\tau)$ be the probability that $x(t)$ is in bin i and $x(t+\tau)$ is in bin j. Then the mutual information for time delay τ reads

$$I(\tau) = \sum_{i,j} p_{ij}(\tau) \ln p_{ij}(\tau) - 2 \sum_i p_i \ln p_i. \tag{16.22}$$

The value of the mutual information is independent of the particular choice of histogram, as long as it is fine enough.[3] The first minimum of $I(\tau)$ marks the time

---

[3] Throughout this chapter 512 bins have been used.

delay where $x(t+\tau)$ adds maximal information to the knowledge we have from $x(t)$. This is the time delay used in this chapter.

Embedding dimension: What is the appropriate value of d to use as the embedding dimension? The procedure used here identifies the number of *false nearest neighbors*, points that appear to be nearest neighbors because the embedding space is too small, of every point on the attractor associated with the orbit $y(n)$, $n = 1, 2, \ldots, N$. When the number of false nearest neighbors drops to 0, we have unfolded or embedded the attractor in $\mathbb{R}^d$, a d-dimensional Euclidean space.

If we are in d dimensions and we denote the rth nearest neighbor of $y(n)$ by $y^{(r)}(n)$, then from Equation (16.21), the square of the Euclidean distance between the point $y(n)$ and this neighbor is

$$R_d^2(n,r) = \sum_{k=0}^{d-1} [x(n+k\tau) - x^{(r)}(n+k\tau)]^2. \qquad (16.23)$$

In going from dimension d to dimension $d+1$ by time-delay embedding we add a $(d+1)$th coordinate onto each of the vectors $y(n)$. This new coordinate is just $x(n+\tau d)$. The Euclidean distance, as measured in dimension $d+1$, between $y(n)$ and the same rth neighbor as determined in dimension d is given by

$$R_{d+1}^2(n,r) = R_d^2(n,r) + [x(n+\tau d) - x^{(r)}(n+\tau d)]^2. \qquad (16.24)$$

A natural criterion for catching embedding errors is that the increase in distance between $y(n)$ and $y^{(r)}(n)$ is large when going from dimension d to $d+1$. The increase in distance can be stated quite simply from Equations (16.23) and (16.24). We state this criterion by designating as a false neighbor any neighbor for which

$$\sqrt{\frac{R_{d+1}^2(n,r) - R_d^2(n,r)}{R_d^2(n,r)}} = \frac{|x(n+\tau d) - x^{(r)}(n+\tau d)|}{R_d(n,r)} > R_{\text{tol}}, \qquad (16.25)$$

where R_{tol} is some threshold. In practical settings the number of data points is often not large, and the following criterion handles the issue of limited data set size: If the nearest neighbor to $y(n)$ is not close ($R_d(n) \approx R_A$) and it is a false neighbor, then the distance $R_{d+1}(n)$ resulting from adding on a $(d+1)$th component to the data vectors will be $R_{d+1}(n) \approx 2R_A$ [13]. That is, even distant but nearest neighbors will be stretched to the extremities of the attractor when they are unfolded from each other, if they are false nearest neighbors. We write this second criterion as

$$\frac{R_{d+1}(n)}{R_A} > A_{\text{tol}}, \qquad (16.26)$$

where R_A denotes the size of the attractor. Both criterions in Equations (16.25) and (16.26) are used jointly throughout the determination of d_E (also see Fig. 16.1). As a measure of R_A we have chosen the standard deviation $\sigma(\mathbf{x})$ of the observed data $\mathbf{x}$ according to [12]. This source also gives us the recommended values $R_{\text{tol}} = 15.0$ and $A_{\text{tol}} = 2.0$.
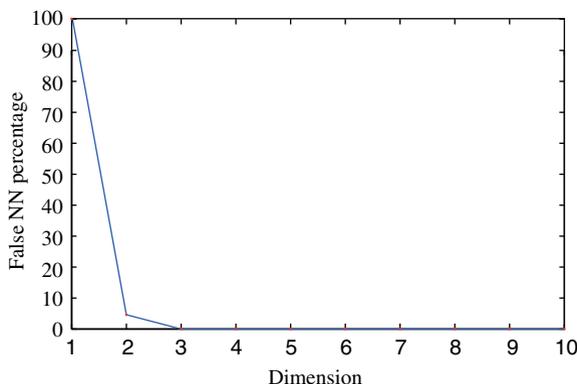
Fig. 16.1: The percentage of false nearest neighbors for 10,000 data points from the Lorenz equations (see Section 2.2.2). The data were output at $\Delta t = 0.01$ during the integration. A time lag $\tau = 12\Delta t = 0.12$, which is the location of the first minimum in the average mutual information for this system, was used in forming the time-delayed vectors.

From the point of view of the mathematics of the embedding process it does not matter whether one uses the minimum embedding dimension d_E or any $d \geq d_E$, since once the attractor is unfolded, the theorem's work is done. For a physicist the story is quite different. Working in any dimension larger than the minimum required by the data leads to excessive computation when investigating the Lyapunov exponents. It also enhances the problem of contamination by roundoff or instrumental error since this noise will populate and dominate the additional $d - d_E$ dimensions of the embedding space where no dynamics is operating. We should add that in going through the data set and determining which points are near neighbors of the point $y(n)$ we use the sorting method of a k-dimensional tree to reduce the computation time from $\mathcal{O}(n^2)$ to $\mathcal{O}(N\log_{10}(N))$.

16.4 Models Used in the Computational Experiments

We evaluate the algorithm performance using the signals simulated from four well-known dynamical mathematical models: Lorenz, Rössler, Hénon, and Hénon–Heilers. Brief descriptions of the models are given below.

16.4.1 Lorenz Attractor

We begin our study of Lyapunov exponents with the Lorenz equations:

$$\begin{aligned}
\dot{x} &= \sigma(y - x), \\
\dot{y} &= Rx - y - xz, \\
\dot{z} &= xy - bz.
\end{aligned}$$

Here σ, R, $b > 0$ are parameters. Edward Lorenz derived this 3D system from a simplified model of convection rolls in the atmosphere. Roughly one can say that the equations describe the flow of a fluid in a box which is heated along the bottom. This simple-looking deterministic system can have extremely erratic dynamics, the solutions oscillate irregularly over a wide range of parameters. In his original experiments he fixed the values of the parameters to $\sigma = 10$, $R = 28$, and $b = 8/3$ for which the system has chaotic behavior. These are the parameter values we will be using in the comparative study between the optimization approach described in [25] and the algorithms in [28, 34]. Figure 16.2a shows a 3D view of the Lorenz system.
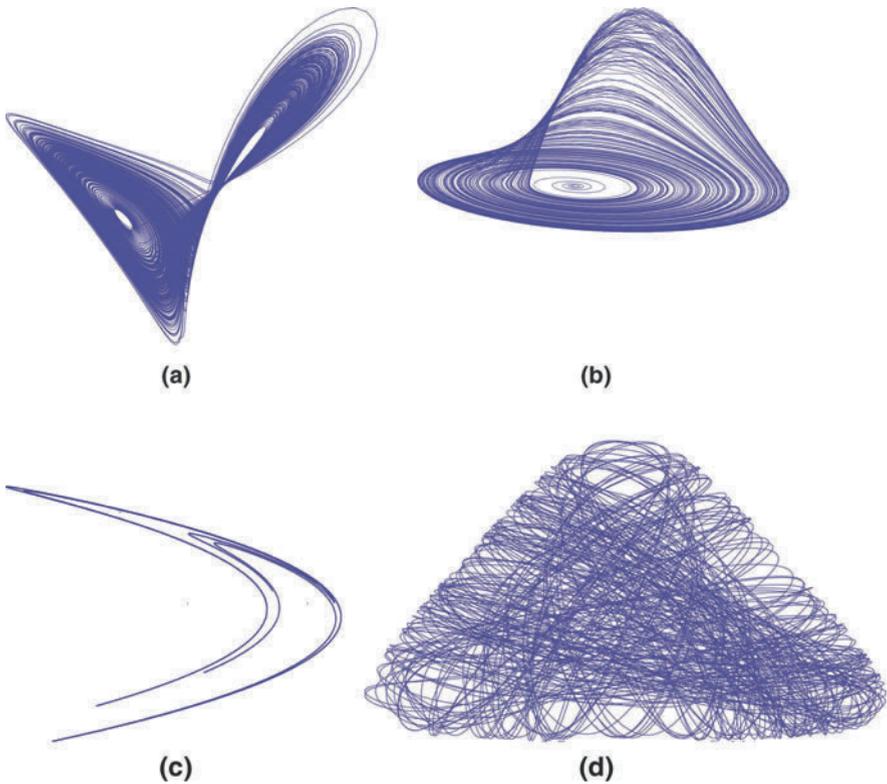

(a) (b)

(c) (d)

Fig. 16.2: The algorithm has been tested on the (**a**) Lorenz attractor, (**b**) Rössler attractor, (**c**) Hénon map and (**d**) Hénon–Heiles system (plotted in the x–y plane).

16.4.2 Rössler Attractor

In 1976, the Swiss mathematician Otto Rössler was studying oscillations in chemical reactions and discovered another set of equations with a chaotic attractor:

$$\dot{x} = -y - z$$
$$\dot{y} = x + ay$$
$$\dot{z} = b + z(x - c).$$

Both the Lorenz and Rössler equations are involved with the study of Navier–Stokes equations. Rössler is acclaimed to have used the parameter values $a = 0.2$, $b = 0.2$, and $c = 5.7$, which we will also use. This system of equations looks easier than the Lorenz system, with only one nonlinearity xz, but it is harder to analyze. Figure 16.2b shows the Rössler attractor.

16.4.3 Hénon Map

The Hénon map was devised by the theoretical astronomer Michel Hénon to illuminate the microstructure of strange attractors in 1976. Previous scientists had encountered numerical difficulties when tackling the Lorenz system, so instead Hénon sought a mapping that captured its essential features but which also had an adjustable amount of dissipation. Hénon chose to study mappings rather than differential equations because maps are faster to simulate and their solutions can be followed more accurately and for a longer time. The Hénon map is given by

$$x_{n+1} = y_n + 1 - ax_n^2$$
$$y_{n+1} = bx_n,$$

where a and b are adjustable parameters which are chosen as $a = 1.4$, $b = 0.3$.

16.4.4 The Hénon–Heiles Equations

The Hénon–Heiles model was introduced in 1964 by Michel Hénon and Carl Heiles as a model for the motion of a star inside a galaxy. With the Hamiltonian

$$H = \frac{1}{2}\left(p_1^2 + q_1^2 + p_2^2 + q_2^2\right) + q_1^2 q_2 - \frac{1}{3}q_2^3,$$

and if we let $q_1 = x$, $q_2 = y$, $p_1 = p_x$, and $p_2 = p_y$, then the Hamiltonian can be interpreted as a model for a single particle moving in two dimensions under the action of a force described by a potential energy function $V(x,y)$ [9]. Hamilton's equations for this system lead to the following equations for the dynamics of the system:

$$\dot{x} = \frac{\partial H}{\partial p_x} = p_x \qquad \dot{p}_x = -\frac{\partial H}{\partial x} = -x - 2xy$$

$$\dot{y} = \frac{\partial H}{\partial p_y} = p_y \qquad \dot{p}_y = -\frac{\partial H}{\partial y} = -y - x^2 + y^2.$$

It can be shown that the potential supports bounded motion for the particle for
H< 1/6. Thus, to fulfill this condition, the initial values are chosen as $x_0 = 0$, $y_0 = -0.15$, $p_{x,0} = 0.50$, and $p_{y,0} = 0$.

16.5 Computational Experiments

This section presents the main results of the implemented algorithm and compares
these with two other algorithms described in [28, 34]. Some enhancements are made
to the algorithm, which is finally tested for robustness against change in parameter
values and noise.

16.5.1 Numerical Computations

The differential equations are integrated with the *Runge–Kutta (RK4)* method using
a fixed step δt. This method is reasonably simple and robust, even without the adap-
tive step-size routine. The RK4 method is a fourth-order method, meaning that the
error per step is $\mathcal{O}((\delta t)^5)$, while the total accumulated error has order $\mathcal{O}((\delta t)^4)$.
    Table 16.2 summarizes the values for the computed Lyapunov exponents, which
has been estimated using three different implemented algorithms described in

Table 16.2: Results of the preliminary computational experiments for $n = 2,000$. For the Lorenz
attractor, the parameter values have been chosen as: $\tau = 9$, $\Delta t = 5$, $\delta t = 0.01$, $\varepsilon = 1.20$. The
parameter values for the Rössler attractor are: $\tau = 6$, $\Delta t = 5$, $\delta t = 0.12$, $\varepsilon = 0.28$

System with		$n = 2,000$			$n = 4,000$		
initial condition		Pardalos	Sano	Wolf	Pardalos	Sano	Wolf
Lorenz							
$x_0 = 0$	λ_1	1.08547	1.08546	0.86241	1.15313	1.15312	0.92140
$y_0 = 1.0$	λ_2	−0.30421	−0.30421		−0.31226	−0.31232	
$z_0 = 0$	λ_3	−10.61753	−10.61754		−10.57328	−10.57319	
Rössler							
$x_0 = 0.1$	λ_1	0.06928	0.06924	0.06642	0.06891	0.06881	0.06593
$y_0 = 0.1$	λ_2	0.00132	0.00156		0.00122	0.00131	
$z_0 = 0.1$	λ_3	−1.27202	−1.27183		−1.28908	−1.28849	
Hénon							
$x_0 = 0.1$	λ_1	0.41672	0.41672	0.40445	0.41669	0.41672	0.41547
$y_0 = 0.1$	λ_2	−1.57647	−1.57647		−1.57765	−1.57773	
Hénon–Heiles							
$x_0 = 0$	λ_1	0.15207	0.15209	0.16012	0.14875	0.14874	0.15493
$y_0 = -0.15$	λ_2	0.01950	0.01949		0.01382	0.01385	
$p_{x,0} = 0.50$	λ_3	−0.04242	−0.04245		−0.04874	−0.04871	
$p_{y,0} = 0$	λ_4	−0.23309	−0.23395		−0.22142	−0.22141	

[24, 28, 34]. The rows show the Lyapunov exponents in a decreasing manner, i.e., $\lambda_1 > \lambda_2 > \ldots > \lambda_n$, for given initial conditions. The algorithm by Wolf et al. [34] only gives us an estimate of the largest exponent. As mentioned earlier, a positive Lyapunov exponent measures sensitive dependence on initial conditions, or how much our forecasts can diverge based upon different estimates of starting conditions. Another way to view Lyapunov exponents is the loss of predictability as we look forward in time. Thus, it is interesting to know a measure of information loss for avoiding possible misinterpretations.

If we assume that the true starting point x_0 of a time series is possibly displaced by an ε, we know only the information area I_0 about the starting point. After some steps the time series is in the information area at time t, I_t. The information about the true position of the data decreases due to the increase of the information area. Consequently, we get a bad predictability. The largest Lyapunov exponent can be used for the description of the average information loss; $\lambda_1 > 0$ leads to bad predictability. Therefore, the exponent values in Table 16.2 are given in units of nats/s.[4]

Of all the N displacement vectors found inside the sphere of radius ε, only five to seven vectors with the smallest norm are chosen. This has practically no noticeable effect on the exponent values, but speeds up the algorithm. It is further enhanced by introducing another constraint which enables us to search for displacement vectors close in phase space (Equation (16.13)), but far away in time

$$|t_j - t_i| > \frac{\varepsilon}{\delta t}, \quad \forall i, j, \quad i \neq j. \tag{16.27}$$

The Gauss–Newton algorithm is used to solve the nonlinear least-squares problem in Equation (16.16), while Sano [28] uses a linear approach to solve the same problem. By examining Table 16.2 we see that there is hardly any difference in the estimated exponent values between Pardalos's algorithm and the one described by Sano. This behavior is due to the small values of the evolution time Δt. During this short evolution, the mapping between t_j and $t_j + \Delta t$ does not show any stronger nonlinear properties, therefore, the results are similar. The value Δt should be kept small enough so that orbital divergence is monitored at least a few times[5] per (mean) orbit. A larger Δt has been shown to increase the difference between these two algorithms, as expected.

The displacement vectors y_i have been chosen to lie inside a sphere of radius $\varepsilon \lesssim 0.02 L_A$, where L_A is the horizontal extent of the attractor. The choice of ε is good as long as we fulfill the condition of finding a minimum of five vectors inside the sphere. Though theory says this value should be infinitesimal, the optimization algorithm described in Section 16.2 is robust against small increase in ε. Figure 16.3 shows how the Lyapunov exponents for the examined systems converge.

The results from Pardalos's and Sano's algorithms, though different from the estimated values computed by the Wolf algorithm, are in good agreement with other numerical experiments performed in [30, 27, 4, 26, 8].

---

[4] 1 nat/s $\approx$ 1.44 bits/s.

[5] We have computed Equation (16.16) between 30 and 40 times per mean orbit.
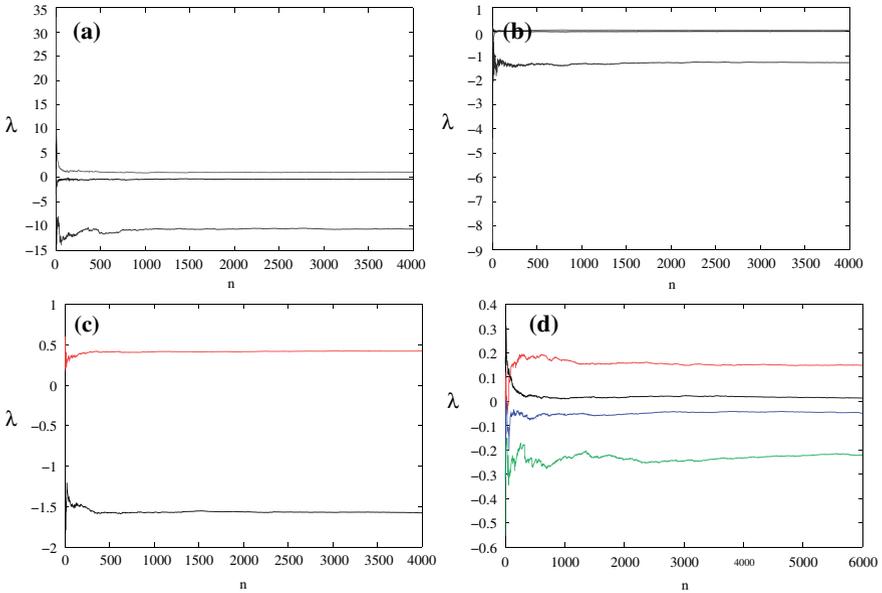
Fig. 16.3: Convergence of the spectrum of Lyapunov exponents for (**a**) the Lorenz attractor, (**b**) Rössler attractor, (**c**) Hénon map, and (**d**) the Hénon–Heiles system. The graphs show the results of the algorithm described in Section 16.2.

16.5.2 Sensitivity Analysis

Two parameters, namely the evolution time Δt and the time delay τ, have been chosen for further investigation for robustness. We mentioned in Section 2.2.2 that τ is determined as the lag which gives us the first minimum for the mutual average information for our observed data. Figure 16.4 shows how the spectrum of Lyapunov

Fig. 16.4: The Lyapunov exponents as a function of (**a**) τ and (**b**) Δt for the Lorenz attractor.

exponent values depend on Δt and τ. The two largest Lyapunov exponents, λ_1 and λ_2, seem to be rather stable in the vicinity of the chosen values $\Delta t = 5$ and $\tau = 9$ while λ_3 is unstable for all values in the interval $1 < \Delta t < 30$, $1 < \tau < 40$. What is important to be reminded of here is that the systems are extremely sensitive due to their chaotic nature, and the observed "errors" due to perturbations of parameter values do not have to be entirely blamed on this specific algorithm. The science of choosing the right parameter values for these kind of problems is not general and depends on which system you are examining.

Figure 16.5 shows how the Lyapunov exponents for the Rössler attractor depend on τ and Δt. Again, λ_1 and λ_2 are stable in the neighborhood of the chosen values $\tau = 6$ and $\Delta t = 5$, while λ_3 is very irregular throughout the interval. This behavior is given a deeper theoretical explanation in [28].
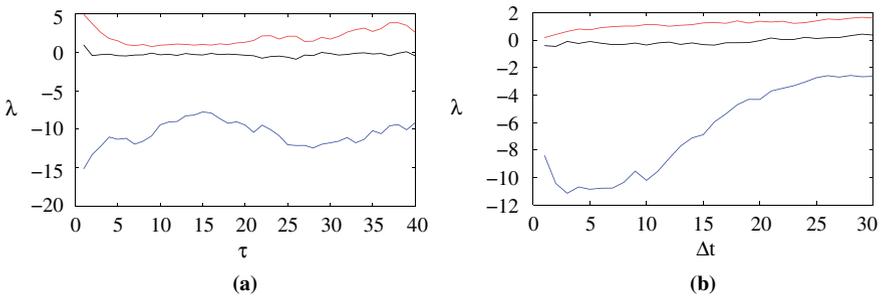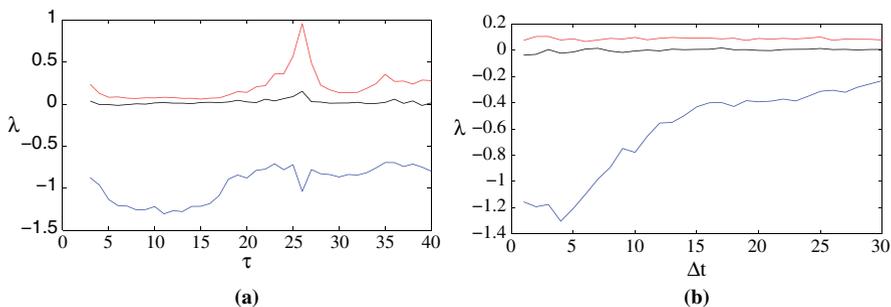

Fig. 16.5: The Lyapunov exponents as a function of (**a**) τ and (**b**) Δt for the Rössler attractor.

The algorithm has also been tested for noise contaminated data. We have added Gaussian white noise $w(t)$ to the solutions of the Lorenz and Rössler systems according to

$$x_{\text{noise}} = x_{\text{clean}} + w(t)\sigma(x)s$$
$$y_{\text{noise}} = y_{\text{clean}} + w(t)\sigma(y)s$$
$$z_{\text{noise}} = z_{\text{clean}} + w(t)\sigma(z)s,$$

where s is a scaling factor for the standard deviation σ.

Figure 16.6 shows the dependence between the Lyapunov exponents and the scaling factor within the interval $0 \leq s \leq 0.1$. We see that the exponents λ_1 and λ_2 for both systems are quite stable within the interval $0 < s < 0.01$, i.e., they are not sensitive to data contaminated with up to 0.01σ of Gaussian white noise. Once again we can confirm the sensitive nature of the smallest Lyapunov exponent λ_3 of the Lorenz attractor.

Many physical signals, including the time series studied in this chapter, are fundamentally different from linear time-invariant signals in that they are invariant to
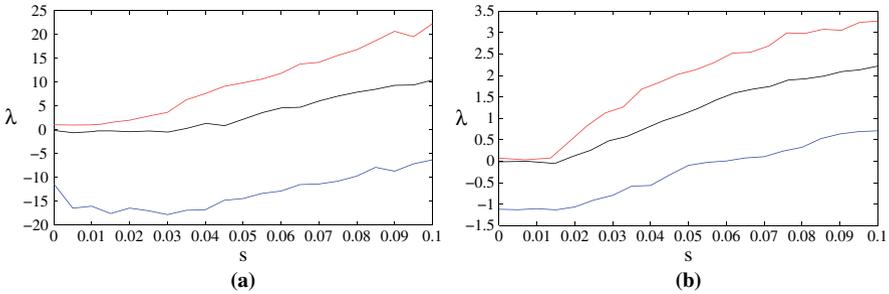
Fig. 16.6: The Lyapunov exponent values as a function of the scaling factor *s* for (**a**) the Lorenz attractor and (**b**) the Rössler attractor.

scale rather than to translation. The signals are often mixed with noise, and the separation may be very difficult if both the signal and the noise are broadband. The problem becomes inherently difficult when the signal is chaotic because its power spectrum is indistinguishable from a broadband noise, as in our case. Since there is a strong relationship between these *fractal* signals and the wavelet transform, the latter appears to be the natural signal processing technique, just as the Fourier transform is natural for the linear time-invariant signals [14]. Many new filtering techniques to handle these problems are still under development.

16.6 Summary and Conclusion

The Lyapunov exponents are conceptually the most basic indicators of deterministic chaos of dynamical systems. For the analysis of such dynamics, many numerical algorithms to determine the spectrum of the Lyapunov exponents have been proposed. In this chapter, we considered an optimization technique for calculating tangent maps with the aim of developing a robust algorithm. We have described a method which is shown to behave well in the perturbation of certain parameter values, but slightly sensitive in the presence of noise. This method uses the Gauss–Newton algorithm to solve the least-squares problem that arises, which is no more complicated to implement than the linear method. By using the new optimization method, we could obtain good estimates of the Lyapunov spectrum from the observed time series in a very systematic way.

References

1. Abarbanel, H., Brown, R., Kennel, M. Variation of Lyapunov exponents in chaotic systems: Their importance and their evaluation using observed data. J Nonlinear Sci **2**, 343–365 (1992)
2. Chen, G., Lai, D. Making a dynamical system chaotic: Feedback control of Lyapunov exponents for discrete-time dynamical systems. IEEE Trans Circuits Syst I Fundam Theory Appl **44**, 250–253 (1997)

3. Cvitanovic, P., Artuso, R., Mainieri, R., Tanner, G., Vattay, G. Classical and Quantum Chaos. 10th ed., ChaosBook.org Niels Bohr Institute, Copenhagen (2003) Webbook: http://chaosbook.org/.
4. Djamai, L., Coirault, P. Estimation of Lyapunov exponents by using the perceptron. Proceedings of the American Control Conference **6**, 5150–5155 (2002)
5. Eckmann, J.P., Kamphorst, S.O., Ruelle, D., Ciliberto, S. Lyapunov exponents from time series. Phys Rev A **34**, 4971–4979 (1986)
6. Eckmann, J.P., Ruelle, D. Ergodic theory of chaos and strange attractors. Rev Modern Phys **57**, 617–657 (1985)
7. Elger, C.E., Lehnertz, K. Seizure prediction by non-linear time series analysis of brain electrical activity. Eur J Neurosci **10**, 786–789 (1998)
8. Golovko, V. Estimation of the Lyapunov spectrum from one-dimensional observations using neural networks. Proceedings of the Second IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 95–98 (2003)
9. Hilborn, R.C. Chaos and Nonlinear Dynamics. Oxford University Press, New York (2000)
10. Iasemidis, L.D., Pardalos, P.M., Sackellares, J.C., Yatsenko, V.A. Global optimization approaches to reconstruction of dynamical systems related to epileptic seizures. In: Fotiadis, D., Massalas, C.V. (eds.) Scattering and Biomedical Engineering: Modeling and Applications, World Scientific, Singapore, pp. 308–318 (2002)
11. Kantz, H., Schreiber, T. Nonlinear Time Series Analysis. Cambridge University Press, New York (1997).
12. Kelliher, J.: Lyapunov Exponents and Oseledec's Multiplicative Ergodic Theorem. http://www.ma.utexas.edu/users/kelliher/Geometry/Geometry.html (2005)
13. Kennel, M.B., Brown, R., Abarbanel, H.D.I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A **45**, 3403–3411 (1992)
14. Kinsner, W. Characterizing chaos through Lyapunov metrics. Second IEEE International Conference on Cognitive Informatics (ICCI'03), 189–201 (2003)
15. Lehnertz, K., Andrzejak, R.G., Arnhold, J., Kreuz, T., Mormann, F., Rieke, C., Widman, G., Elger, C. Nonlinear EEG analysis in epilepsy: Its possible use for interictal focus localization, seizure antipilation, and prevention. J Clin Neurophysiol **18**, 209–222 (2001)
16. Nair, S. Brain Dynamics and Control with Applications in Epilepsy. Dissertation at the University of Florida, Gainesville, FL (2006)
17. Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S. Geometry from a time series. Phys Rev Lett **45**, 712–715 (1980)
18. Pardalos, P.M., Boginski, V., Vazakopoulos, A. (eds.) Data Mining in Biomedicine. Springer, New York (2007)
19. Pardalos, P.M., Chaovalitwongse, W., Iasemidis, L.D., Sackellares, J.C., Shiau, D-S., Carney, P.R., Prokopyev, O.A., Yatsenko, V.A. Seizure warning algorithm based on optimization and nonlinear dynamics. Math Program **101**, 365–385 (2004)
20. Pardalos, P.M., Principe, J. (eds.) Biocomputing. Kluwer Academic Publishers, Dordrecht (2002)
21. Pardalos, P.M., Sackellares, C., Carney, P., Iasemidis, L. (eds.) Quantitative Neuroscience. Kluwer Academic Publishers, Dordrecht (2004)
22. Pardalos, P.M., Sackellares, J.C., Iasemidis, L.D., Yatsenko, V.A., Yang, M., Shiau, D.-S., Chaovalitwongse,W. Statistical information approaches to modelling and detection of the epileptic human brain. Comput Stat Data Anal **43**(1), 79–108 (2003)
23. Pardalos, P.M., Sackellares, J.C., Yatsenko, V.A., Butenko, S.I. Nonlinear dynamical systems and adaptive filters in biomedicine. Ann Oper Res **119**, 119–142 (2003)
24. Pardalos, P.M., Yatsenko, V.A. Optimization approach to the estimation and control of Lyapunov exponents. J Optim Theory Appl **128**, 29–48 (2006)
25. Pardalos, P.M., Yatsenko, V.A., Sackellares, J.C., Shiau, D.-S., Chaovalitwongse, W., Iasemidis, L. Analysis of EEG data using optimization, statistics, and dynamical systems techniques. Comput Stat Data Anal **44**, 391–408 (2003)
26. Ramasubramanian, K., Sriram, M.S. A comparative study of computation of Lyapunov spectra with different algorithms. Physica D **139**, 72–86 (2000)

27. Rosenstein, M.T., Collins, J.J., De Luca, C.J. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D **65**, 117–134 (1993)
28. Sano, M., Sawada, Y. Measurement of the Lyapunov spectrum from a chaotic time series. Phys Rev Lett **55**, 1082–1085 (1985)
29. Serfaty DeMarkus, A. Detection of the onset of numerical chaotic instabilities by Lyapunov exponents. Discrete Dyn Nat Soc, **6**, 121–128 (2001)
30. Shimada, I., Nagashima, T. A Numerical approach to ergodic problem of dissipative dynamical systems. Progr Theor Phys, **61**, 1605–1616 (1979)
31. Strogatz, S.H. Nonlinear Dynamics and Chaos. Perseus Books Publishing, Cambridge, MA (1994)
32. Takens, F. Detecting strange attractors in turbulence. In: Rand, D.A., Young, L.-S. (eds.) Dynamical Systems and Turbulence, *Lecture Notes in Mathematics*, Vol. **898**, pp. 366–381. Springer-Verlag, New York (1981)
33. Wiesel,W.E. Modal feedback control on chaotic trajectories. Phys Rev E **49**, 1990–1996 (1994)
34. Wolf, A., Swift, J.B., Swinney, H.L., Vastano, J.A. Determining Lyapunov exponents from a time series. Physica D **16**, 285–317 (1985)
35. Zeni, A.R., Gallas, J.A.C. Lyapunov exponents for a Duffing oscillator. Physica D **89**, 71–82 (1995)

Chapter 17

Dynamical Analysis of the EEG and Treatment of Human Status Epilepticus by Antiepileptic Drugs

Aaron Faith, Shivkumar Sabesan, Norman Wang, David Treiman, Joseph Sirven, Konstantinos Tsakalis, and Leon Iasemidis

Abstract An estimated 42,000 epileptic patients die from status epilepticus (SE) every year in the United States alone. Evaluation of antiepileptic drugs and protocols for SE treatment, in terms of the dynamics of concurrently monitored electroencephalogram (EEG), may lead to the design of new, more effective treatment paradigms for successfully controlling SE. Such monitoring techniques may have a profound effect in the treatment of SE in the emergency department (ED) and intensive care unit (ICU), where antiepileptic drugs (AEDs) are given in rapid succession in the hope of patient recovery, or even in the epilepsy monitoring unit (EMU), where occasionally a patient may progress to SE. In the past, using techniques from nonlinear dynamics and synchronization theory, we have shown that successful treatment with AEDs results in dynamical disentrainment (desynchro-

Aaron Faith
The Harrington Department of Bioengineering, Arizona State University, Tempe, AZ 85287, USA, e-mail: `atfaith@asu.edu`

Shivkumar Sabesan
The Harrington Department of Bioengineering, Arizona State University, Tempe, AZ 85287, USA; Barrow Neurological Institute, Phoenix, AZ 85013, USA, e-mail: `ssabesa@asu.edu`

Norman Wang
Barrow Neurological Institute, Phoenix, AZ 85013, USA

David Treiman
Barrow Neurological Institute, Phoenix, AZ 85013, USA, e-mail: `dtreiman@chw.edu`

Joseph Sirven
Mayo Clinic, Phoenix, AZ 85054, USA, e-mail: `joseph.sirven@mayo.edu`

Konstantinos Tsakalis
Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA, e-mail: `tsakalis@asu.edu`

Leon Iasemidis
The Harrington Department of Bioengineering, Arizona State University, Tempe, AZ 85287, USA; Mayo Clinic, Phoenix, AZ 85054, USA; Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA, e-mail: `leon.iasemidis@asu.edu`

nization) of entrained brain sites in SE, a phenomenon we have called dynamical resetting. We herein apply this nonlinear dynamical analysis to scalp EEG recordings from two patients, one admitted to the EMU and the other to the ED and ICU and both treated with AEDs, to show that successful administration of AEDs dynamically disentrains the brain and correlates well with the patients' recovery. This result further supports our hypothesis of dynamical resetting of the brain by AEDs into the recovery regime, and indicates that the proposed measures/methodology may assist in an objective evaluation of the efficacy of current and the design of future AEDs for the treatment of SE.

17.1 Introduction

Status epilepticus (SE) is a life-threatening neurological emergency. SE is characterized by recurrent epileptic seizures without recovery of normal brain function between seizures. Out of the 200,000 cases of SE diagnosed each year in the United States, the 30-day and 60-day mortality rate in the adult cases are well into the 40% range [2]. It is estimated that SE accounts for more than $4B annual health-care costs in USA alone. SE affects all age groups, with higher morbidity and mortality in older aged adults.

The most perplexing aspect about clinical management of SE is that SE can become refractory to initial, or sometimes any, treatment. In such cases, prompt treatment is the key to preventing catastrophic outcomes. It has been shown that mortality in children and adults is minimized when SE lasts less than 1 h; however, thereafter, the odds of mortality jump dramatically to close to 38% [19]. Therefore, the goal of SE treatment is to stop the seizure activity as quickly as possible. The clinical standard for deciding a successful clinical response of SE to AED treatment is by visual inspection of EEG to determine complete cessation of all seizure (ictal) activity. Typically, in SE patients who respond to AED medication, successful cessation of ictal EEG activity occurs within 20 min following AED treatment. On the other hand, in SE patients who do not respond to AED treatment, patterns of ictal EEG activity continue or reappear within 60 min following AED treatment [18]. Unfortunately, all too frequently, it is extremely difficult to differentiate such EEG patterns from those associated with other abnormalities, such as metabolic encephalopathy. Moreover, it is also difficult to distinguish the relapse of SE ictal activity due to wearing away of a treatment from other abnormal non-SE EEG patterns. Therefore, an independent measure of "ictalness," that could help differentiate morphological patterns on the EEG that appear ictal is needed. If a brain dynamical analysis correlates well with the electroencephalographer's assessment of the presence or absence of ictal patterns on the EEG, it could lead to the development of a clinically useful tool that might independently from the visual analysis of the EEG determine the presence or absence of SE.

In the last decade, substantial progress has been made toward the study of the human brain by utilizing concepts and measures from nonlinear dynamics [6]. Within this framework, a significant amount of effort has been made toward understanding

the mechanisms underlying the spontaneous initiation and termination of epileptic seizures. The central concept is that seizures represent transitions of the epileptic brain from its "normal" (less ordered/chaotic) state to an abnormal (more ordered) state and back to a "normal" state, along the lines of spatiotemporal chaos-to-order-to-chaos transitions. The hallmark of this research is the ability to predict epileptic seizures, in the order of tens of minutes prior to their clinical or electrographic onset. This research has provided useful insights into the progressive preictal (before a seizure) entrainment, and the subsequent post-ictal (after a seizure) disentrainment of the epileptic brain's spatiotemporal EEG activity, under the hypothesis of "dynamical resetting of the epileptic brain" [12, 16]. According to this hypothesis, seizures do not occur as long as there is no need for the brain to reset. In status epilepticus though, seizures may continue to occur as the entrainment of normal brain sites with the focus persists and the internal seizure resetting mechanism is not effective enough to disrupt it. Thus, the brain typically resets its dynamics after the occurrence of a seizure except when it is confined in status epilepticus.

In this study, we further validate that SE is due to the non-resetting of the pathology of the dynamics of epileptic brain's electrical activity. This pathology of dynamics is characterized by an intense and long-term entrainment (the term synchronization may be used selectively herein instead of entrainment) of the dynamics of normal brain sites with the ones of the epileptogenic focus (foci) and could be reset by successful external intervention. Our preliminary results from mathematical analysis of the available "almost continuous" and "relatively short" scalp EEG, in two patients with SE, one from each of two participating medical centers, show that the above described pathology of epileptic brain dynamics can be reset by external successful intervention, such as administration of antiepileptic drugs (AEDs).

The organization of the rest of this chapter is as follows. The EEG data and the measures of brain dynamics utilized for the analysis of EEG are described in Section 17.2. In Section 17.3, results from the application of this analysis to scalp EEG data from two patients with SE are presented. Discussion of these results and conclusions are given in Section 17.4.

17.2 Materials and Methods

17.2.1 Recording Procedure and EEG Data

We test our dynamical resetting hypothesis on EEG data from two epilepsy centers, namely the Barrow Neurological Institute in Phoenix Arizona, and the Mayo Clinic Hospital in Scottsdale, Arizona. Two patients (one from each center), who had an episode of SE and were subsequently treated successfully via AEDs, were chosen for dynamical analysis of their stored, "almost continuous," scalp EEG recordings. The available recordings were of duration of about 2 h in one patient and 14 h in the other. The EEGs were analyzed with the methodology described in the next section. We have shown in the past [9, 8, 7], that EEG segments of 10.24 s in duration would be sufficient for the estimation of measures of dynamics from the nonstationary

EEG in epilepsy. By choosing such apparently "long" segments we have been able to detect spatiotemporal changes of dynamics over time that lead to prediction of epileptic seizures tens of minutes ahead of their onset, at a respectable degree of sensitivity and specificity [10, 11].

17.2.1.1 EEG from Barrow Neurological Institute, Phoenix, Arizona

Patient 1 was a 6-year-old patient admitted to the Epilepsy Monitoring Unit at Barrow Neurological Institute, St. Joseph's Hospital, Phoenix, AZ. The EEG was recorded with a standard International 10–20 scalp electrode montage (see Fig. 17.1) at an A/D rate of 400 Hz. At the beginning of the recording, the EEG was characterized by continuous ictal discharges associated with SE stage III and progressed into ictal discharges punctuated by periods of flattening characteristic of SE stage IV. During the EEG recording of SE, the attending physicians administered two AEDs. The first AED (diazepam 10 mg) was administered rectally at 18 min into the recording; the second AED (lorazepam 0.1 mg/kg) was administered intravenously at 54 min into the recording. The proprietary EEG data were converted into 16-bit signed binary format for further off-line nonlinear dynamical analysis.
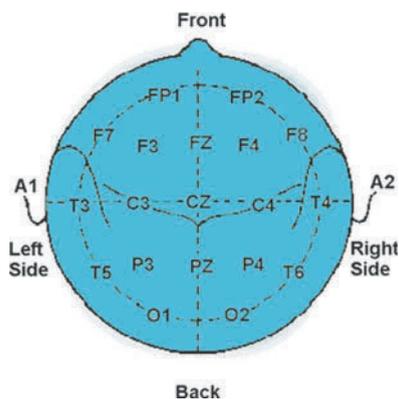
Fig. 17.1: Schematic diagram showing a standard scalp electrode placement, according to the international 10–20 system as seen from above the head. A = Ear lobe, C = central, P = parietal, F = frontal, Fp = frontal polar, O = occipital.

17.2.1.2 EEG from Mayo Clinic Hospital, Scottsdale, Arizona

The patient from this medical center (Patient 2) was a 75-year-old male with no history of seizures, brought to the emergency department after he was found unresponsive at home with his right arm twitching and his head deviated to his right. When the patient arrived at the emergency department he was observed having rhythmic

contractions in his right thigh adductor muscles, followed by tonic–clonic movements on his right side and right gaze deviation and nystagmus. The patient was initially given 2 mg of Ativan IV followed by 20 mg/kg of phenytoin equivalence IV and a second fosphenytoin bolus of 5 mg/kg. Approximately half an hour after the beginning of the EEG recording, the patient was transferred to the ICU. The patient did not respond to the fosphenytoin, so he was intubated and placed on a propofol drip. Soon after, the EEG showed a burst suppression pattern. The last EEG file was recorded when the patient remained off propofol. The etiology of the patient's seizures was noted as encephalomalacia in the left frontal cortex likely due to a prior stroke. The EEG was recorded using 24 scalp electrodes that were placed according to the standard international 10–20 montage (see Fig. 17.1) at a sampling rate of 200 Hz. Due to the emergency nature of the clinical situation, the recorded EEG data were available in five separate files (A through E) with some gaps (e.g., when recording was stopped in order to transport the patient) between the recordings. These files were of 1.10, 3.54, 3.03, 3.03, and 2.81 h in duration respectively (i.e., total duration of 13.51 h). The proprietary EEG data were converted into 16-bit signed binary format for further off-line nonlinear dynamical analysis. Each file was first analyzed separately using the measures of brain dynamics that are described next.

17.2.2 Measures of Brain Dynamics

17.2.2.1 Measure of Chaos(STL_{max})

Under certain conditions, through the method of delays described by Packard et al. [14] and Takens [17], sampling of a single variable of a system over time can determine all state variables of the system that are related to an observed state variable. In the case of the EEG, this method can be used to reconstruct a multidimensional state space of the brain's electrical activity from a single EEG electrode that referentially records from a brain site. Thus, in such an embedding, each state in the state space is represented by a vector $\mathbf{X}(t)$, whose components are the delayed versions of the original single-channel EEG time series $x(t)$, that is:

$$\mathbf{X}(t) = (x(t), x(t+\tau), \ldots, x(t+(d-1)\tau)), \tag{17.1}$$

where τ is the time delay between successive components of $\mathbf{X}(t)$, and d is a positive integer denoting the embedding dimension of the reconstructed state space. Plotting $\mathbf{X}(t)$ in the created state space produces the state portrait of a spatially distributed system using the subsystem (brain's portion) where $x(t)$ is recorded from. The most complicated steady state a nonlinear deterministic system can exhibit is a strange and chaotic attractor, whose complexity is measured by its dimension D, and its chaoticity by its Kolmogorov entropy (K) and Lyapunov exponents (Ls) [4, 3]. A steady state is chaotic if at least the maximum of these Lyapunov exponents ($L\text{max}$) is positive.

According to Takens, in order to properly embed a signal in the state space, the embedding dimension d should at least be equal to $(2D + 1)$. Of the many different methods used to estimate D of an object in the state space, each has its own practical problems [13]. The measure most often used to estimate D is the state space correlation dimension v. Methods for calculating v from experimental data have been described in [1] and were employed in our work to approximate D in the ictal state. The brain, being nonstationary, is not expected to be in a steady state in the strict dynamical sense at any location. Arguably, activity at brain sites is constantly moving through steady states, which are functions of certain parameter values at a given time. According to bifurcation theory [5], when these parameters change slowly over time (e.g., when the system is close to a bifurcation), dynamics slow down and conditions of stationarity are better satisfied. In the ictal state, temporally ordered and spatially synchronized oscillations in the EEG usually persist for a relatively long period of time (in the range of minutes). Dividing the ictal EEG into short segments ranging from 10.24 to 50 s in duration, estimation of v from ictal EEG has produced values between 2 and 3 [7], implying the existence of a low-dimensional manifold in the ictal state, which we have called "epileptic attractor." Therefore, an embedding dimension d of at least 7 has to be used to properly reconstruct this epileptic attractor.

Although d of interictal (between seizures) EEG data is expected to be higher than that of the ictal state, a constant embedding dimension $d = 7$ has been used to reconstruct all relevant state spaces over the ictal and interictal periods at different brain locations. The advantages of this approach are: (a) existence of irrelevant information in dimensions higher than 7 might not influence much the estimated dynamical measures, and (b) reconstruction with high d requires longer data segments, which may interfere with the nonstationary nature of the EEG. The disadvantage is that possibly existing relevant information about the transition to seizures in higher than $d = 7$ dimensions may not be captured.

The Lyapunov exponents measure the information flow (bits/s) along local eigenvectors as the system moves through such attractors. Theoretically, if the state space is of d dimensions, we can estimate up to d Lyapunov exponents. However, as expected, only $D + 1$ of these will be real. The others are spurious [15]. Methods for calculating these dynamical measures from experimental data have been published in [8]. The estimation of the largest Lyapunov exponent (Lmax) in a chaotic system has been shown to be more reliable and reproducible than the estimation of the remaining exponents [20], especially when D is unknown and changes over time, as in the case of high-dimensional and nonstationary data (e.g., interictal EEG). A method to estimate an approximation of Lmax from nonstationary data, called STL (short-term Lyapunov) [8, 7], has been developed via a modification of the Wolf's algorithm used to estimate Lmax from stationary data [21]. The STL_{max} algorithm is applied to sequential EEG segments recorded from electrodes in multiple brain sites to create a set of STL_{max} profiles over time (one STL_{max} profile per recording site) that characterize the spatiotemporal chaotic signature of the epileptic brain. The consistent observation across seizures and patients is the convergence of STL_{max} values between electrode sites prior to seizures. We have called this phe-

nomenon *dynamical entrainment* (synchronization), and it has constituted the basis for the development of epileptic seizure prediction algorithms.

17.2.2.2 Measure of Dynamical Entrainment

A statistical measure of synchronization of the dynamics between two electrodes i and j has been developed in the past. Specifically, the T_{ij} between electrode sites i and j for a measure of dynamics (e.g., $\text{STL}_{\max}$) at time t is defined as

$$T_{ij}^t = \frac{|\overline{D}_{ij}^t|}{\hat{\sigma}_{ij}^t / \sqrt{m}}, \tag{17.2}$$

where $\overline{D}_{ij}^t$ and $\hat{\sigma}_{ij}^t$ denote the sample mean and standard deviation respectively of all m differences between a measure's values at electrodes i and j within a window $w_t = [t, t - m^* 10.24\,\text{s}]$ moving over the measure's profiles. If the true mean μ_{ij}^t of the differences D_{ij}^t is equal to 0, and σ_{ij}^t are independent and normally distributed, T_{ij}^t is asymptotically distributed as the t-distribution with $(m-1)$ degrees of freedom. We have shown that these independence and normality conditions are satisfied for $\text{STL}_{\max}$ [10]. Therefore, we define desynchronization between electrode sites i and j when T_{ij} is significantly different from 0 at a significance level α. The desynchronization condition between the electrode sites i and j, as detected by the paired t-test, is

$$T_{ij}^t > t_{\alpha/2, m-1} = T_{\text{th}}, \tag{17.3}$$

where $t_{\alpha/2, m-1} = T_{\text{th}}$ is the $100(1 - \alpha/2)$ critical value of the t-distribution with $m-1$ degrees of freedom. If $T_{ij}^t \leq t_{\alpha/2, m-1}$ (which means that we do not have satisfactory statistical evidence at the α level for the differences of values of a measure between electrode sites i and j be nonzero within the time window w_t), we consider sites i and j be synchronized with each other at time t. Using $\alpha = 0.01$ and $m = 60$, the threshold $T_{\text{th}} = 2.662$.

We then estimate the average level of synchronization, as measured by the average T-index across all possible entrained (synchronized) pairs of electrode sites within a predefined time period over time. This average T-index is followed over time to monitor the response of SE to AED treatment. In the following sections, for simplicity, we denote these spatially averaged T-index values by "T-index." Results from this analysis in SE patients are presented next.

17.3 Results

From the EEG data per SE patient, the $\text{STL}_{\max}$ profiles per recording site and the T-index profiles over time per pair of recording sites were estimated. All synchronized pairs of electrode sites within 10 min prior to administration of the first drug
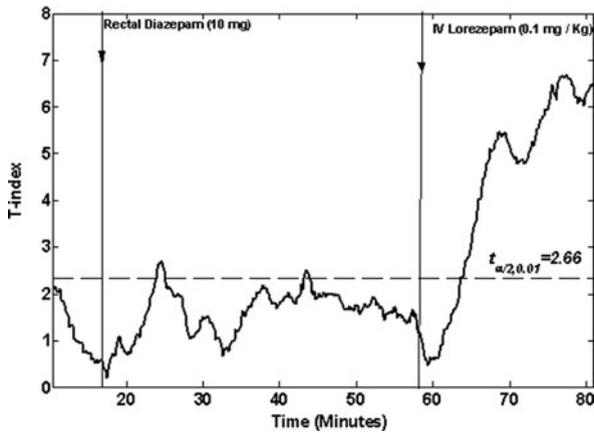
Fig. 17.2: AED resetting of brain dynamics in SE (Patient 1 – child). The average T-index profile of all entrained pairs of electrode sites, selected prior to SE onset, over time (80 min scalp EEG record). The patient developed SE while he was in the EMU at BNI undergoing presurgical evaluation. The patient was administered diazepam rectally 24 min into the recording, and 0.1 mg/kg of lorezepam intravenously 61 min into the recording (times of AEDs administration are denoted by the *vertical arrows*). Both drugs desynchronized the brain in the short term. However, the one administered intravenously had the maximum and most enduring effect on the desynchronization of the brain. The patient recovered by the end of this record.

were selected for the estimation of the average T-index and followed over time. For Patient 2, due to the discontinuity of the available EEG data, the STL_{max} and the T-index profiles were first estimated separately per available file using the approach given in Section 17.2; the results were then concatenated. Figure 17.2 shows resetting of brain dynamics (from low to high T-index values) after a successful administration of AEDs in Patient 1. Figure 17.3 illustrates a successful treatment and resetting of the EEG dynamics by AEDs in Patient 2 over a considerably longer period than the one in Patient 1.

During SE, and prior to the administration of the first AED, the average T-index was lower than the statistical threshold of entrainment, suggesting that the corresponding critical brain sites were entrained. Within minutes of the first AED treatment (diazepam in the first patient and fosphenytoin in the second patient), signs of disentrainment of the brain dynamics were observed in both patients. Within 10–20 min after the administration of the first AED, in both patients the brain became entrained again (T-index reversed its route toward above the statistical threshold of entrainment and started to assume lower values). The administration of subsequent AED(s) (lorazepam in the first patient, fentanyl and propofol in the second patient) dynamically disentrained the brain (statistically high T-index values were attained and sustained). Both patients recovered. Once the patients progressed out
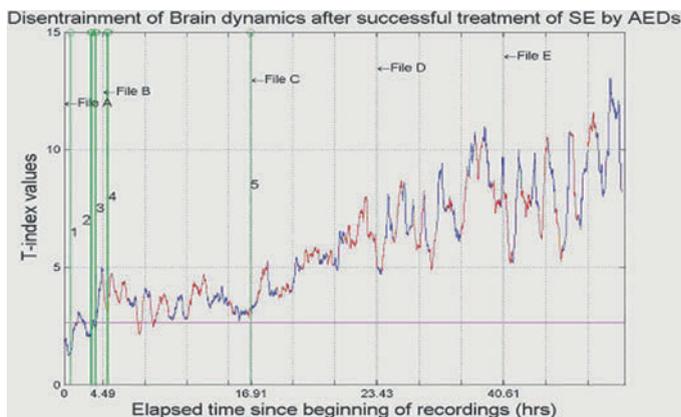
Fig. 17.3: Long-term AED resetting of the brain dynamics in SE (Patient 2 – adult). The effectiveness of AEDs on the patient's recovery is correlated well with the exhibited trends of the average T-index of all entrained pairs of brain sites, selected 10 min before the beginning of the treatment of the patient with the first AED (a scalp EEG dynamical analysis). The file vertical lines A through E mark the beginning of these five available EEG segments for the dynamical analysis (the existing gaps in time between the files are not shown – the results from the dynamical analysis of the files are concatenated in time). The dynamical analysis shows that the patient was safe (high T-index values) and stable (high T-index values for a long time) approximately 1 day following admission to the emergency room (beginning of the record). The *vertical green lines* 1 through 5 denote the times of AED administration. Line 1 marks the completion of fosphenytoin infusion. Line 2 marks the administration of fentanyl and etomidate. Line 3 marks administration of 50 mg of propofol and the beginning of a 30 mcg/kg/min propofol infusion. Line 4 marks a fosphenytoin 350 mg infusion and propofol administration at 50 mcg/kg/min. Finally, line 5 marks where propofol is running at 35 mcg/kg/min. In the last EEG recording available to us (file E), the patient remained off propofol.

of SE, the average T-index values remained high, denoting that the involved brain sites remained disentrained. The correspondence of the T-index values and trends to the changing medical condition of these patients over time is remarkable; T-index values were low when AEDs failed, and high when they succeeded in getting the patients out of SE.

17.4 Conclusion

We have shown a very good correlation of the measures derived by mathematical analysis of EEG with the treatment efficacy of AEDs in stopping status epilepticus. The above results indicate that the proposed measure/methodology, as well as

possibly other measures within the framework of nonlinear dynamics and chaos theory, may assist in an objective evaluation of the efficacy of current and future AEDs for the treatment of SE. While larger scale studies are contemplated for further validation of these results, it appears that this methodology could be clinically valuable as an independent online and real-time monitoring of the state of the brain and evaluation of the efficacy of the administered AEDs in SE.

References

1. Abarbanel, H. Analysis of Observed Chaotic Data. Springer Verlag, New York (1996)
2. Alldredge, B., Treiman, D., Bleck, T., Shorvon, S. Treatment of Status Epilepticus. Epilepsy: A Comprehensive Textbook, p. 1357. Philadelphia: Wolters Kluwer, Lippincott Williams & Wilkins (2007)
3. Grassberger, P., Procaccia, I. Characterization of strange attractors. Phys Rev Lett **50**(5), 346–349 (1983)
4. Grassberger, P., Procaccia, I. Measuring the strangeness of strange attractors. Physica D **9**(1–2), 189–208 (1983)
5. Haken, H. Principles of Brain Functioning: A Synergetic Approach to Brain Activity. Springer, New York (1996)
6. Iasemidis, L. Epileptic seizure prediction and control. IEEE Trans Biomed Eng **50**(5), 549–558 (2003)
7. Iasemidis, L., Principe, J., Sackellares, J. Measurement and quantification of spatio-temporal dynamics of human epileptic seizures. Nonlinear Biomedical Signal Processing **2**, 294–318 (2000)
8. Iasemidis, L., Sackellares, J. The temporal evolution of the largest Lyapunov exponent on the human epileptic cortex. In: Duck, D.W., Pritchard, W.S. (eds.) Measuring Chaos in the Human Brain, pp. 49–82. World Scientific, Singapore (1991)
9. Iasemidis, L., Sackellares, J., Zaveri, H., Williams, W. Phase space topography of the electrocorticogram and the Lyapunov exponent in partial seizures. Brain Topogr **2**, 187–201 (1990)
10. Iasemidis, L., Shiau, D., Chaovalitwongse, W., Sackellares, J., Pardalos, P., Principe, J., Carney, P., Prasad, A., Veeramani, B., Tsakalis, K. Adaptive epileptic seizure prediction system. IEEE Trans Biomed Eng **50**(5), 616–627 (2003)
11. Iasemidis, L., Shiau, D., Pardalos, P., Chaovalitwongse, W., Narayanan, K., Prasad, A., Tsakalis, K., Carney, P., Sackellares, J. Long-term prospective online real-time seizure prediction. Clin Neurophysiol **116**(3), 532–544 (2005)
12. Iasemidis, L., Shiau, D., Sackellares, J., Pardalos, P., Prasad, A. Dynamical resetting of the human brain at epileptic seizures: Application of nonlinear dynamics and global optimization techniques. IEEE Trans Biomed Eng **51**(3), 493–506 (2004)
13. Kostelich, E.J. Problems in estimating dynamics from data. Physica D **58**(1–4), 138–152 (1992)
14. Packard, N., Crutchfield, J., Farmer, J., Shaw, R. Geometry from a time series. Phys Rev Lett **45**(9), 712–716 (1980)
15. Panter, P. Modulation, Noise, and Spectral Analysis: Applied to Information Transmission. McGraw-Hill, New York (1965)
16. Sabesan, S., Chakravarthy, N., Tsakalis, K., Pardalos, P., Iasemidis, L. Measuring resetting of brain dynamics at epileptic seizures: Application of global optimization and spatial synchronization techniques. J Comb Optim **17**(1), 74–97 (2009)
17. Takens, F. Detecting strange attractors in turbulence. In: Rand, D.A., Young, L.S. (eds.) Proceedings of Symposium on Dynamical Systems and Turbulence. Coventry (1980)

18. Treiman, D., Meyers, P.,Walton, N., Collins, J., Colling, C., Rowan, A., Handforth, A., Faught, E., Calabrese, V., Uthman, B., et al. A comparison of four treatments for generalized convulsive status epilepticus. Veterans Affairs Status Epilepticus Cooperative Study Group. N Engl J Med **339**(12), 792–798 (1998)
19. Treiman, D., Walker, M. Treatment of seizure emergencies: Convulsive and non-convulsive status epilepticus. Epilepsy Research **68**, 77–82 (2006)
20. Vastano, J., Kostelich, E. Comparison of algorithms for determining Lyapunov exponents from experimental data. In: Mayer-Kress, G. (ed.) Dimensions and Entropies in Chaotic Systems: Quantification of Complex Behavior, Vol. 11. Springer, New York (1986)
21. Wolf, A., Swift, J., Swinney, H., Vastano, J. Determining Lyapunov exponents from a time series. Physica D **16**(3), 285–317 (1985)

Chapter 18
Analysis of Multichannel EEG Recordings Based on Generalized Phase Synchronization and Cointegrated VAR

Alla R. Kammerdiner and Panos M. Pardalos

Abstract Synchronization is shown to be a characteristic feature of electroencephalogram data collected from patients affected by neurological diseases, such as epilepsy. Phase synchronization has been applied successfully to investigate synchrony in neurophysiological signal. The classical approach to phase synchronization is inherently bivariate. We propose a novel *multivariate* approach to phase synchronization, by extending the bivariate case via cointegrated vector autoregression, and then apply the new concept to absence epilepsy data.

keywords Electroencephalogram, Phase synchronization, Cointegrated vector autoregressive processes

18.1 Introduction

The temporal integration of various functional areas in different parts of the brain is believed to be essential for normal cognitive processes. Many studies stress a significant role of neural synchrony in such large-scale integration [6, 42, 41, 43]. Specifically, it was discovered that oscillation of various neuronal groups in given frequency bands leads to temporary phase-locking between such groups of neurons. This discovery prompted the development of robust approaches that allow one to measure the phase synchrony in a given frequency band from experimentally recorded biomedical signals such as EEG.

Alla R. Kammerdiner
Department of Industrial and Systems, Engineering, University of Florida, Gainesville, FL, USA, e-mail: `alla.ua@gmail.com`

Panos M. Pardalos
Department of Industrial and System Engineering, Center for Applied Optimization, University of Florida, Gainesville, FL, USA, e-mail: `pardalos@ufl.edu`

In particular, the importance of synchronization of neuronal discharges has been shown by a variety of animal studies using microelectrode recordings of brain activity [38, 33], and even at coarser levels of resolution by other studies in animals and humans [9]. The phase synchronization in the brain extracted from EEG data using Hilbert or wavelet transforms has recently been shown to be an especially promising tool in analysis of EEG data recorded from patients with various types of epilepsy [35].

In this chapter, we introduce a novel concept of generalized phase synchronization, which is based on vector autoregressive modeling. This new notion of phase synchronization is constructed as an extension of the classical definition of phase synchronization between two systems. Indeed, the phase synchronization is usually defined by imposing the condition that some integer combination of the instantaneous phases of two signals is constant. Often this condition is further relaxed by allowing for a bounded linear combination of two phases, in order to account for noise in the measurements. This classical approach to phase synchronization is clearly bivariate. Since we are interested in investigating synchrony among several areas in the brain, we would like to generalize the bivariate phase synchronization to a multivariate case.

To construct a more general multivariate notion of phase synchronization, we extend the classical definition by considering such a linear combination of phases for a finite number of signals that represent a stationary process. All the individual signals together form a common system described by some multivariate process. We note that a vector process, such that a linear combination of its individual components is a stationary process, can be modeled as a cointegrated vector autoregressive time series.

Furthermore, we show that the cointegrated rank of the regression determines how restricted the behavior of such system is. This means that the rank r of cointegrated autoregressive model, estimated from the multiple time series of the instantaneous phases, measures how large the vector subspace, which generates the changes in the phase values, is.

This new measure of cointegration is also applied to absence epilepsy EEG data. The data sets collected from the patients with other types of epilepsy are currently being investigated.

This chapter is organized as follows. Section 18.2 introduces cointegrated vector autoregressive processes, and various related testing procedures. In Section 18.3, we discuss role of synchronization in brain dynamics, and give a definition of classical phase synchronization. Section 18.4 presents the Hilbert transform method for extracting instantaneous phases from time series. To develop our multivariate approach to studying phase synchrony in a complex system, such as brain, we extend the classical bivariate concept of phase synchronization based on cointegrated vector autoregression in Section 18.5, and test our method on absence epilepsy data.

18.2 Integrated and Cointegrated VAR

Let p be a positive integer, and let y_t denote the K-variate time series (i.e., realizations of K-dimensional process $Y(t)$). A *vector autoregressive* model of order p, denoted VAR(p), is formally defined as follows:

$$y_t = v + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{18.1}$$

where $y_t = (y_{1t}, \ldots, y_{Kt})'$ is a $(K \times 1)$ random vector, $v = (v_1, \ldots, v_K)'$ is a fixed $(K \times 1)$ vector representing a nonzero mean $EY(t)$, the A_i, $i = 1, \ldots, p$ are fixed $(K \times K)$-dimensional coefficient matrices, and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})'$ is a K-dimensional *white noise* process (i.e., $E[\varepsilon_t] = 0$, $E[\varepsilon_s \varepsilon_t'] = 0$, for $s \neq t$, and $E[\varepsilon_s \varepsilon_t'] = \Sigma_\varepsilon$).

It is assumed that the covariance matrix Σ_ε is nonsingular. In addition, the following three important conditions are imposed on the time series in the VAR model:

- $Y(t)$ is a stable process;
- $Y(t)$ is stationary;
- the underlying white noise process ε_t is *Gaussian*.

However, in practice, many time series data are fit better by unstable non-stationary processes. For instance, integrated and cointegrated processes are found especially useful in econometric studies, and for such processes the stability and stationarity conditions are violated.

Note that the VAR(p) process (18.1) satisfies the *stability condition* when its *reverse characteristic polynomial* $\det(I_K - A_1 z - \ldots A_p z^p)$ has no roots on and inside a complex unit circle. If an unstable process has a single unit root and all the other roots outside of the complex unit circle, then such process exhibits a behavior similar to that of a random walk. In other words, the variance of such process increases linearly to infinity, and the correlation between the variables $Y(t)$ and $Y(t \pm h)$ tends to 1 as $t \to \infty$. On the other hand, when the root of reverse characteristic polynomial lies inside the unit circle, the process becomes explosive, i.e., its variance increases exponentially. In real-life applications, the former case is of the most practical interest.

This renders the following definition of an integrated process.

A 1D process with d roots on the unit circle is said to be *integrated of order d* (denoted as $I(d)$).

It can be shown [17] that the integrated $I(d)$ process $Y(t)$ of order d with all roots of its reverse characteristic polynomial being equal to 1 can be made stable by differencing the original process d times. For example, the integrated $I(1)$ process $Y(t)$ becomes stable after taking the first differences $(1 - L)Y(t) = Y(t) - Y(t-1)$, where L represents the lag operator. More generally, for the $I(d)$ process $Y(t)$, its transformation $(1 - L)^d Y(t)$ is stable. An example of an integrated $I(d)$ process in the univariate case is an *autoregressive integrated moving average* process ARIMA(p, d, q).

It is noteworthy to point out that taking differences may distort the relationship among the variables (i.e., 1D components) in some VAR(p) models. In particular,

this is the case for systems with cointegrated variables. It turns out that fitting VAR(p) model after differencing the original *cointegrated* process produces inadequate results.

Suppose that sampled values y_{it} of K different variables of interest $Y_i(t)$ are combined into the K-dimensional vectors $y_t = (y_{1t}, \ldots, y_{Kt})'$. In addition, suppose that the variables are in a *long-run* equilibrium relation:

$$c\, Y(t) := c_1 \cdot Y_1(t) + \ldots + c_K \cdot Y_K(t) = 0, \qquad (18.2)$$

where $c = (c_1, \ldots, c_K)'$ is a K-dimensional real vector. During any given time interval, the relation (18.2) may not necessarily be satisfied precisely by the sample y_t, instead we may have

$$c\, y_t := c_1 \cdot y_{1t} + \ldots + c_K \cdot y_{Kt} = \varepsilon_t, \qquad (18.3)$$

where ε_t is a stochastic process that denotes the deviation from the equilibrium relation at time t. If our assumption about the long-run equilibrium among individual variables $Y_i(t)$, $i = 1, \ldots, K$ is valid then it is reasonable to expect that the variables $Y_i(t)$ move together, i.e., the stochastic process ε_t is stable. On the other hand, this does not contradict the possibility that the variables deviate substantially as a group. Therefore, it is possible that although each individual component $Y_i(t)$ is integrated, there is a linear combination of $Y_i(t)$, $i = 1, \ldots, K$, which is stationary. Integrated processes with such property are called *cointegrated*.

Without loss of generality, we assume that all individual 1D components $Y_i(t)$ ($i = 1, \ldots, K$) are either $I(1)$ or $I(0)$ processes. Then the combined K-dimensional VAR(p) process

$$Y(t) = v + A_1 Y(t-1) + \ldots + A_p Y(t-p) + \varepsilon_t \qquad (18.4)$$

is said to be *cointegrated of rank r*, when the correspondent matrix

$$\Pi = I_K - A_1 - \ldots - A_p \qquad (18.5)$$

has rank r.

Since some 1D components of the cointegrated VAR(p) process are integrated processes, one may be interested in testing the presence of a unit root in the univariate series. In the following section, we present a commonly used unit root test, which was derived by Dickey and Fuller [7].

18.2.1 Augmented Dickey–Fuller Test for Testing the Null Hypothesis of a Unit Root

The augmented Dickey–Fuller (or ADF) test is a widely used statistical test for detecting the existence of a unit root of the reverse characteristic polynomial in a

univariate time series. The limiting distribution of the ADF test for $p \leq k-1$ was derived by Dickey and Fuller [7], and it can be shown that this distribution is the same for $k > 1$ and for $k = 1$. Fuller tabulated the approximate critical values for the ADF test with $k \geq 1$ and $p \leq k-1$ for *specific* sample sizes.

Finite-sample critical values for the ADF test for *any* sample size were obtained by means of response surface analysis by MacKinnon [18], who also showed that an approximate asymptotic distribution function for the test can be derived via response surface estimation of quantiles [19].

Although the asymptotic distribution of the ADF test statistic does not depend on the lag order, it is noted by Cheung et al. [5] that empirical applications must deal with finite samples, in which case the distribution of the ADF test statistic can be sensitive to the lag order. Taking this into account, they closely examined the roles of the sample size and the lag order in finding the finite-sample critical values of the ADF test.

As we noted above, the limiting distribution of the ADF test statistic is the same for $k > 1$ and $k = 1$. Hence, for simplicity, we consider the case of $k = 1$. In fact, let Y denote the autoregressive AR(1) model:

$$Y(t) = c\,Y(t-1) + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{18.6}$$

where $Y(0) = 0$, c is a real number, and $\varepsilon_t \sim N(0, \sigma^2)$ (i.e., ε_t is normally distributed with zero mean and variance σ^2 for all $t = 1, 2, \ldots$).

From the AR(1) model (18.6), one can see that the condition $c = 1$ in (18.6) is equivalent to the requirement that the reverse characteristic polynomial $\det(1 - cz) = 1 - z$ of AR(1) has a unit root. In other words, to determine whether an autoregressive time series AR(1) has a unit root, we must test the null hypothesis $H_0 : c = 1$.

Let $y_1, y_2, \ldots, y_T$ denote a sample of T consecutive observations of the AR(1) process $Y(t)$, then the maximum likelihood estimator of c is the least squares estimator:

$$\widehat{c} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}. \tag{18.7}$$

Note that $\widehat{c}$ is a consistent estimator of the regression coefficient c.

Then the ADF statistic is given by

$$T(\widehat{c} - c) = \frac{\frac{1}{T} \sum_{t=1}^{T} y_{t-1} \varepsilon_t}{\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2}. \tag{18.8}$$

Dickey and Fuller [7] derived the following representation of the limiting distribution for statistic $T(\widehat{c} - c)$:

$$T(\widehat{c} - c) \Rightarrow \frac{1}{2} \Gamma^{-1}(W^2 - 1), \quad \text{as} \quad T \to \infty, \tag{18.9}$$

where

$$\Gamma = \sum_{i=1}^{\infty} d_i^2 X_i^2, \tag{18.10}$$

$$W = \sum_{i=1}^{\infty} \sqrt{2}\, d_i X_i, \tag{18.11}$$

$$d_i = \frac{2(-1)^{i+1}}{\pi(2i-1)}, \tag{18.12}$$

random variables X_i, $i = 1, 2, \ldots$, are independent and identically distributed according to the normal distribution with zero mean and variance σ^2, and $\Rightarrow$ denotes convergence in distribution.

In [7], Dickey and Fuller considered the following "Studentized" statistic based on the likelihood ratio test of the hypothesis $H_0: c = 1$:

$$\widehat{\tau} = \frac{\widehat{c} - 1}{S} \left(\sum_{t=2}^{T} y_{t-1}^2 \right)^{\frac{1}{2}}, \tag{18.13}$$

where

$$S^2 = \frac{1}{T-2} \left(\sum_{t=2}^{T} (y_t - \widehat{c} y_{t-1})^2 \right), \tag{18.14}$$

and $\widehat{c}$ is computed from (18.7).

Tables of the critical values for the asymptotic distributions of the ADF test statistic $T(\widehat{c} - 1)$ and the statistic $\widehat{\tau}$ can be found in Fuller [10].

18.2.2 Estimation of Cointegrated VAR(p) Processes

Several methods can be employed to estimate the parameters of a cointegrated VAR(p) model, including modifications of the approaches used for estimation of the standard VAR(p) processes.

In this section we present the maximum likelihood approach to estimating a *Gaussian* cointegrated VAR(p) process. Suppose y_t is a realization of a K-dimensional VAR(p) process with cointegration rank r, such that $0 < r < K$. Without loss of generality, we assume that $Y(t)$ has zero mean, i.e., the intercept $v = 0$ in (18.4).

Given a realization y_t, $t = 1, 2, \ldots$, of $Y(t)$, one seeks to determine the coefficients of the following model:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t+p} + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{18.15}$$

subject to the constraint

$$\text{rank}(\Pi) = \text{rank}(I_K - A_1 - \ldots - A_p) = r. \tag{18.16}$$

Note that ε_t is assumed to be a Gaussian white noise with a nonsingular covariance matrix Σ_ε. Furthermore, the initial conditions $y_{-p+1}, \ldots, y_0$ are supposed to be fixed.

In order to impose the cointegration constraint, the model (18.15) is reparameterized in the following fashion [17]:

$$\Delta y_t = D_1 \Delta y_{t-1} + \ldots + D_{p-1} \Delta y_{t-p+1} + \Pi y_{t-p} + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{18.17}$$

where $\Delta y_t = y_t - y_{t-1}$, and matrix Π can be represented as a product $\Pi = HC$ of matrices of rank r, i.e., H is $(K \times r)$ and C is $(r \times K)$.

Consider

$$\Delta Y := [\Delta y_1, \ldots, \Delta y_T],$$

$$\Delta X_t := \begin{bmatrix} \Delta y_t \\ \vdots \\ \Delta y_{t-p+2} \end{bmatrix}, \tag{18.18}$$

$$\Delta X := [\Delta X_0, \ldots, \Delta X_{T-1}],$$

$$D := [D_1, \ldots, D_{p-1}],$$

$$Y_{-p} := [y_{1-p}, \ldots, y_{T-p}].$$

Then the log-likelihood function for a sample of size T can be written as

$$\ln l = -\frac{KT}{2} \ln[2\pi] - \frac{T}{2} \ln[\det \Sigma_\varepsilon]$$

$$- \frac{1}{2} \text{trace} \left((\Delta Y - D\Delta X + HCY_{-p})' \Sigma_\varepsilon^{-1} (\Delta Y - D\Delta X + HCY_{-p}) \right). \tag{18.19}$$

The proof of the following theorem on the maximum likelihood estimators of a cointegrated VAR process can be found in [17] (Proposition 11.1).

Theorem 18.1. *(reproduced from [17])*
Define

$$M := I - \Delta X'(\Delta X \Delta X')^{-1} \Delta X,$$

$$R_0 := \Delta Y M,$$

$$R_1 := Y_{-p} M,$$

$$S_{ij} := \frac{1}{T} R_i R_j', \quad i = 0, 1.$$

Let G be the lower triangular matrix with positive diagonal such that $GS_{11}G' = I_K$.
Denote $\lambda_1 \geq \ldots \geq \lambda_K$ to be the eigenvalues of $GS_{10}S_{00}^{-1}S_{01}G'$,
and
$v_1, \ldots, v_2$ be the corresponding orthonormal eigenvectors.

Then the log-likelihood function in (18.19) *is maximized for*

$$
\begin{aligned}
C &:= [v_1, \ldots, v_r]' G, \\
H &:= -\Delta Y M Y_{-p'} C' \left(C Y_{-p} M Y_{-p'} C' \right)^{-1} \\
&= -S_{01} C' \left(C S_{11} C' \right)^{-1}, \\
D &:= \left(\Delta Y + H C Y_{-p} \right) \Delta X \left(\Delta X \Delta X' \right)^{-1}, \\
\Sigma &:= \frac{1}{T} \left(\Delta Y - D \Delta X + H C Y_{-p} \right) \left(\Delta Y - D \Delta X + H C Y_{-p} \right)'.
\end{aligned}
$$

The maximum is

$$
\max[\ln l] = -\frac{KT}{2} \ln[2\pi] - \frac{T}{2} \left(\ln[\det S_{00}] + \sum_{i=1}^{r} \ln(1 - \lambda_i) \right) - \frac{KT}{2}. \qquad (18.20)
$$

18.2.3 Testing for the Rank of Cointegration

Based on Theorem 18.1, one can easily derive the likelihood ratio statistic for testing a candidate value r_0 of the cointegration rank r of a VAR(p) process against a larger cointegration rank r_1.

Given a VAR(p) process $y(t)$ defined by (18.4), suppose we wish to test a hypothesis H_0 against an alternative H_1, where

$$
H_0 : r = r_0 \qquad \text{against} \qquad H_1 : r_0 < r \le r_1. \qquad (18.21)
$$

Under assumption that the noise ε_t is a Gaussian process, the maximum of the likelihood function for a cointegrated VAR(p) model with cointegration rank r is computed in Theorem 18.1. From that result, the value of the LR statistic for testing (18.21) can be determined in the following manner:

$$
\begin{aligned}
\lambda_{\mathrm{LR}}(r_0, r_1) &= 2 \left[\ln L_{\max}(r_1) - \ln L_{\max}(r_0) \right] \qquad (18.22) \\
&= T \left[-\sum_{i=1}^{r_1} \ln(1 - \lambda_i) + \sum_{i=1}^{r_0} \ln(1 - \lambda_i) \right] \\
&= -T \sum_{i=r_0+1}^{r_1} \ln(1 - \lambda_i),
\end{aligned}
$$

where $L_{\max}(r_i)$, $i = 0, 1$, denotes the maximum of the Gaussian likelihood function for cointegration rank r_i. The advantage of this test is in the simplicity with which the LR statistic can be computed. On the other hand, the asymptotic distribution of the LR statistic (18.22) is nonstandard. Specifically, the LR statistic is not asymptotically distributed according to χ^2-distribution. Nevertheless, the asymptotic distribution of the cointegration rank test statistic λ_{LR} depends only on two factors:

- the difference $K - r$ between the process dimensionality and the cointegration rank; and
- the alternative hypothesis.

As a result, the selected percentage points of the asymptotic distribution of the test statistic λ_{LR} were tabulated by Johansen and Juselius in [13].

18.3 The Role of Phase Synchronization in Neural Dynamics

The word "synchrony" originates from a combination of two Greek words $\sigma \upsilon \nu$ (syn, meaning common) and $\chi \rho o \nu o \varsigma$ (chronos, meaning time), and it can be translated as "happening at the same time." A concept of synchronization can be defined as a process of active adjustment between the rhythms of different oscillating systems due to some kind of interaction or coupling between them [28]. Synchronization phenomena were discovered in the late seventeenth century by C. Huygens who first observed synchronization between two pendulum clocks hanging from a common support [12]. Since then, the study of synchronization between dynamical systems became an active field of research in many scientific and technical disciplines, including solid state physics [24], plasma physics [34], communication [3], electronics [27, 22], laser dynamics [8, 36, 39], and control [30, 37].

Complex physiological systems, such as heart and brain, also display synchronization. The presence of synchronization processes in physiological systems was discovered by B. van der Pol in the beginning of the twentieth century. In particular, he first applied oscillation theory to the human heart [29]. The role of synchronization in neural dynamics is an important area of research in neuroscience. Much effort is given to investigation of synchronization phenomena on all different levels of organization of brain tissue, starting with pairs of individual neurons to larger scales, such as within a given area of the brain or between distinct parts of the brain. Recent findings indicate that long-range synchronization can be detected not only in microelectrode studies [38, 33], but also in the studies using surface recordings [32].

It has been shown that synchronization is a significant attribute of the signal recorded from the patients affected by several neurological disorders. In particular, researchers have found that epilepsy [20] and Parkinson's disease [40] manifest as a pathological form of the synchronization process.

Several studies in neuroscience emphasize major difference between synchrony as an appropriate estimate of phase relation, and the classical measures of coherence or spectral covariance [2, 1]. Le Van Quyen et al. discuss two important limitations of coherence [31]. The first limitation arises because the standard approaches for measuring coherence [4] based on Fourier analysis are known to be highly dependent on the stationarity of the measured signal, whereas the signals recorded from the brain, such as EEG, appear to be clearly nonstationary. The second limitation stems from the fact that classical coherence is a measure of spectral covariance. Hence, it is not able to separate the effects of amplitude and phase in the relations between two signals. Thus, coherence gives only an indirect and approximate indication of phase synchrony.

Classical concept of the synchronization of two oscillators is described as an active adjustment of their rhythmicity that manifests in phase locking between the synchronized oscillators. Specifically, given two signals $X_1(t)$ and $X_2(t)$, and their corresponding instantaneous phases $\phi_1(t)$ and $\phi_2(t)$, the basic definition of the *phase locking* states that

$$n\phi_1(t) - m\phi_2(t) = C \equiv \text{const}, \tag{18.23}$$

where integers n and m specify the phase locking ratio.

When investigating phase synchrony in neurophysiological signals, one must assume that the constant phase locking ratio is valid within a limited time interval T, which usually means a few hundreds of milliseconds. As noted in [31], as a consequence of volume conduction effects in brain tissues, the activity of a single neuronal population can be recorded by two distant electrodes, which results in spurious phase locking between their signals. Furthermore, in noninvasive EEG, the true synchronies are hidden in a significant background noise. Hence, in the synchronous state, the phase shifts back and forth around some constant value, and so the signals can be viewed as *synchronous* or *not synchronous* only in a statistical sense. Therefore, the condition (18.23) must be adjusted to account for the noise as follows:

$$C - \varepsilon \le n\phi_1(t) - m\phi_2(t) \le C + \varepsilon, \tag{18.24}$$

where n, m, C are constants from (18.23), and ε denotes a small positive constant.

To investigate phase synchrony, first the instantaneous phases need to be extracted from the data, and then statistical approaches are applied to evaluate the degree of phase synchronization. The following two methods for estimating the phases applied to neuronal signals have recently been considered in the literature. Tass and colleagues [40] extracted the instantaneous phases from original signals by means of the *Hilbert transform*, and then applied to magnetoencephalographic (MEG) motor data in patients affected by Parkinson's disease [40]. On the other hand, Lachaux et al. [16] estimated the phases from the original signals by means of convolution with a complex wavelet, and then applied it to EEG and intracranial data recorded during cognitive tasks [32, 15].

The first step in quantifying phase synchronization between two time series X and Y is the determination of their instantaneous phases $\phi_X(t)$ and $\phi_Y(t)$. This is achieved either via the Hilbert transform or via the wavelet transform. Next, we present phase estimation approach based on Hilbert transform.

18.4 Phase Estimation Using Hilbert Transform

The first method used to extract the instantaneous phase from the time series is based on the *analytic signal approach*, which was first introduced by D. Gabor [11] and later extended for model systems and experimental data [35].

The *Hilbert transform* of a given real-valued function $f(t)$ with domain T is defined as a real-valued function $\hat{f}(t)$ on T as follows:

$$\hat{f}(t) = \text{CPV} \int_{-\infty}^{+\infty} f(\tau)g(t-\tau)d\tau = \text{CPV} \int_{-\infty}^{+\infty} g(\tau)h(t-\tau)d\tau, \qquad (18.25)$$

where

$$g(t) := \frac{1}{\pi t}, \qquad t \in T,$$

and symbol CPV signifies that the integral is taken in the sense of *Cauchy principal value*.

Notice that $\hat{f}(t)$ can be viewed as a convolution $g(t) \times f(t)$ of the original function $f(t)$ with the function $g(t)$. This means that the Hilbert transform can be performed by applying an ideal filter, whose amplitude response equals to 1, and phase response is a constant $\pi/2$ lag at all frequencies.

Given an arbitrary continuous real-valued time series $X(t)$, the corresponding *analytic signal* is defined as the following complex-valued function:

$$\xi_X(t) = X(t) + \iota \cdot \hat{X}(t) = a_X(t) \cdot \exp\{\iota \cdot \phi_X(t)\}, \qquad (18.26)$$

where t denotes time, ι is a unit on the complex axis, $\hat{X}(t)$ denotes the Hilbert transform of the time series $X(t)$, $a_X(t)$ is the corresponding instantaneous amplitude, and $\phi_X(t)$ represents the instantaneous phase of the signal via Hilbert convolution.

It follows from (18.26) that the *instantaneous phase* $\phi_X(t)$ of $X(t)$ can be computed as

$$\phi_X(t) = \arctan\left\{\frac{\hat{X}(t)}{X(t)}\right\}. \qquad (18.27)$$

A key advantage of the analytic approach is that the phase can be easily computed for an arbitrary broadband signal. On the other hand, instantaneous amplitude and phase have a clear physical meaning only if $X(t)$ is a narrowband signal. Therefore, filtration is required in order to separate the frequency band of interest from the background brain activity.

Various measures of phase synchrony between two signals are proposed based on the phases extracted via the Hilbert and the wavelet transforms, including standard deviation, mutual information, and Shannon entropy [31, 14]. However, most of the currently used measures of phase synchronization are based on *bivariate* indexes. In the next section, we propose a novel *multivariate* approach to detecting phase synchronization in the phases extracted from multiple time series, such as multichannel EEG.

18.5 Multivariate Approach to Phase Synchrony via Cointegrated VAR

We develop a new method for measuring the synchrony among the instantaneous phases extracted from multivariate time series. Our technique is based on the cointegrated VAR modeling of time series.

Given the signal represented formally as a multiple time series $X(t)$, one can extract the instantaneous phases $\phi_{X_i}(t)$ from each 1D component $X_i(t)$ of the signal as shown in Section 18.4 (either via a convolution with the Morlet wavelet or by applying the Hilbert transform). The phase extraction procedure produces a new multiple time series $\phi_X(t)$ of the correspondent phases.

Next, we derive new measures of phase synchrony of the signal based on the concepts introduced in Section 18.3. Let us observe that the left-hand side of Equation (18.23) represents the linear combination of the respective phases $\phi_{X_1}(t)$ and $\phi_{X_2}(t)$ with integer coefficients. Also recall that condition (18.23), which defines phase locking between two signals $X_1(t)$ and $X_2(t)$, needs to be modified in practice to account for the noise in the signal. Taking into account presence of the stochastic noise in the phase series, let us introduce a *modified* concept of the phase synchrony between two signals *by relaxing the integrality condition on the coefficients in the linear combination* as follows.

Two signals $X_1(t)$ and $X_2(t)$ are considered to be *generally phase synchronized*, if the correspondent instantaneous phases $\phi_{X_1}(t)$ and $\phi_{X_2}(t)$ satisfy the condition below:

$$\exists\, c_1, c_2 \quad : \quad c_1\phi_{X_1}(t) + c_2\phi_{X_2}(t) = z_t, \tag{18.28}$$

where $z_t \sim N(C, \sigma\prime)$ is a stochastic variable that represents the deviation from the constant level C as a result of the noise. Notice that in the contrast to condition (18.23) in the classic definition of phase synchronization, the coefficients c_1 and c_2 in the definition of generalized phase synchrony (18.28) do not need to be integer.

Furthermore, it is straightforward that the new condition (18.28) means that a 2D process $X(t) = (X_1(t), X_2(t))\prime$ is cointegrated. Based on this observation, we can extend our modified concept of phase synchronization between two signals to the multivariate case in the following manner.

The multichannel signal $X(t) = (X_1(t), \ldots, X_K(t))$ is considered to be *phase-synchronized of rank r*, if the process $\phi_X(t)$ composed of the correspondent instantaneous phases $\phi_{X_i}(t)$, $i = 1, \ldots, K$ is cointegrated of rank r.

In the subsequent subsections, we first discuss the role of the cointegration rank in the framework of multivariate phase synchronization, and then apply this approach to multichannel EEG data collected from the patients with absence epilepsy.

18.5.1 Cointegration Rank as a Measure of Synchronization among Different EEG Channels

Note that integrated autoregressive processes $I(d)$ are shown to exhibit behavior similar to that of a random walk. In a short paper [21], Michael Murray used an example of drunkard and her dog to illustrate the concept of the cointegration. To explain our reasoning behind the rank of cointegration as a measure of synchrony, we briefly summarize and then further extend his analogy.

Random walk process is often described to students using an example of the drunkard's walk. The drunkard wonders aimlessly, so that the direction of each step is random and completely independent of her previous steps. In other words, the meandering of the drunkard is described by a random walk:

$$x_t - x_{t-1} = \varepsilon_t, \quad t = 1, 2, \ldots, \tag{18.29}$$

where x_t represents the position of the drunk at time t, and ε_t is a stationary white-noise, which models the drunk's step at time t.

As Murray noticed [21], an unleashed puppy is another creature, whose behavior reminds a random walk. Indeed, each new scent that puppy's nose comes upon dictates a direction for the pup's next step so strongly that the last scent along with its direction is forgotten as soon as the new scent appears. Having shown that the puppies follow the random walk y_t, $t = 1, 2, \ldots$, let us represent the puppy's walk as:

$$y_t - y_{t-1} = \varepsilon_t, \quad t = 1, 2, \ldots, \tag{18.30}$$

where ε_t is a stationary white noise (i.e., puppy's step at time t).

For a random walk, the best predictor of the future value is the most recently observed one. In other words, the longer it has been since we had seen the drunk, or the dog, the further away from the initial place, on average, they are at the moment. As a result, even if the drunk and the dog crossed their walks at some location, as the time goes on, they tend to wander further away from each other.

However, if the puppy belongs to the drunkard, then they will remain relatively close to each other at all the time, similarly to the individual integrated processes that together form a cointegrated process. Indeed, the drunk would still wonder aimlessly in a random walk fashion, as would her puppy. However, from time to time she would remember about her dog and call for it, the puppy would recognize her voice and bark. They would hear each other and make their next step in each other's direction.

The paths of the drunk and her dog are still nonstationary, but they are no longer independent from each other. As a matter of fact, at each time, the puppy and its master are likely to be found not far from each other. If this is true, then the distance between two paths is stationary, and the walks of the drunk x_t and her dog y_t are said to be *cointegrated*, i.e., x_t and y_t are integrated $I(1)$, and there is a *linear combination* of x_t and y_t (with nonzero weights) that is $I(0)$, i.e., stationary.

Mathematically, the *cointegrating relationship* between a lady and her puppy can be written as

$$x_t - x_{t-1} = \varepsilon_t + c(y_{t-1} - x_{t-1}), \tag{18.31}$$
$$y_t - y_{t-1} = \varepsilon_t + d(x_{t-1} - y_{t-1}), \tag{18.32}$$

at time $t = 1, 2, \ldots$. Note that ε_t, as before, represent the stationary white noise steps of the drunk and her dog.

Since Equation (18.31) can be easily rewritten in form of (18.17) as follows:

$$\Delta \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ \varepsilon_t \end{bmatrix} - \begin{bmatrix} c & -c \\ -d & d \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix}, \tag{18.33}$$

then

$$\Pi = \begin{bmatrix} c & -c \\ -d & d \end{bmatrix},$$

and so, $\text{rank}(\Pi) = 1$. This shows that the cointegrating relationship between the drunk lady and her puppy has the cointegration rank 1.

Note that $\text{rank}(\Pi) = 0$, if and only if $c = d = 0$. In such case, (18.31) becomes simply a system of Equations (18.29) and (18.30), which models *two* independent random walks driven by independent white noise process ε. On the other hand, when at least one of the coefficients c and d is nonzero, then by multiplying system (18.33) by a vector $[d, c]'$, we have

$$d \Delta x_t + c \Delta y_t = d \varepsilon_t + c \varepsilon_t, \quad t = 1, 2, \ldots, \tag{18.34}$$

which means that the model is driven by a *single* common stochastic trend $d\varepsilon_t + c\varepsilon_t$.

Although the example described by Murray is clearly a bivariate cointegrated VAR(1), it can be extended to an illustration of the multivariate cointegrated process. Consider, for example, a herd of sheep guarded by two dogs, where the sheep wonder aimlessly in the field, while the dogs run around and bring the sheep that have strayed too far back into the flock. Say, for example, a faster dog guards sheep from the east, south, and west, whereas a slower dog – from the north, then the cointegrated process appears to have the cointegration rank of 2. Clearly, two dogs are able to keep a flock of sheep closer together, than a single dog can. In other words, the higher cointegration rank the more restrictive it is.

In fact, let us consider a K-dimensional cointegrated vector autoregressive process, and let r denote the cointegration rank of the process. Similarly to the bivariate example above, we can see that when the rank is zero ($r = 0$), the univariate components of the process are independent, and the model is driven by K independent white noise processes (i.e., there is no cointegration). In the case of $r = 1$, we can decompose the multivariate process onto $K - 2$ independent components, and two dependent components that form a common stochastic trend. Hence, in the case $r = 1$, the cointegrated model is driven by $(K - 2) + 1 = K - 1$ independent stochastic processes. By induction, we can show that for a cointegrated VAR process with the cointegration rank r, $0 < r < K - 1$, the VAR model is generated by $K - r$ independent stochastic trends.

Therefore, the smaller is the cointegration rank r, the larger is the number $K - r$ of the underlying independent stochastic trends, and so (the larger) is the vector space in which our cointegrated model can travel. And the other way around, increasing the cointegration rank of the model shrinks the underlying domain of the process, i.e., makes it bounded to a smaller hyperplane. For $r = K$, the VAR(p) is a stable process, which clearly has the most constrained domain. For $r = 0$, the VAR process is not cointegrated and unrestricted.

Thus, in the framework of generalized phase synchronization introduced above, the cointegration rank represents a fundamental measure of synchrony in the multichannel signal, such as EEG. In particular, we say that the signal is *completely asynchronous*, if the cointegration rank r is 0. On the other hand, when the multivariate process is stable (i.e., the rank coincides with the dimension of the process, $r = K$), the signal is said to be *perfectly synchronous*.

18.5.2 Absence Seizures

Absence seizures (or petit mal seizures) are known to occur in several forms of epilepsy, whereas absence epilepsy refers to a type of epilepsy in which only the absence seizures occur. Absence epilepsy is usually characterized by age of onset, and often affects teenage population. Absence seizures usually begin in childhood or adolescence, and often run in families, which may suggest a genetic predisposition. Absence seizures are marked by momentary lapses of consciousness. Absence seizures often have no visible symptoms, although some patients may have purposeless movements during a seizure, such as rapidly blinking eyes. Absence seizures often have a brief duration, and a person may resume the previous activity immediately after the seizure [23]. These brief seizures can happen several times during a day, but in some patients, the frequency of absence seizures can be as high as hundred of times a day, which interferes with the daily activities of a child such as school. In some cases of childhood absence epilepsy, the seizures stop when a child reaches puberty. Absence seizures exhibit a characteristic spike-and-wave EEG pattern at a 3 Hz frequency [23].

Figure 18.1 displays a multichannel EEG recording that includes an absence seizure. The duration of the seizure is approximately 4 s. The figure vividly illustrates a characteristic spike-and-wave activity during the seizure.

18.5.3 Numerical Study of Synchrony in Multichannel EEG Recordings from Patients with Absence Epilepsy

The proposed approach to studying synchronization among multiple channels was applied to analysis of EEG data recorded from the patience with absence epilepsy.

First, the multiple time series of the instantaneous phases were extracted from the raw EEG data using the Hilbert transform approach as described in Section 18.4. In particular, we took advantage of the functions `hilbert` and `angle` readily available in the MATLAB R 2006a environment.

The VAR modeling and testing were implemented using the R 2.6.1 statistical software. In our analysis of the instantaneous phases, we incorporated `ar`, `adf.test`, `po.test`, `cajolst` and other functions found in packages *tseries* and *urca*.
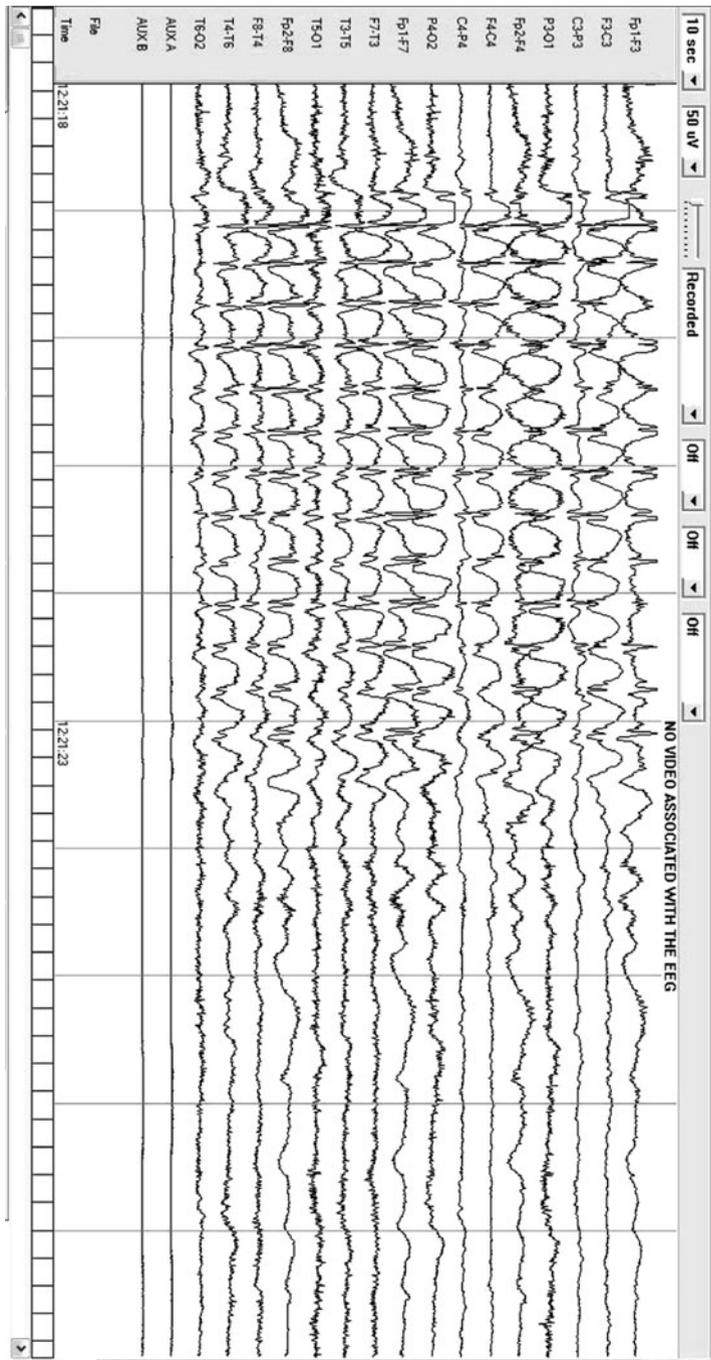
Fig. 18.1: Segment of multichannel EEG long absence seizure.

Next, we illustrate our approach on the example of the EEG data file that includes three seizure intervals. The file contains a 16-channel recording of scalp EEG sampled at the 200 Hz frequency as well as two auxiliary channels, which were discarded. The instantaneous phase values were estimated from the EEG time series by means of Hilbert transform, and the resulting phase series were tested using the ADF test introduced in Section 18.2.1. Specifically, we applied the augmented Dickey–Fuller procedure to test the presence of a unit root in the individual univariate components of the multiple time series of estimated phases.

The results of our experiments for three consecutive seizures are presented in Tables 18.1, 18.2, and 18.3, where each table, respectively, summarizes results for one of the following types of EEG segments:

- during approximately 2 s immediately preceding a seizure;
- during a seizure;
- during approximately 2 s immediately after a seizure.

The channels, for which the ADF unit root test has detected a presence of a unit root at the significance level $\alpha = 0.025$, are listed as *integrated*. Whereas the channels, for which the null hypothesis of a unit root has been rejected by the ADF at the 2.5% level, are denoted by *stationary*. Interestingly, when the ADF is applied at a 0.025 significance level, all three seizure segments are considered stable.

Table 18.1: Pre-ictal: Results of the ADF unit root tests for each channel during the segments 2 s immediately *before a seizure* for three consecutive seizures. (The significance level is set at 2.5%)

Seizure #	Stationary	Integrated
Seizure 1	3,4,5,7,9,10,11,15	1,2,6,8,12,13,14,16
Seizure 2	3,4,6,7,9,10,11,13,15,16	1,2,5,8,12,14
Seizure 3	3,7,9,11,12,13,14,15	1,2,4,5,6,8,10,16

Table 18.2: Ictal: Results of the ADF unit root tests for each channel *during a seizure* for three consecutive seizures. (The significance level is set at 2.5%)

Seizure #	Stationary	Integrated
Seizure 1	1–16	none
Seizure 2	1–16	none
Seizure 3	1–16	none

Next, we fit vector autoregression to the multiple time series of phase estimates, for each of three different segments (before, during, and after a seizure) in order to determine appropriate lag length parameter p. To find appropriate lags p, the

Table 18.3: Post-ictal: Results of the ADF unit root tests for each channel during the segments 2 s immediately *after each seizure* for three consecutive seizures. (The significance level is set at 2.5%)

Seizure #	Stationary	Integrated
Seizure 1	1,3,5,6,7,8,9,10,12,14,15	2,4,13,16
Seizure 2	2,3,5,7,8,10,11,12,13,14,15,16	1,4,6,9
Seizure 3	1,2,4,5,7,8,11,12,13,15,16	3,6,9,10,14

Akaike information criteria (AIC) was used. This led us to choose several lag length for each segment and each seizure. Finally, Johansen cointegration rank procedure was applied to determine the values of cointegration rank r for each case. The results are summarized in Tables 18.4, 18.5, and 18.6.

Table 18.4: Pre-ictal cointegration rank: Results of the Johansen procedure for the multiple series during 2 s immediately before seizure for three consecutive seizures. Significance level is 1%. Full rank is denoted by †

Seizure #	Short lag		Long lag	
	p	r	p	r
Seizure 1	$p = 2, r = 13$		$p = 22, r = 12$	
Seizure 2	$p = 2, r = 16$†		$p = 21, r = 14$	
Seizure 3	$p = 2, r = 9$		$p = 24, r = 13$	

Table 18.5: Ictal cointegration rank: Results of the Johansen procedure for the multiple series during a seizure. Significance level is 1%. Full rank is denoted by †

Seizure #	Short lag		Long lag	
	p	r	p	r
Seizure 1	$p = 2, r = 16$†		$p = 23, r = 11,$ $p = 20, r = 13$	
Seizure 2	$p = 3, r = 16$†		$p = 26, r = 12,$ $p = 20, r = 9$	
Seizure 3	$p = 2, r = 16$†		$p = 26, r = 16$†, $p = 20, r = 16$†	

Notice that during the seizure the system becomes stable, especially when modeled using a short estimate of the lag parameter. Since the durations of the seizure 1 and seizure 2 are rather short, and only include 440–500 sample points, the models

Table 18.6: Post-ictal cointegration rank: Results of the Johansen procedure for the multiple series during 2 s after seizure for three consecutive seizures. Significance level is 1%. Full rank is denoted by †

Seizure #	Short lag		Long lag	
	p	r	p	r
Seizure 1	$p = 2, r = 10$		$p = 20, r = 12$	
Seizure 2	$p = 2, r = 13$		$p = 20, r = 10$	
Seizure 3	$p = 2, r = 16$†		$p = 20, r = 13$	

estimated under a long lag parameter may not adequately represent the underlying processes in seizures 1 and 2. On the other hand, seizure 3 is estimated based on almost 1,200 sample values, and therefore, the long lag model of a longer seizure 3 may be more realistic, than the long lag models for shorter seizures 1 and 2. Overall, the models based on a short lag p for all three seizures provide an evidence of absolute synchronization among the channels. Whereas, the preseizure and postseizure models are more likely to be less restricted, and seem to exhibit a cointegration rank between 9 and 16.

18.6 Conclusion

Recent success in application of phase synchronization to analysis of dynamic processes in epileptic brain motivated us to develop a concept of *generalized synchronization*. This new concept based on our original idea to extend the condition of classical synchronization from the classical bivariate case to a more general multivariate case by studying a cointegrating relationship in the multiple time series. The proposed approach allows one to analyze the synchrony among different parts of the common interrelated system (such as a human brain), by modeling the phases extracted from a finite number of signals in the systems by means of cointegrated vector autoregression. Interestingly, the cointegration rank in the cointegrated VAR model of the phase time series can be viewed as a measure of synchrony among the phases of different components of the EEG signal. We applied our multivariate approach to phase synchronization on the EEG data recorded from the patients with absence epilepsy. The results of our experiments indicate that the new method is capable of capturing phase synchronization in multivariate EEG during seizures.

Not only this new measure of multivariate phase synchrony can be tested on various biomedical data, such as multichannel EEG recorded from an epileptic brain, but also the new multiple phase synchronization can be employed in different areas of applied and theoretic research (including physics, communication, electronics, laser dynamics, and control) for studying synchronization among several dynamical systems or a system that consists of several parts.

18.6.1 Phillips–Ouliaris Cointegration Test

The unit root tests based on analysis of residuals were introduced by Phillips [25]. In particular, in his study Phillips first considered two statistics Z_α and Z_t for testing the null of no cointegration in time series.

Because many unit root tests, constructed before 1987, were founded on the assumption that the errors in the regression are independent with common variance (which is rarely met in practice), Phillips wanted to relax the rather strict condition that the time series are driven by independent identically distributed innovations. In other words, he wanted to develop the testing procedures based on the least squares regression estimation and the associated regression t statistic, which would allow for rather general weakly dependent and heterogeneously distributed sequence of error terms.

The properties of asymptotic distributions of residual-based tests for the presence of cointegration in multiple time series were thoroughly investigated by Phillips and Ouliaris [26]. The characteristic feature of these tests is that they utilize the residuals computed from regressions among the univariate components of multivariate series. The residual-based procedures developed by Phillips and Ouliaris are designed to test the null of *no cointegration* by means of testing the null hypothesis of the unit root presence in the residuals against the alternative of a root that lies inside the complex unit circle. The hypothesis H_0 of the absence of cointegration is rejected, if the null of a unit root in the residuals is rejected. In the nutshell, the procedures are simply residual-based unit root tests.

As noted in [26], the residual-based unit root tests are asymptotically similar, and can be represented via the standard Brownian motion. Moreover, the ADF and Z_t tests are proved to be asymptotically equivalent. However, these two tests are not as powerful as the test based on statistic Z_α, because it was shown by Phillips and Ouliaris [26] that the rate of divergence under cointegration assumption is slower for the ADF and Z_t than other tests, such as the Z_α-statistic test. The later test (i.e., the cointegration test based on Z_α) is also widely known as the *Phillips–Ouliaris cointegration test*.

It is noteworthy that the null hypothesis for the Phillips–Ouliaris test is that of *no cointegration* (instead of cointegration). This formulation is chosen because of some major pitfalls found in procedures that are designed to test the null of cointegration in multiple time series. These defects (discussed in more detail in [26]) are significant enough to be a strong argument against the indiscriminate use of the test formulations based on the null of cointegration, and to support the continuing use of residual based unit root tests.

Consider the K-dimensional vector autoregressive process $Y(t)$. Let us partition $Y(t) = (U_t, V_t')'$ into the univariate component $U_t = Y_1(t)$ and the $(K-1)$-dimensional $V_t = (Y_2(t), \dots, Y_K(t))'$.

The residuals are determined by fitting linear cointegrating regression:

$$U(t) = c\, V(t) + \xi_t, \qquad t = 1, 2, \dots \qquad . \tag{18.35}$$

Residual-based tests are formulated to test the null hypothesis that the multiple time series $Y(t)$ are not cointegrated using the *scalar* unit root tests, such as the ADF test, which are applied to the residuals ξ_t, $t = 1, 2, \ldots$ in (18.35)

In [26], the ADF test as well as two additional tests Z_α and Z_t, developed earlier by Phillips [25], were applied to check for the presence of a unit root in the residuals ξ_t. In order to perform the unit root test, we fit an AR(1) model to ξ_t, $t = 1, 2, \ldots$ according to

$$\xi_t = \hat{\alpha}\,\xi_{t-1} + \rho_t, \qquad t = 1, 2, \ldots \qquad . \tag{18.36}$$

Then the statistic Z_α in Phillips–Ouliaris test is defined as follows:

$$Z_\alpha = T(\hat{\alpha} - 1) - \frac{1}{2} \cdot \frac{s_{Tl}^2 - s_\rho^2}{\frac{1}{T^2} \sum_{t=2}^{T} \xi_{t-1}^2}, \tag{18.37}$$

whereas the Z_t statistic is given by the following formula:

$$Z_t = \left(\sum_{t=2}^{T} \xi_{t-1}^2 \right)^{\frac{1}{2}} \cdot \frac{(\hat{\alpha} - 1)}{s_{Tl}} - \frac{1}{2} \cdot \frac{s_{Tl}^2 - s_\rho^2}{s_{Tl} \left(\frac{1}{T^2} \sum_{t=2}^{T} \xi_{t-1}^2 \right)^{\frac{1}{2}}}, \tag{18.38}$$

where

$$s_\rho^2 = \frac{1}{T} \sum_{t=1}^{T} \rho_t^2, \tag{18.39}$$

$$s_{Tl}^2 = \frac{1}{T} \sum_{t=1}^{T} \rho_t^2 + \frac{2}{T} \sum_{s=1}^{T} w_{sl} \sum_{t=s+1}^{T} \rho_t \rho_{t-s}, \tag{18.40}$$

$$w_{sl} = 1 - \frac{s}{l+1}. \tag{18.41}$$

Note that s_ρ^2 and s_{Tl} are consistent estimators for the variance σ_ρ^2 of ρ_t and the partial sum variance $\sigma^2 = \lim_{T \to \infty} E\left(\frac{1}{T} S_T^2 \right)$, where $S_T = \sum_{t=1}^{T} \xi_t$ is the partial sum of the error terms in (18.36).

The critical values for Z_α and Z_t statistics can be found in [26] (Tables I and II). Phillips and Ouliaris tabulated the values for cointegrating regressions with at most 5 explanatory variables. Some estimates of the critical values for the Phillips–Ouliaris test (Z_α) are listed in Table 18.7.

Table 18.7: Critical values of the asymptotic distributions of the Z_α statistic for testing the null of no cointegration (Phillips–Ouliaris demeaned, reproduced from [26]). Parameter n ($n = K - 1$) represents the number of explanatory variables

n	90%	95%	99%
1	−17.0390	−20.4935	−28.3218
2	−22.1948	−26.0943	−34.1686
3	−27.5846	−32.0615	−41.1348
4	−32.7382	−37.1508	−47.5118
5	−37.0074	−41.9388	−52.1723

Table 18.8: Percentage points of the asymptotic distributions of the $\lambda_{\mathrm{LR}}(r,K)$ for testing the cointegration rank (reproduced from [13])

$K-r$	90%	95%	99%
1	6.69	8.08	11.58
2	15.58	17.84	21.96
3	28.44	31.26	37.29

Table 18.9: Percentage points of the asymptotic distributions of the $\lambda_{\mathrm{LR}}(r,r+1)$ for testing the cointegration rank (reproduced from [13])

$K-r$	90%	95%	99%
1	6.69	8.08	11.58
2	12.78	14.60	18.78
3	18.96	21.28	26.15

References

1. Bressler, S.L., Coppola, R., Nakamura, R. Episodic multiregional cortical coherence at multiple frequencies during visual task performance. Nature **366**, 153–156 (1993)
2. Bullock, T.H., McClune, M.C. Lateral coherence of the electrocorticogram: A new measure of brain synchrony. Electroencephalogr Clin Neurophysiol **73**, 479–498 (1989)
3. Carroll, T.L., Pecora, L.M. Cascading synchronized chaotic systems. Physica D **67**, 126 (1993)
4. Carter, G.C. Coherence and time delay estimation. Proc IEEE **75**, 236–255 (1987)
5. Cheung, Y.W., Lai, K.S. Lag order and critical values of the augmented dickey-fuller test. J Bus Econ Stat **13**(3), 277–280 (1995)
6. Damasio, A.R. Synchronous activation in multiple cortical regions: A mechanism for recall. Semin Neurosci **2**, 287–296 (1990)
7. Dickey, D.A., Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc **74**, 427–431 (1979)
8. Fabiny, L., Colet, P., Roy, R. Coherence and phase dynamics of spatially coupled solid-state lasers. Phys Rev A **47**, 4287 (1993)
9. Freeman, W.J. Spatial properties of an EEG event in the olfactory bulb and cortex. Electroencephalogr Clin Neurophysiol **44**, 586–605 (1978)
10. Fuller, W.A. Introduction to Statistical Time Series, 2nd edn. John Wiley, New York (1996)
11. Gabor, D. Theory of communication. Proc IEEE Lond **93**, 429 (1946)
12. Huygens, C. Horoloquium Oscilatorium. Parisiis, Paris (1673)
13. Johansen, S., Juselius, K. Maximum likelihood estimation and inference on cointegration – with applications to the demand for money. Oxford Bulletin of Economics and Statistics **52**, 169–210 (1990)
14. Kreuz, T. Measuring synchronization in model systems and electroencephalographic time series from epilepsy patients. Ph.D. thesis. Dissertation NIC Series, Vol. 21 (2001)
15. Lachaux, J.P., Rodriguez, E., Martinerie, J., Adam, C., Hasboun, D., Varela, F.J. Gamma-band activity in human intracortical recordings triggered by cognitive tasks. Eur J Neurosci **12**, 2608–2622 (2000)
16. Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J. Measuring phase synchrony in brain signals. Hum Brain Mapp **8**, 194 (1999)
17. Lütkepohl, H. Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin, Heidelberg (1991)

18. MacKinnon, J.G. Critical values for cointegration tests. In: Long-Run Economic Relationships: Readings in Cointegration, pp. 266–276. Oxford University Press, New York (1991)

19. MacKinnon, J.G. Approximate asymptotic distribution functions for unit-root and cointegration tests. J Bus Econ Stat **12**, 167–176 (1994)

20. Mormann, F., Lehnertz, K., David, P., Elger, C.E. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epileptic patients. Physica D **144**, 358–369 (2000)

21. Murray, M.P. A drunk and her dog: An illustration of cointegration and error correction. Am Stat **48**(1), 37–39 (1994)

22. Parlitz, U., Junge, L., Lauterborn, W., Kocarev, L. Experimental observation of phase synchronization. Phys Rev E **54**, 2115 (1996)

23. Patten, J.B. Neurological Differential Diagnosis, 2nd edn. Springer-Verlag, New York (1996)

24. Peterman, D.W., Ye, M., Wigen, P.E. High frequency synchronization of chaos. Phys Rev Lett **74**, 1740 (1995)

25. Phillips, P.C.B. Time series regression with a unit root. Econometrica **55**, 277–301 (1987)

26. Phillips, P.C.B., Ouliaris, S. Asymptotic properties of residual based tests for cointegration. Econometrica **58**, 165–193 (1990)

27. Pikovsky, A.S. Phase synchronization of chaotic oscillations by a periodic external field. Sov J Commun Technol Electron **30**, 85 (1985)

28. Pikovsky, A.S., Rosenblum, M.G., Kurths, J. Synchronization. A universal concept in nonlinear sciences. Cambridge University Press, Cambridge (2001)

29. van der Pol, B., van der Mark, J. The heartbeat considered as a relaxation oscillation, and an electrical model of the heart. Phil Mag **6**, 763 (1928)

30. Pyragas, K. Continuous control of chaos by self-controlling feedback. Phys Lett A **170**, 421 (1992)

31. Quyen, M.L.V., Foucher, J., Lachaux, J.P., Rodriguez, E., Lutz, A., Martinerie, J., Varela, F.J. Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. J Neurosci Methods **111**, 83–98 (2001)

32. Rodriguez, E., George, N., Lachaux, J.P., Martinerie, J., Varela, F.J. Perceptions shadow: Long-distance synchronization in the human brain. Nature **397**, 340–343 (1999)

33. Roelfsema, P.R., Engel, A.K., König, P., Singer, W. Visuomotor integration is associated with zero time-lag synchronization among cortical areas. Nature **385**, 157–161 (1997)

34. Rosa, E., Jr. Pardo, W.B., Ticos, C.M., Walkenstein, J.A., Monti, M. Phase synchronization of chaos in a plasma discharge tube. Int J Bifurc Chaos **10**, 2551 (2000)

35. Rosenblum, M.G., Pikovsky, A.S., Schäfer, C., Tass, P., Kurths, J. Phase synchronization: From theory to data analysis. In: Handbook of Biological Physics, *Neuro-informatics*, Vol. 4. Elsevier Science, Amsterdam (1999)

36. Roy, R., Thornburg, K.S. Experimental synchronization on chaotic lasers. Phys Rev Lett **72**, 2009 (1994)

37. Rulkov, N.F., Tsimring, L.S., Abarbanel, H.D.I. Tracking unstable orbits in chaos using dissipative feedback control. Phys Rev E **50**, 314 (1994)

38. Singer, W., Gray, C.M. Visual feature integration and the temporal correlation hypothesis. Annu Rev Neurosci **18**, 555–586 (1995)

39. Tang, D.Y., Dykstra, R., Hamilton, M.W., Heckenberg, N.R. Experimental evidence of frequency entrainment between coupled chaotic oscillations. Phys Rev E **57**(3), 3649 (1998)

40. Tass, P., Rosenblum, M.G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., et al. Detection of n:m phase locking from noisy data: Application to magnetoencephalography. Phys Rev Lett **81**, 3291–3294 (1998)

41. Tononi, G., Edelman, G. Consciousness and complexity. Science **282**, 1846–1851 (1998)

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

43. Varela, F.J., Lachaux, J.P., Rodriguez, E., Martinerie, J. The brain web: Phase synchronization and large-scale integration. Nat Rev Neurosci **2**, 229–239 (2001)

Chapter 19
Antiepileptic Therapy Reduces Coupling Strength Among Brain Cortical Regions in Patients with Unverricht–Lundborg Disease: A Pilot Study

Chang-Chia Liu, Petros Xanthopoulos, Vera Tomaino, Kazutaka Kobayashi, Basim M. Uthman, and Panos M. Pardalos

Abstract The unified myoclonus rating scale (UMRS) has been utilized to assess the severity of myoclonus and the efficacy of antiepileptic drug (AED) treatment in patients with Unverricht–Lundborg disease (ULD). Electroencephalographic (EEG) recordings are normally used as a supplemental tool for the diagnosis of epilepsy disorders. In this study, mutual information and nonlinear interdependence measures were applied to the EEG recordings in an attempt to identify the effect of treatment

Chang-Chia Liu
J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611, USA, e-mail: `iamjeff@ufl.edu`

Petros Xanthopoulos
Department of Industrial and Systems Engineering, University of Florida Gainesville, FL 32601, USA, e-mail: `petrosx@ufl.edu`

Vera Tomaino
Bioinformatics Laboratory, Experimental and Clinical Medicine Department, Magna Graecia University, viale Europa 88100, Catanzaro, Italy; Department of Industrial and Systems Engineering, University of Florida, Center for Applied Optimization, Gainesville, FL, USA, e-mail: `vera.tomaino@gmail.com`

Kazutaka Kobayashi
Department of Neurological Surgery, Nihon University School of Medicine, Tokyo, Japan; Division of Applied System Neuroscience, Department of Advanced Medical Science, Nihon University School of Medicine, Tokyo, Japan

Basim M. Uthman
Department of Neurology, University of Florida, Gainesville, FL 32611, USA; Department of Neuroscience, University of Florida, Gainesville, FL 32611, USA; The Evelyn F. and William L. McKnight Brain Institute, University of Florida, Gainesville, FL 32611, USA; Neurology Services, North Florida/South Georgia Veterans Health System, Gainesville, FL 32605, USA, e-mail: `basim.uthman@med.va.gov`

Panos M. Pardalos
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA; J. Crayton Pruitt Family Department of Biomedical Engineering University of Florida, Gainesville, FL 32611, USA; The Evelyn F. and William L. McKnight Brain Institute, University of Florida, Gainesville, FL 32611, USA, e-mail: `pardalos@ufl.edu`

on the coupling strength and directionality of mutual information and nonlinear interdependences between different brain cortical regions. Two 1-h EEG recordings were acquired from four ULD subjects; one prior and one after a minimum of 2 months treatment with an add-on AED. Subjects in this study were siblings of same parents and suffered from ULD for approximately 37 years. Our results indicated that the coupling strength was low between different brain cortical regions in the patients with disease of less severity. Adjunctive AED treatment was associated with significant decrease of the coupling strength in all subjects. The mutual information between different brain cortical regions was also reduced after treatment. These findings could provide a new insight for developing a novel surrogate outcome measure for patients with epilepsy when clinical tools or observations could potentially fail to detect a significant difference.

19.1 Introduction

The EEG is an essential tool used to corroborate the diagnosis of epilepsy and other neurological disorders. Changes in the frequency and amplitude of EEG activity arise from spontaneous interactions between excitatory and inhibitory neurons in the brain. The underlying mechanism of brain function, studied by researchers, suggested the importance of the EEG coupling strength between different brain cortical regions. For example, it has been shown that the synchronization of EEG activity is important for the memory [12, 13] and the learning processes [30] of brain. In one study, different brain synchronization/desynchronization of EEG patterns were reportedly induced by hippocampal atrophy in subjects with mild cognitive impairment [18].

Several authors have suggested a direct relationship between changes in synchronization of EEG and the onset of epileptic seizures. Using intracranial EEG recordings, Iasemidis et al. reported that the nonlinear dynamical entrainment of cortical regions is a necessary condition for onset of seizures in patients with temporal lobe epilepsy [10, 11, 22]. Le Van Quyen et al. showed that epileptic seizures might be predicted by nonlinear analysis of dynamical similarity between EEG channels [28]. Mormann et al. claimed that a preictal (before a seizure) state could be characterized by a decrease in synchronization between some EEG channels [20, 19]. A normal brain state is associated with a higher degree of complexity in EEG; transition into a lower degree of complexity may suggest pathology in the brain. In a recent study, using linear and nonlinear synchronization measures, Aarabi et al. indicated that during the interictal state, the degree of interdependence between EEG channels was significantly less than that observed in the ictal state in typical absence seizure EEG recordings. In some cases, the authors reported that they could identify preictal states by a significant decrease in the synchronization level with respect to

interictal states [1]. Synchronization patterns were also found to depend on epileptic syndromes with primary generalized absence seizures displaying more long-range synchrony in frequency bands (3–5 Hz) than generalized tonic motor seizures of secondary (symptomatic) generalized epilepsy or frontal lobe epilepsy [6]. In addition, we hypothesize that coupling strength and amount of mutual information between different brain cortical regions exit in the patients with higher severity of ULD. In addition, the coupling strength and amount of mutual information in brain cortical regions are positively correlated with the degree of severity (higher UMRS scores).

Unverricht–Lundborg disease is one type of progressive myoclonic epilepsy (PME); a rare epilepsy disorder with complex inheritance. It was first described by Unverricht in 1891 and Lundborg in 1903 [36, 17].

AEDs are the mainstay treatments of ULD with overall unsatisfactory efficacy. Due to the progression of the severity of myoclonus, the efficacy of AED treatment is difficult to clinically measure especially in the later stages of the disease. EEG recordings of ULD subjects usually demonstrate abnormal slow background rhythms and frequent generalized high-amplitude 3–5 Hz spike waves or poly spike and wave complexes. Sometimes, normal background EEG can be observed between generalized spike and wave discharges (see Figs. 19.1 and 19.2 for EEG examples). Studies have shown increased background slowing of EEG or no change in patients with more advanced ULD stages [4, 9]. Generalized slowing of EEG
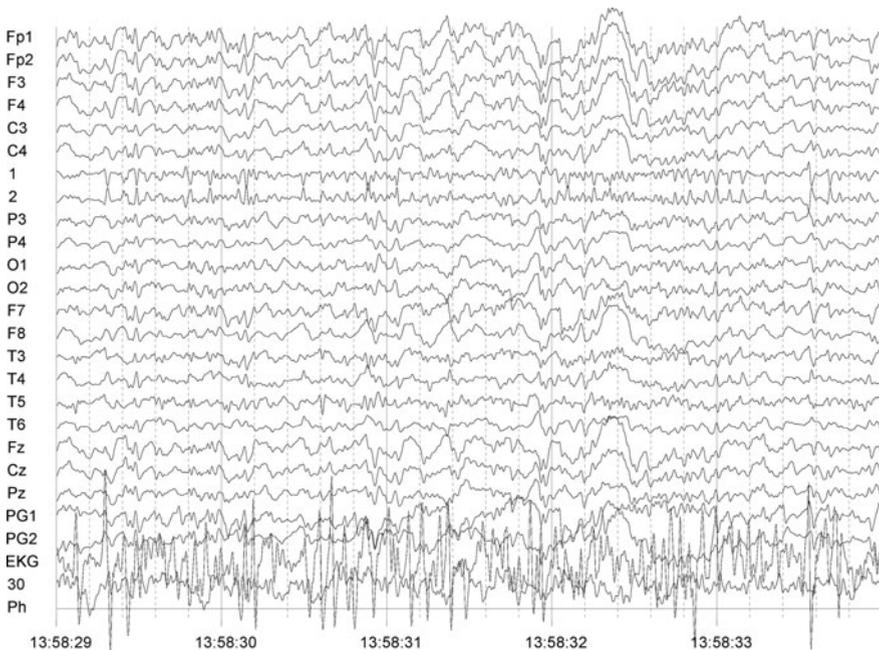

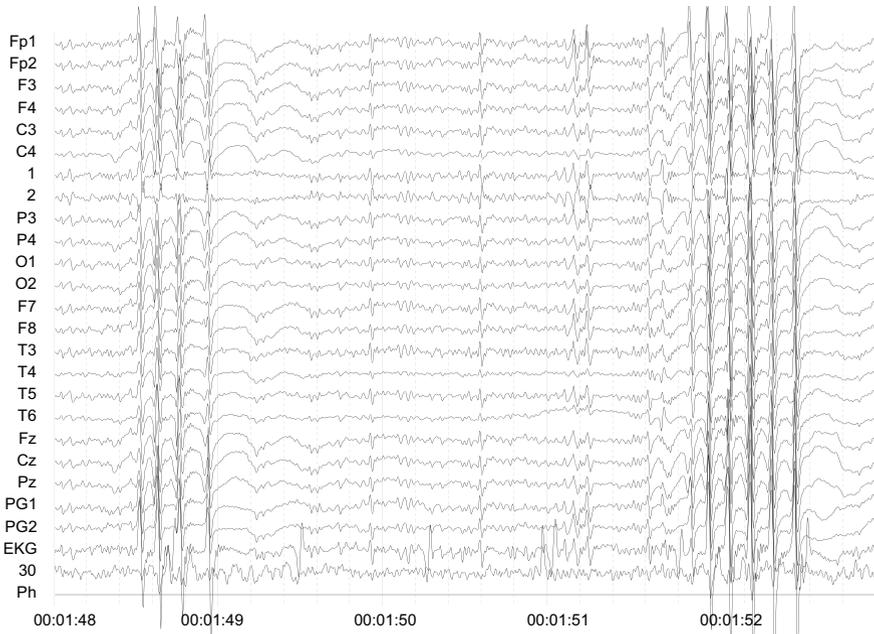Fig. 19.1: Five-second baseline EEG recording.

Fig. 19.2: Ten-second EEG recoding with spike and wave discharges.

background rhythms induced by AED treatment has been reported with highly vari-ability from patient to patient [29]. Furthermore, it has been difficult to determine if and how the strength of correlation between EEG slowing and disease progression since the intensification of drug treatment during the later stages of illness may also contribute to the EEG slowing [4]. Clinical observations have been the most com-mon method for evaluating the influence and effectiveness of AED interventions in patients with epilepsy and other neurological disorders. More specifically, efficacy of treatment is usually measured by comparing seizure frequency during treatment to a finite baseline period. EEG recordings are mainly used as supplemental diag-nostic tools in management of seizure disorders. Other than counting the frequency of occurrence of seizures as a measure for treatment effect, there is currently no reliable tool for evaluating treatment effects in patients with seizure disorders. A quantitative surrogate outcome measure using EEG recordings for patients with epilepsy is desired, especially when it is difficult to count seizures reliably, such as the case in ULD. ULD is characterized by severe myoclonus (usually triggered by some stimulus), generalized tonic-clonic seizures and the aforementioned EEG patterns [16]. In this chapter, we propose that the coupling strength between cortical regions may be used as a surrogate measure of drug efficacy in patients suffering from ULD.

The rest of this chapter is organized as follows. Background information on the patients and parameters of EEG recordings are given in Section 19.2. The methods

for identifying the coupling strength and directionality between different brain cortical regions are described in Section 19.3. The quantitative analysis, statistical tests, and results are presented in Section 19.4. The conclusion and discussion are given in Section 19.5.

19.2 Data Information

EEG recordings were acquired using a Nicolet BMSI recording system and the international 10–20 electrode placement system (Fp1, Fp2, F3, F4, C3, C4, A1, A2, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz). The EEG recordings were band-pass filtered at 0.1–70 Hz. The sample frequency was set to 250 Hz. An additional 60 Hz notch filter was applied to reduce the artifact induced by city alternating current. One hour EEG recordings were obtained before and after an add-on AED treatment. All the EEG recordings were reviewed by a board certified electroencephalographer and artifact-free baseline EEG segments were selected for the quantitative analysis. The clinical information about the ULD patients in this study is summarized in Table 19.1. One set of EEG recordings was acquired before treatment started and the other set was acquired after at least 8 weeks of treatment with an adjunctive AED. All EEG recordings were recorded approximately at the same time of day after the first dose of treatment with subjects lying supine in a relaxed state.

Table 19.1: Patient information and UMRS scores

Patient	Gender	Age	ULD Onset age	UMRS Score (before)	UMRS Score (after)
1	Female	47	9	98	48
2	Male	45	10	80	65
3	Male	50	12	50	66
4	Male	51	11	68	54

In this study, the severity of the ULD patients was clinically evaluated by performing the unified myoclonus rating scale (UMRS), a statistically validated clinical rating instrument for evaluating individuals with myoclonus. Low UMRS score indicates less severity of ULD and vice versa. UMRS in patient 2 to patient 4 was bed ridden and no points for arising, standing, or walking was included in their UMRS scores. Baseline UMRS scores suggested that patient 1 had the highest disease severity followed by patient 2, patient 4, and patient 3, respectively. After treatment, the UMRS scores indicated that severity was highest in patient 3 followed by patient 4, patient 2, and patient 1. However, clinical observations indicated that patient 1 had the mildest severity of disease before and after the treatment in this study. We speculated that EEG dynamical analysis might reconcile this discrepancy too.

19.3 Synchronization Measures

19.3.1 Mutual Information

Mutual information and nonlinear interdependence measures were applied on the EEG recordings to identify the effect of treatment on the coupling strength between different brain cortical regions [26, 27, 7, 21, 25].

In this section, we first describe the approach for estimating mutual information [15]. Let us denote the time series of two observable variables as $X = \{x_i\}_{i=1}^N$ and $Y = \{y_j\}_{j=1}^N$, where N is the fixed length of the discrete time, and the time between consecutive observations (i.e., *sampling period*) is fixed. Then the mutual information is given by

$$I(X;Y) = \sum_i \sum_j P_{x,y}(x_i, y_j) \log \left(\frac{P_{x,y}(x_i, y_j)}{P_x(x_i) P_y(y_j)} \right). \tag{19.1}$$

One can obtain the mutual information between X and Y using the following equation [5]:

$$I(X;Y) = H(X) + H(Y) - H(X,Y), \tag{19.2}$$

where $H(X), H(Y)$ are the entropies of X, Y and $H(X,Y)$ is the joint entropy of X and Y. Entropy for X is defined by

$$H(X) = - \sum_i p(x_i) \log p(x_i). \tag{19.3}$$

The units of the mutual information depends on the choice on the base of logarithm. The natural logarithm (base e) is used in the study, therefore, the unit of the mutual information is *nat*. For X and Y time series we define $d_{ij}^{(x)} = \|x_i - x_j\|, d_{ij}^{(y)} = \|y_i - y_j\|$ as the distances between x_i and y_i and every other point in matrix spaces X and Y. One can rank these distances and find the k-nearest neighbor (knn) for every x_i and y_i. In the space spanned by X, Y, similar distance rank method can be applied for $Z = (X, Y)$ and for every $z_i = (x_i, y_i)$ one can also compute the distances $d_{ij}^{(z)} = \|z_i - z_j\|$ and determine the knn according to some distance measure. The maximum norm is used in this study:

$$d_{ij}^{(z)} = \max\{\|x_i - x_j\|, \|y_i - y_j\|\}, \quad d_{ij}^{(x)} = |x_i - x_j|. \tag{19.4}$$

Next, let $\frac{\varepsilon(i)}{2}$ be the distance between z_i and its kth neighbor. In order to estimate the joint probability density function ($p.d.f.$), we consider the probability $P_k(\varepsilon)$ which is the probability that for each z_i the kth nearest neighbor is at a distance $\frac{\varepsilon(i)}{2} \pm d\varepsilon$ from z_i. This $P_k(\varepsilon)$ represents the probability for $k-1$ points to have the distance less than the kth nearest neighbor and $N-k-1$ points have distance greater than $\frac{\varepsilon(i)}{2}$ and $k-1$ points have distance less than $\frac{\varepsilon(i)}{2}$. $P_k(\varepsilon)$ is obtained using the multinomial distribution:

$$P_k(\varepsilon) = k \binom{N-1}{k} \left(\frac{dp_i(\varepsilon)}{d\varepsilon}\right) p_i^{k-1}(1-p_i)^{N-k-1}, \tag{19.5}$$

where p_i is the mass of the ε-ball. Then the expected value of $\log p_i$ will be

$$E(\log p_i) = \psi(k) - \psi(N), \tag{19.6}$$

where $\psi(\cdot)$ is the *digamma function*:

$$\psi(t) = \Gamma(t)^{-1} \frac{d\Gamma(t)}{dt}, \tag{19.7}$$

where $\Gamma(\cdot)$ is the gamma function. It holds when $\psi(1) = C$ where C is the Euler – Mascheroni constant ($C \approx 0.57721$). The mass of the ε-ball can be approximated if the *p.d.f* inside the ε-ball is uniform (*epsilon* is chosen sufficiently small so that the uniform distribution is valid) by

$$p_i(\varepsilon) \approx c_{d_x} \varepsilon_x^d P(X = x_i), \tag{19.8}$$

where c_{d_x} is the number of points within the unit ball in the d_x-dimensional space. From Equation (19.8) we can find an estimator for $P(X = x_i)$:

$$\log[P(X = x_i)] \approx \psi(k) - \psi(N) - dE(\log \varepsilon(i)) - \log c_{d_x}. \tag{19.9}$$

Finally we obtain the entropy estimator for X [14]:

$$\hat{H}(X) = \psi(N) - \psi(k) + \log c_{d_x} + \frac{d_x}{N} \sum_{i=1}^{N} \log \varepsilon(i), \tag{19.10}$$

where $\varepsilon(i)$ is twice the distance from x_i to its k-th neighbor in the d_x-dimensional space. For the joint entropy we have

$$\hat{H}(X,Y) = \psi(N) - \psi(k) + \log(c_{d_x} c_{d_y}) + \left(\frac{d_x + d_y}{N}\right) \sum_{i=1}^{N} \log(\varepsilon(i)). \tag{19.11}$$

The $I(X;Y)$ is now ready to be estimated by Equation (19.2). The problem with this estimation is that a fixed number k is used in all estimators but the distance metric in different scaled spaces (marginal and joint) are not comparable . To avoid this problem, instead of using a fixed k, $n_x(i) + 1$ and $n_y(i) + 1$ are used in obtaining the distances (where $n_x(i)$ and $n_y(i)$ are the number of samples contained in the bin $[x(i) - \frac{\varepsilon(i)}{2}, x(i) + \frac{\varepsilon(i)}{2}]$ and $[y(i) - \frac{\varepsilon(i)}{2}, y(i) + \frac{\varepsilon(i)}{2}]$, respectively) in the *x*–*y* scatter diagram. Equation (19.10) becomes

$$\hat{H}(X) = \psi(N) - \psi(n_x(i) + 1) + \log c_{d_x} + \frac{d_x}{N} \sum_{i=1}^{N} \log \varepsilon(i). \tag{19.12}$$

Finally the Equation (19.2) is rewritten as

$$I_{\text{knnr}}(X;Y) = \psi(k) + \psi(N) - \frac{1}{N}\sum_{i=1}^{N}[\psi(n_x(i)+1) + \psi(n_y(i)+1)]. \qquad (19.13)$$

19.3.2 Nonlinear Interdependencies

Arnhold et al. [2] introduced the nonlinear interdependence measures for characterizing directional relationships (i.e., driver and response) between two time sequences [2]. Given two time series x and y, using the method of delay we obtain the delay vectors $x_n = (x_n, ..., x_{n-(m-1)\tau})$ and $y_n = (x_n, ..., x_{n-(m-1)\tau})$, where $n = 1, ...N$, m is the embedding dimension and τ denotes the time delay [34]. Let $r_{n,j}$ and $s_{n,j}$, $j = 1, ..., k$ denote the time indices of the k nearest neighbors of x_n and y_n. For each x_n, the mean Euclidean distance to its k neighbors is defined as

$$R_n^k(X) = \frac{1}{k}\sum_{j=1}^{k}(x_n - x_{r_{n,j}})^2, \qquad (19.14)$$

and the Y-conditioned mean squared Euclidean distance is defined by replacing the nearest neighbors by the equal time partners of the closest neighbors of y_n:

$$R_n^{(k)}(X|Y) = \frac{1}{k}\sum_{j=1}^{k}(x_n - x_{s_{n,j}})^2. \qquad (19.15)$$

For EEG, the delay $\tau = 5$ is estimated using auto mutual information function, the embedding dimension $m = 10$ is obtained using Cao's method and the Theiler correction is set to $T = 50$ (Theiler correction corresponds to the T first sample points omitted from our analysis) [3, 35]. If x_n has an average Euclidean radius $R(X) = (1/N)\sum_{n=1}^{N}R_n^{(N-1)}(X)$, then $R_n^{(k)}(X|Y) \approx R_n^{(k)}(X) < R(X)$ if the systems are strongly correlated, while $R_n^{(k)}(X|Y) \approx R(X) > R_n^{(k)}(X)$ if they are independent [24]. Accordingly, the interdependence measure $S^{(k)}(X|Y)$ can be defined as

$$S^{(k)}(X|Y) = \frac{1}{N}\sum_{n=1}^{N}\frac{R_n^{(k)}(X)}{R_n^{(k)}(X|Y)}. \qquad (19.16)$$

Since $R_n^{(k)}(X|Y) \geq R_n^{(k)}(X)$ by construction,

$$0 < S^{(k)}(X|Y) \leq 1. \qquad (19.17)$$

Low values of $S^k(X|Y)$ indicate independence between X and Y, while high values indicate synchronization. Arnhold et al. [2] introduced another nonlinear interdependence measure $H^{(k)}(X|Y)$ as

$$H^{(k)}(X|Y) = \frac{1}{N}\sum_{n=1}^{N}log\frac{R_n(X)}{R_n^{(k)}(X|Y)}. \qquad (19.18)$$

$H^{(k)}(X|Y) = 0$ if X and Y are completely independent, while it is possible if closest that closest in Y implies also closest in X for equal time indexes. $H^{(k)}(X|Y)$ would be

negative if close pairs in Y would correspond mainly to distant pairs in X. $H^{(k)}(X|Y)$ is a linear measure thus it is more sensitive to weak dependencies compared to mutual information. Arnhold et al. [2] also showed H was more robust against noise and easier to interpret than S. Since H is not normalized Quian Quiroga et al. [26] introduced another $N(X|Y)$:

$$N^{(k)}(X|Y) = \frac{1}{N} \sum_{n=1}^{N} \frac{R_n(X) - R_n^{(k)}(X|Y)}{R_n(X)}, \qquad (19.19)$$

which is normalized between 0 and 1. The opposite interdependencies $S(Y|X)$, $H(Y|X)$, and $N(Y|X)$ are defined in complete analogy and they are in general not equal to $S(X|Y)$, $H(X|Y)$, and $N(X|Y)$, respectively. Using nonlinear interdependencies on several chaotic models (Lorenz, Roessler, and Hénon models) Quian Quiroga et al. [24] showed the measure H is more robust than S.

The asymmetry of above nonlinear interdependencies is the main advantage over other synchronization measures. This asymmetry property can give directionality of nonlinear interdependence between different cortical regions, and reflects different properties of brain functions when it is important to detect causal relationships. It should be clear that the above nonlinear interdependencies measures were bivariate measures. Finally, although directional measures quantify the "driver-response" relationship for a given input, the system under study might be driven by other unobserved sources.

19.4 Statistical Tests and Data Analysis

All mutual informations between all pairs of electrodes were computed (excluding reference channels and channels with themselves). In Fig. 19.3, we present the amount of mutual information for every patient before and after treatment. For every heatmap figure, every axis corresponds to the channels and the intensity of each pixel correspond to the amount of mutual information (in nats). Qualitatively, we can see that the first column plots are darker than the second column plots, implying that a mutual information decoupling occurs. In order to statistically validate this assumption we performed a paired t-test with replacement.

For this we used bootstrap resampling technique [8] to investigate the variability of the strength of interdependence among different brain cortical areas. In bootstrap resampling, we randomly sample, with replacement, 10.24 s continuous EEG recordings. We emphasize that resample should be performed on the parts of EEG where no SWD is presented.

The reference A1 and A2 channels (inactive regions) were excluded from the analysis. Two sample t-test ($N = 30$, $\alpha = 0.05$) was used to test the statistical differences on mutual information and nonlinear interdependence during, before, and after treatment. Low mutual information between different cortex regions were observed in our subjects with less severity of ULD. Furthermore, for each patient both mutual information between different brain cortical regions decreased after 2
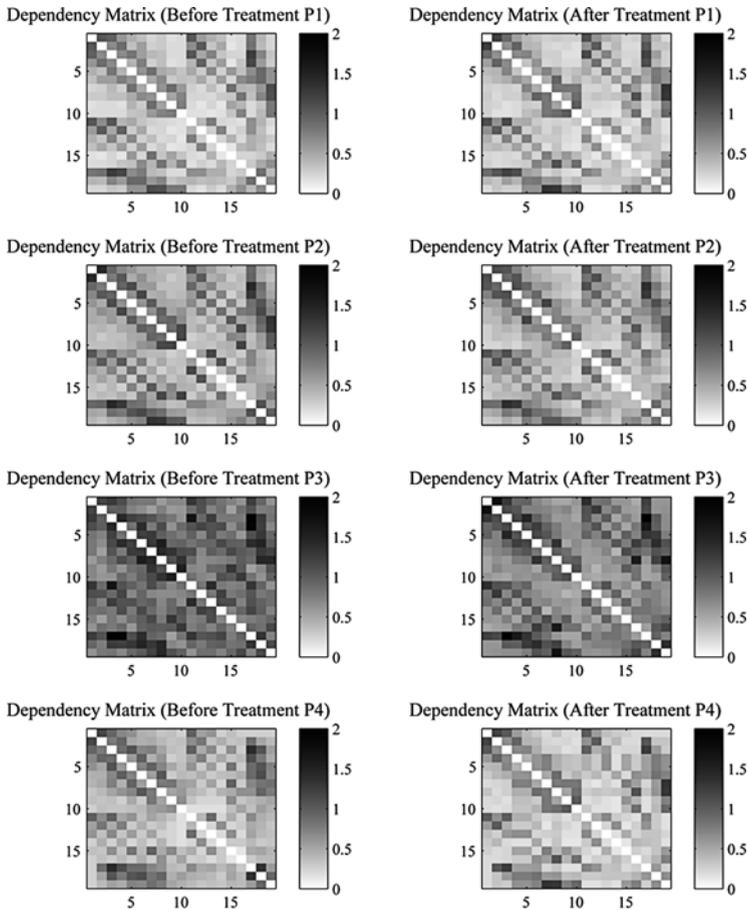
Fig. 19.3: Pairwise mutual information between for all electrodes: *Before* v.s. After Treatment. The reference electrodes A1, A2 are excluded from the above plot. For every heatmap each axis corresponds to the electrodes and the intensity of each pixel is associated with the amount of the mutual information (in nats) with reference to the colorbar. Correspondence between numbers (in axes) and electrodes can be found in the first column of Table 19.2.

months of adjunctive AED treatment. In Table 19.2, we present all the electrode pairs that were decoupled (mutual information after treatment was statistically significant lower than before treatment).

Also in order to visualize the topological distribution of the decoupled channels we divided (F-Frontal, C-central, T-temporal, P-pariental, O-occipital) and we counted the number of pairs that were decoupled between each of these regions. Results are presented in Table 19.3.

The significant "driver–response" relationship is revealed by t-test. After t-test the significant coupling strengths between Fp1 and other brain cortical regions are

Table 19.2: In this table we present all the electrode pairs that were decoupled for each one of the four patients (columns 2–4). One electrode pair is decoupled when mutual information after treatment is statistically lower (based on paired t-test) than mutual information before treatment. For example, electrodes of the first line and fourth column indicate the decoupled electrodes between electrode Fp1 (first line) for the third patient (fourth column)

Electrode	DE for P1	DE for P2	DE for P3	DE for P4
(1) Fp1	F3, C4, P4, F7 T4, T5, O1	Fp2, F3, F8, T5	F3, F7	F3, P3, Fz, T5 F8, T4, Fz
(2) Fp2	C3, C4, F8, T4 T5, Pz	Fp1, F4, T6, O2	F8	C3, C4, T5, P4 Cz
(3) F3	Fp1, C4, P4, O2	Fp1, C3, P3, Pz	Fp1	Fp1, F7
(4) F4	C4, P4 O2	Fp2, P4, O2, Fz	Cz	C4, Fz
(5) C3	Fp2, C4, P3, O1	F3, P3, O1	P3, O1	Fp2, P3, O1
(6) C4	Fp1, Fp2, F3, F4, C3	P4, T6, O2	P4, O2	Fp1, Fp2, F3
A1	N/A	N/A	N/A	N/A
A2	N/A	N/A	N/A	N/A
(7) P3	C3, O2	F3, C3, T3, Pz	O1	Fp1, C3, Cz
(8) P4	Fp1, F3, F4	F4, C4	C4	Fp2,
(9) O1	Fp1, C3	C3, T5, T3, F7	P3, Pz	C3
(10) O2	F3, F4, P3	Fp2, F4, C4	C4	Pz
(11) F7	Fp1, Fz, Pz	C3, T5, O1	Fp1	F3
(12) F8	Fp2	Fp1, C4, P4	Fp2	Fp2,
(13) T3	None	P3, O1	None	None
(14) T4	Fp1, Fp2, Cz	Fp2, O1	None	Fp2,
(15) T5	Fp1, Fp2	Fp2, C4	None	Fp2, Cz
(16) T6	None	C4	None	None
(17) Fz	F7	Fp1, F4	Cz	Fp1, Fp2, F3
(18) Pz	Fp2, F7, Cz	F3, P3, Cz	O2	O2
(19) Cz	T4, Pz	Pz	F4, Fz	Fp2, T5, P3

Table 19.3: Number of channels per patient per site that were decoupled after treatment. F, T, C, P, O correspond to the frontal, temporal, central, pariental, and occipital channels correspondingly. It is easy to see that for patient 1 we have the most decoupled channels

	Patient 1	Patient 2	Patient 3	Patient 4
F	20	20	7	18
T	5	7	0	3
C	11	7	6	9
P	9	9	3	5
O	5	7	3	2
Total	50	32	19	37

shown in Fig. 19.4. The edges with an arrow starting from Fp1 to other channels denote $N(X|Y)$, which is significant larger then $N(Y|X)$, therefore, Fp1 is the driver, and vice versa. The above results suggest the existing treatment effects on the coupling strength and the directionality between different cortex regions.
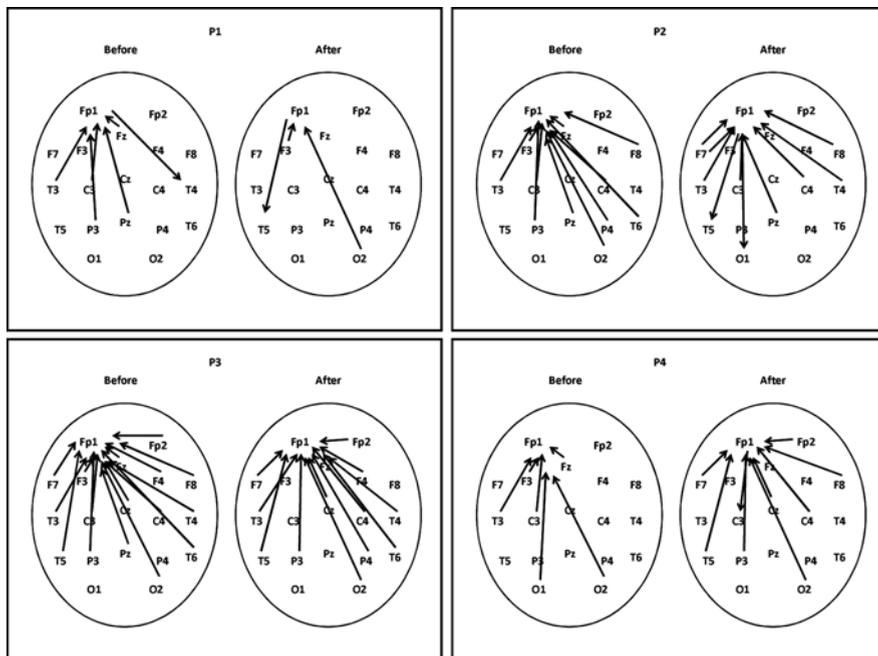
Fig. 19.4: This is the connectivity plots for the four patients derived from the nonlinear interdependence metrics that involve the channel Fp1.

19.5 Conclusion and Discussion

Effectiveness of an add-on AED in treatment of ULD was measured by the degree of reduction of UMRS after treatment. Patients with 25% or greater in UMRS were considered responders to the add-on AED treatment. As mentioned above, it is not easy to precisely perform such evaluation scheme especially in the later stages of the disease. Furthermore, the UMRS is a skewed measure that may not detect functional changes in a patient when these changes may be clinically important. In fact, changes in the UMRS scores after treatment were significant ($\geq 25\%$) only in patient 1, then patient 2, patient 3, and patient 4 had clinically meaningful improvement. Furthermore, the patient with least disease severity (patient 1) had paradoxically the higher baseline UMRS score only because patient 2–patient 4 were bed ridden and their UMRS did not include arising, standing, and walking. For this reason the UMRS scores for patient 2–patient 4 could be artificially lowered by 48 points ($4 \times 4 = 16$ points per test, 3 tests $= 48$ points). We believe that the UMRS may not be the appropriate tool for measuring AED improvements in myoclonus in ULD patients and maybe misleading at times. The present study objectively measures the mutual information and nonlinear interdependencies in the cortical network before and after an add-on AED treatment.

Results from our study agreed with clinical observations that patient 1 had the least severe ULD and patient 3 had the most severe one. The highest coupling strength in brain cortical networks was found in patient 3 before and after treatment. However, patient 3 received 50 points on the UMRS before treatment and 66 after treatment due to the fact that patient 3 was able to engage in more testing after treatment.

Although the results indicate that the mutual information and nonlinear interdependencies measures could be useful in determining the treatment effects for patients with ULD, their limitations must be mentioned. It has been reported that it is necessary to take into account the interdependence between thalamus and cortex [33, 32]. By applying Granger causality in animal studies, Sitnikova et al. suggested that onset of spike and wave discharges was associated with a rapid and significant increase of coupling strength between frontal cortex and thalamus in both directions. Furthermore, the strength of the thalamus to cortex coupling remained constantly high during seizures [31]. The decoupling between frontal and occipital cortical regions of our data after AED treatment may also be caused by decrease of a driving force deep inside the brain. The effect of the treatment may thus reduce the coupling strength between thalamus and cortex in ULD subjects.

It has been pointed out by several authors that nonlinear interdependence measures need to be applied with care [23]. For example, the embedding parameters often play important roles for nonlinear analysis involving with state space reconstruction. In this study, we used a false nearest neighbor algorithm and the mutual information function for finding the embedding dimension m and delay τ. However, it is also known that there is no guarantee that these embedding parameters are the optimal choices. Besides the intrinsic nonlinear properties, there are other sources of noise underlying the real-world EEG recordings. Our strategy, in dealing with the above potential drawbacks, is to fix the embedding parameters for the same set of EEG recordings. By fixing the embedding parameters the underlying dynamics for 10.24 s SWD-free EEG recording, and therefore attractor, can consistently quantify.

At this point we would like to point out as a future research direction the need to reproduce the same study using sleep EEG recordings and confirm if the number and the distribution of decoupled electrode sites are the same and independent of the state of vigilance. Also to prove the usefulness of the proposed study, a larger patient population is needed.

References

 1. Aarabi, A., Wallois, F., Grebe, R. Does spatiotemporal synchronization of EEG change prior to absence seizures? Brain Res **1188**,207–221 (2008)
 2. Arnhold, J., Grassberger, P., Lehnertz, K., Elger, C.E. A robust method for detecting interdependences: Application to intracranially recorded EEG. Physica D **134**, 419–430 (1999)
 3. Cao, L. Practical method for determining the minimum embedding dimension of a scalar time series. Physica D **110**(1–2), 43–50 (1997)

4. Chew, N.K., Mir, P., Edwards, M.J., Cordivari, C., Martino, D., Schneider, S.A., Kim, H.-T., Quinn, N.P., Bhatia, K.P. The natural history of Unverricht-Lundborg disease: A report of eight genetically proven cases. Mov Dis **23**(1), 107–113 (2007)

5. Cover, T.M., Thomas, J.A. Elements of Information Theory. Wiley, New York (1991)

6. Dominguez, L.G., Wennberg, R.A., Gaetz, W., Cheyne, D., Snead, O.C., Perez Velazquez, J.L. Enhanced synchrony in epileptiform activity? Local versus distant phase synchronization in generalized seizures. J Neurosci **25**(35), 8077–8084 (2005)

7. Duckrow, R.B., Albano, A.M. Comment on performance of different synchronization measures in real data: A case study on electroencephalographic signals. Phys Rev E **67**(6), 063901 (Jun 2003)

8. Efron, B., Gong, G. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am. Stat. **37**(1), 36–48 (1983)

9. Ferlazzoa, E., Magauddaa, A., Strianob, P., Vi-Hongc, N., Serraa, S., Gentonc, P. Long-term evolution of EEG in Unverricht-Lundborg disease. Epilepsy Res **73**, 219–227 (2007)

10. Iasemidis, L.D., Pappas, K.E., Gilmore, R.L., Roper, S.N., Sackellares, J.C. Preictal entrainment of a critical cortical mass is a necessary condition for seizure occurrence. Epilepsia **37S**(5), 90 (1996)

11. Iasemidis, L., Shiau, D.-S., Sackellares, J.C., Pardalos, P.M., Prasad, A. Dynamical resetting of the human brain at epileptic seizures: Application of nonlinear dynamics and global optimization tecniques. IEEE Trans Biomed Eng **51**(3), 493–506 (2004)

12. Klimesch, W. Memory processes, brain oscillations and EEG synchronization. Int J Psychophysiol, **24**(1–2), 61–100 (1996)

13. Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. Brain Res Rev **29**(2–3), 169–195 (1999)

14. Kozachenko, L.F., Leonenko, N.N. Sample estimate of entropy of a random vector. Problems Inform Transmission **23**, 95–101 (1987)

15. Kraskov, A., Stögbauer, H., Grassberger, P. Estimating mutual information. Phys Rev E **69**, 066138 (2004)

16. Lalioti, M.D., Antonarakis, S.E., Scott, H.S. The epilepsy, the protease inhibitor and the dodecamer: Progressive myoclonus epilepsy, cystatin b and a 12-mer repeat expansion. Cytogenet Genome Res **100**, 213–223 (2003)

17. Lundborg, H.B. Die progressive Myoclonus-Epilepsie (Unverrichts Myoclonie) Almqvist and Wiksell, Uppsala (1903)

18. Moretti, D.V., Miniussi, C., Frisoni, G.B., Geroldi, C., Zanetti, O., Binetti, G., Rossini, P.M. Hippocampal atrophy and EEG markers in subjects with mild cognitive impairment. Clin Neurophysiol, **118**(12), 716–2729 (2007)

19. Mormann, F., Andrzejak, R.G., Kreuz, T., Rieke, C., David, P., Elger, C.E., Lehnertz, K. Automated detection of a preseizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients. Phys Rev E **67**(2), 021912 (2003)

20. Mormann, F., Lehnertz, K., David, P., Elger, C.E. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. Physica D **144**, 358–369 (2000)

21. Nicolaou, N., Nasuto, S.J. Comment on "performance of different synchronization measures in real data: A case study on electroencephalographic signals". Phys Rev E **72**, 063901 (2005)

22. Pardalos, P.M., Sackellares, J.C., Carney, P.R., Iasemidis, L.D. Quantitative Neuroscience. Kluwer Academic Publisher, Boston, MA (2004)

23. Pereda, E., Rial, R., Gamundi, A., Gonzlez, J. Assessment of changing interdependencies between human electroencephalograms using nonlinear methods. Physica D **148**(1–2), 147–158 (2001)

24. Quian Quiroga, R., Arnhold, J., Grassberger, P. Learning driver-response relationships from synchronization patterns. Phys Rev E **61**, 5142–5148 (2000)

25. Quian Quiroga, R., Kraskov, A., Grassberger, P. Reply to "comment on 'performance of different synchronization measures in real data: A case study on electroencephalographic signals'". Phys Rev E **72**, 063902 (2005)

26. Quian Quiroga, R., Kraskov, A., Kreuz, T., Grassberger, P. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. Phys Rev E **65**, 041903 (2002)
27. Quian Quiroga, R., Kraskov, A., Kreuz, T., Grassberger, P. Reply to "comment on 'performance of different synchronization measures in real data: A case study on electroencephalographic signals'". Phys Rev E **67**, 063902 (2003)
28. Le Van Quyen, M., Martinerie, J., Baulac, M., Varela, F.J.. Anticipating epileptic seizures in real time by non-linear analysis of similarity between EEG recordings. NeuroReport **10**, 2149–2155 (1999)
29. Salinsky, M.C., Oken, B.S., Morehead, L. Intraindividual analysis of antiepileptic drug effects on EEG background rhythms. Electroencephalogr Clin Neurophysiol **90**(3), 186–193 (1994)
30. Singer, W. Synchronization of cortical activity and its putative role in information processing and learning. Annual Review of Physiology **55**, 349–374 (1993)
31. Sitnikova, E., Dikanev, T., Smirnov, D., Bezruchko, B., van Luijtelaar, G. Granger causality: Cortico-thalamic interdependencies during absence seizures in WAG/Rij rats. J Neurosci Methods **170**(2), 245–254 (2008)
32. Sitnikova, E., van Luijtelaar, G. Cortical and thalamic coherence during spike-wave seizures in WAG/Rij rats. Epilepsy Res **71**, 159–180 (2006)
33. Steriade, M., Amzica, F. Dynamic coupling among neocortical neurons during evoked and spontaneous spike-wave seizure activity. J Neurophysiol **72**, 2051–2069 (1994)
34. Takens, F. Detecting strange attractors in turbulence. In: Rand, D.A., Young, L.S. (eds.) Dynamical Systems and Turbulence, *Lecture Notes in Mathematics*, Vol. 898, pp. 366–381. Springer-Verlag, New York (1981)
35. Theiler, J. Spurious dimension from correlation algorithms applied to limited time-series data. Phys Rev A **34**, 2427–2432 (1986)
36. Unverricht, H. Die Myoclonie. Franz Deutick, Leipzig (1891)

Chapter 20
Seizure Monitoring and Alert System for Brain Monitoring in an Intensive Care Unit

J. Chris Sackellares, Deng-Shan Shiau, Alla R. Kammerdiner,
and Panos M. Pardalos

Abstract Although monitoring for most organ systems is commonly used in intensive care units (ICU), brain function monitoring relies almost exclusively upon bedside clinical observations. As a result, a large number of nonconvulsive seizures go undiagnosed every day. Recent clinical studies have demonstrated the clinical utility of continuous EEG monitoring in ICU settings. Continuous EEG is a well-established tool for detecting nonconvulsive seizures, cerebral ischemia, cerebral hypoxia, and other reversible brain disturbances in the ICU. However, the utility of EEG monitoring currently depends on the availability of expert medical professionals, and interpretation is labor intensive. Such experts are available only in tertiary care centers. We have designed a seizure monitoring and alert system (SMAS) that utilizes a seizure susceptibility index (SSI) and seizure detection algorithms based on measures that characterize the spatiotemporal dynamical properties of the EEG signal. The SMAS allows distinguishing the organized seizure patterns from more irregular and less organized background EEG activity. The algorithms and initial results in human long-term EEG recordings are described.

J. Chris Sackellares
Optima Neuroscience, Inc., Gainesville, FL 32601, USA, e-mail: csackellares@optimaneuro.com

Deng-Shan Shiau
Optima Neuroscience, Inc., Gainesville, FL 32601, USA, e-mail: dshiau@optimaneuro.com

Alla R.Kammerdiner
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611-6595, USA

Panos M. Pardalos
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611-6595, USA

20.1 Introduction

Although automatic monitoring for most organ systems, including heart, lungs, blood, etc., is common place in general and intensive care units, brain function monitoring relies almost entirely upon bedside clinical observation by medical and nursing staff. As a result, a large number of nonconvulsive seizures, with only subtle or nonspecific behavioral changes, go undiagnosed every day. Recent clinical studies have demonstrated the clinical utility of continuous EEG monitoring in such inpatient settings as emergency department (ED), intensive care unit (ICU), and epilepsy monitoring unit (EMU) [16, 17, 18, 5]. EEG-video monitoring is a standard diagnostic procedure in the EMU for pre-surgical evaluation. Continuous EEG is also well-established tool for detecting nonconvulsive seizures, cerebral ischemia, cerebral hypoxia, and other reversible brain disturbances in the ICU and the ED. Nevertheless, the utility of EEG monitoring currently depends on the availability of expert technical and medical professionals, and the task of interpreting EEG is labor intensive. Such experts are only available in tertiary care centers.

Automatic EEG monitoring has several potential practical applications, including diagnosis and monitoring of patients with epilepsy and other neurological diseases, monitoring of patients under general anesthesia, etc. An epileptic attack is usually characterized by dramatic changes in electrical recordings of the brain activity by multichannel EEG, whereas in the interictal state the EEG recording may appear completely normal or may exhibit only brief rare abnormalities. Generally, the interictal EEG often provides sufficient information to diagnose epilepsy and may even contain evidence about the possible type of epilepsy [1].

The role of long-term continuous EEG recordings in clinical practice cannot be underestimated. Although the main clinical application of continuous EEG involves differential diagnosis between non-epileptic and epileptic seizures [1], there are other useful applications (e.g., localization of site of onset of epileptic seizures, detection of seizures with subtle clinical manifestations, finding the frequency of inconspicuous seizures which may otherwise be overlooked) Development, testing, and implementation of efficient methods for automatic EEG monitoring can be extremely useful in application to monitoring brain functions in clinical settings such as ICU.

In the literature, various combinations of data mining techniques and data preprocessing methods have been applied to EEG monitoring. For instance, an approach for extracting information from the video signal in video/EEG monitoring is presented in [25]. That approach utilizes image compression method to develop a domain change detection algorithm for automatic tracking of patient's movements. Some popular preprocessing techniques applied to EEG involve such feature extraction methods as fast Fourier transform (FFT) [4, 19], wavelet transform (WT) [20, 32], computation of fractal dimensions (FD) [22], calculating different amplitude and frequency features (e.g., the average dynamic range, and frequency vector, respectively) [6], symbolic representation of spike events [23], deviation ratio topography [19], and others. The data mining methods include fuzzy classification [23], regression of transformed data [4], segmentation (ictal, interictal) [22],

event detection and classification [6], patient classification (diagnosis), neural networks [19, 20, 32]. In many studies, EEG monitoring is aimed at detection of epileptic seizures [32, 22, 6], or at detection of spike events [20, 23].

This chapter introduces the approach for automatic EEG monitoring that is based on nonlinear dynamic theory, statistics, and optimization. Based on this approach, a seizure monitoring and alert system (SMAS) is designed as an online system for generating warnings for impeding seizures by the analysis of patient's EEG recordings. The SMAS also incorporates a seizure susceptibility index (SSI) that is based on the seizure warning algorithm. The SMAS is developed with a purpose of providing medical staff information as to the likelihood of ensuing seizure and alerting the staff when seizure occurs.

The remainder of the chapter is organized as follows. Section 20.2 discusses seizure prediction and warning. Section 20.3 presents the methods involved analysis of EEG data. In particular, the methods include application of chaos theory to measure dynamical transitions in the epileptic brain via Lyapunov exponents, statistical approach to quantifying similarity between pairs of measurements, and application of quadratic optimization methods to detect the critical channels in multichannel EEG. In Section 20.4, we propose the SMAS based on an algorithm for generating automatic warnings about impending seizure from EEG, which incorporates the above methods. Finally, the conclusion follows.

20.2 Preictal Transition and Seizure Prediction

The studies investigating the possibilities for prediction of epileptic seizures date back to the late 1970s. During the 1970s and the 1980s, the linear approaches, including linear autoregression, spectral analysis, and pattern recognition techniques, were mostly applied to analysis of epileptic seizures. Some studies conducted at that time reported changes in EEG characteristic of epileptic seizures, which could only be detected a few minutes before the seizure onset. Later, beginning in the late 1980s, various nonlinear approaches based on Lyapunov exponents, correlation dimension, and different entropy measures were introduced to study the dynamical changes in the brain before, during, and after epileptic seizures. Introduction of the nonlinear methods resulted in the findings that showed characteristic changes in EEG minutes to hours before seizure onset. These results were reported in a number of papers in the 1990s, and interpreted as an evidence of existence of interictal state. Beginning in the early 2000s, the multivariate approaches become especially current in analysis of epileptic seizures. Various studies show particular importance of spatiotemporal relations in multichannel EEG data with respect to the transitions in epileptic brain. Another developing research area includes assessment of performance of different algorithms of seizure prediction. The two different techniques are proposed, namely a bootstrap based approach proposed by Andrzejak et al. and a seizure prediction characteristic method introduced by Winterhalder et al.

Beginning in 1988, techniques for analyzing nonlinear chaotic systems were applied by Iasemidis and Sackellares toward the study of the dynamical characteristics of EEG signals from patients with medically refractory epilepsy [15, 14]. Later, they showed that, from a dynamical perspective, seizures evolve in a distinctive way over minutes to hours [11, 9]. Specifically, seizures are preceded by a preictal transition, detectable in the EEG, which has characteristic spatiotemporal dynamical properties. The seizure onset represents an abrupt phase transition from a complex to a less complex (more ordered) state. The spontaneous formation of organized spatial, temporal, or spatiotemporal patterns is often present in various physical, chemical, or biological systems, and the study of these dynamical changes represents one of the most fascinating and challenging areas of scientific investigation [21]. The primary common denominator in such abrupt changes of the state of a deterministic system as above lies in the nonlinear nature of the system.

From analysis of the spatiotemporal dynamics of invasive EEG recordings in patients with medically intractable temporal lobe epilepsy, Sackellares, Iasemidis, and others first discovered and characterized a preictal transition process [10, 8]. The onset of this transition precedes the seizure for periods ranging from 0.5 to 1.5 h. In their observations, the preictal dynamical transition was characterized by

1. progressive convergence of the mean short-term Lyapunov exponents (STLmax) among specific anatomical areas (mean value entrainment), and
2. progressive phase locking of the STLmax profiles among various electrode sites (phase entrainment).

In initial studies, preictal entrainment of EEG dynamics among electrode sites was detected by visual inspection of STLmax versus time plots. More recently, methods have been developed that provide objective criteria for dynamical entrainment among electrode pairs [11, 9]. Based on these findings, an approach is developed for the automatic detection of the preictal state and prediction of impending seizures.

The discovery of the preictal dynamical transition in temporal lobe epilepsy has also been reported by other researchers. Using a modification of the correlation dimension, Elger and Lehnertz reported long-lasting and marked transitions toward low-dimensional states up to 25 min before the occurrence of epileptic seizures [2]. These findings were interpreted by them as evidence for a continual increase in the degree of synchronicity preceding the occurrence of an epileptic seizure. Martinerie et al. also found evidence for a reduction in the correlation dimension calculated from intracranial EEG recordings, beginning 2–6 min prior to seizures [26]. Both studies analyzed intracranial EEG recordings in patients with unilateral mesial temporal epilepsy. The investigators utilized an estimate of the signal complexity (integral correlation dimension). However, we have found that this measure is not reliable when applied to continuous, real-time EEG recordings.

Motivated by the studies of synchrony in communication among various brain structures [36, 37], synchronization measures have recently been applied to EEG data from epilepsy patients [28, 29, 30, 24]. In particular, Mormann et al. used a measure of phase synchronization called mean phase coherence, which is based on

circular variance computed via Hilbert transform, and reported a decease in syn-
chronization preceding epileptic seizures [28, 29]. Whereas in [24], Kraskov et al.
introduced phase synchronization with the phase based on the wavelet transform for
localization of interictal focus in temporal lobe epilepsy.

20.3 Methods

20.3.1 Chaos Theory and Epilepsy

Several studies applied chaos theory to analysis of EEG data [10, 33, 3]. In chaotic
systems, trajectories originating from very close initial conditions diverge exponen-
tially. The system dynamics can be characterized by the rate of the divergence of the
trajectories, which is measured by Lyapunov exponents and dynamical phase.

First, using the method of delays [31], the embedding phase space is constructed
from a data segment $x(t)$ with $t \in [0, T]$ so that the vector X_i of the phase space is
given by

$$X_i = (x(t_i), x(t_i + \tau), \cdots, x(t_i + (p-1)\tau),$$ (20.1)

where $t_i \in [1, T - (p-1)\tau]$, p is a chosen dimension of the embedding phase space,
and τ denotes the time delay between the components of each phase space vector.
Next, the estimate L of the short-term largest Lyapunov exponent STLmax is com-
puted as follows:

$$L = \frac{1}{N\alpha} \sum_{i=1}^{N_a} \log_2 \frac{X(t_i + \Delta t) - X(t_j + \Delta t)}{X(t_i) - X(t_j)},$$ (20.2)

where N_a is the total number of local maximum Lyapunov exponents that are esti-
mated during the time interval $[0, T]$; Δt is the evolution time for the displacement
vector $X(t_i) - X(t_j)$; $X(t_i)$ represents the point of the fiducial trajectory such that
$t = t_i$, $X(t_0) = (x(t_0), x(t_0 + \tau), \cdots, x(t_0 + (p-1)\tau))$, and $X(t_j)$ is an appropriately
selected vector that is adjacent to in the embedding phase space. In [8], Iasemidis
et al. suggested a method of estimating STLmax in the EEG data based on the Wolf's
algorithm for time series [38].

The short term largest Lyapunov exponent STLmax is proved to be an espe-
cially useful EEG feature for studying the dynamics of the epileptic brain [10, 33, 3].
Figure 20.1 shows an example of the STLmax curve derived from an EEG channel
over a 140 min time window that includes a seizure. In this example, the STLmax
values gradually decreases before the seizure and drops to the lowest point dur-
ing the ictal state. It immediately reverses to the highest point after the seizure
stops, a phenomenon that we called "seizure resetting" [12]. In addition, transi-
tions among interictal, preictal, ictal, and postictal states can be characterized by the
spatiotemporal changes in STLmax profiles among EEG channels [34]. Figure 20.2
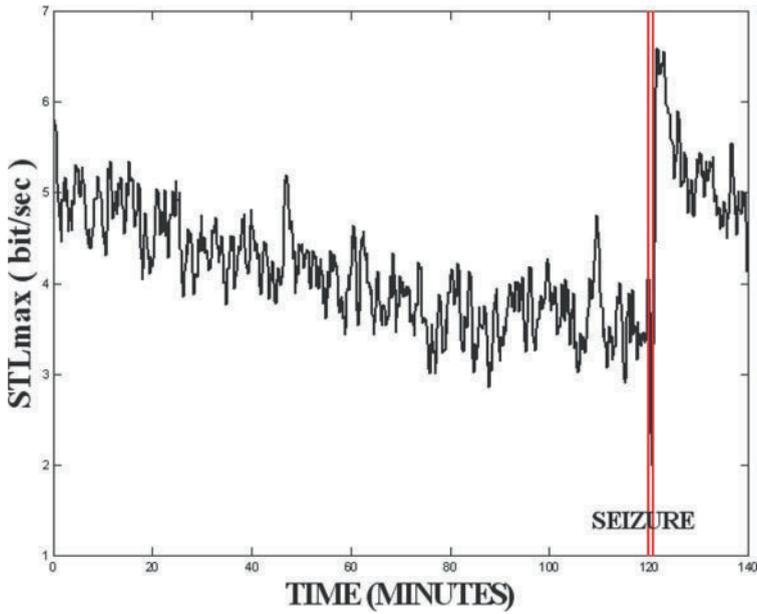shows a typical spatiotemporal pattern of STLmax profiles.

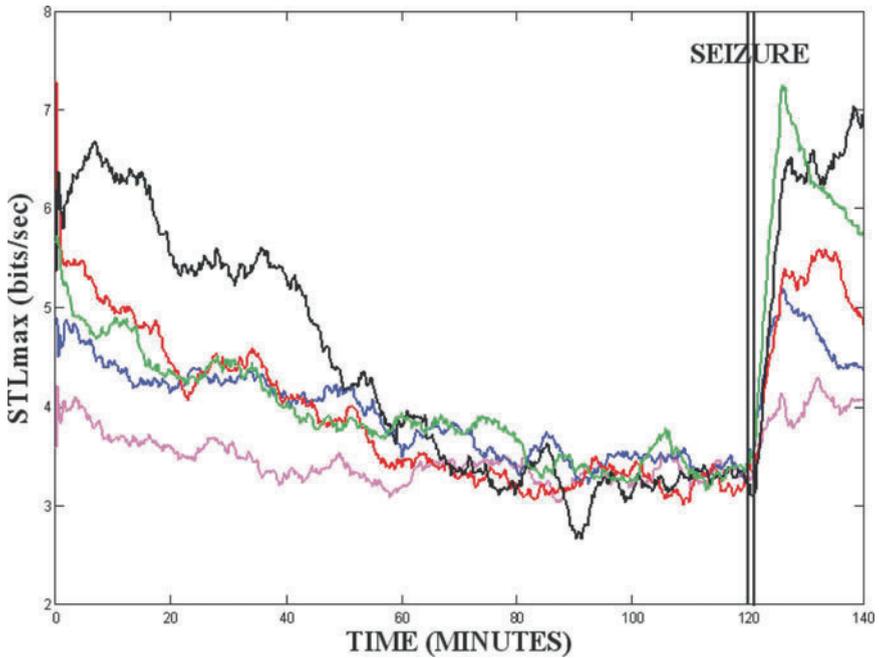Fig. 20.1: STLmax curve derived from an EEG channel over a 140-min time window that includes a seizure.

Fig. 20.2: Spatiotemporal pattern of five STLmax curves over a 140-min time window that includes a seizure.

20.3.2 Statistical Method for Pairwise Comparison of STL_{MAX}

One of the statistical measures of similarity between two different channels was introduced via T-index. It quantifies the statistical mean difference between two EEG channels with respect to their dynamics such as STLmax. As shown in [33], although the critical electrode sites involved in transition into the seizure state vary from patient to patient, the conditions for an impending seizure can be characterized by the fall in the average T-index for pairs of critical electrodes, signifying that on average, all critical electrode sites exhibit convergence in STLmax values, as shown in Fig. 20.3.
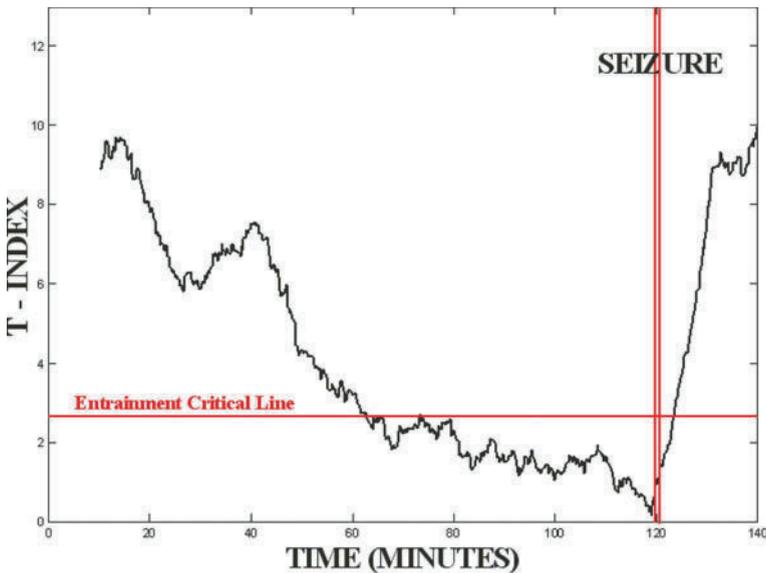
Fig. 20.3: T-index curve that quantifies the spatiotemporal pattern of five STLmax curves over a 140-min time window that includes a seizure.

Mathematically, the value of the T-index statistic at time t for the STLmax values between a pair of electrode sites i and j in a moving window W_t with n points of STLmax is given by the following formula:

$$T^{ij} = \frac{\sqrt{\eta}|\hat{\mu}^{ij}(t)|}{\hat{\sigma}^{ij}(t)}, \tag{20.3}$$

where $\hat{\mu}^{ij}$ and $\hat{\sigma}^{ij}(t)$, respectively, denote the sample mean and the sample standard deviation of the differences $STL_{MAX}^{i} - STL_{MAX}^{j}$ in the values of STLmax for channels i and j estimated successively at times $t, t+1, \cdots, t+(n-1)$.

It is proved in [13] that the T-index statistic $T^{ij}(t)$ is asymptotically distributed according to the t-distribution with $(d-1)$ degrees of freedom, where d denotes the total number of STLmax values per channel contained in a moving window W_i. Let $T^{ij}(t) > t_{\alpha/2,d-1}$ denote $(1-\alpha/2) \times 100\%$ critical value of the t-distribution with $(d-1)$ degrees of freedom. Then using the paired t-test, the sites i and j are considered *disentrained*, if $T^{ij}(t) > t_\alpha/2, d-1$; otherwise, the sites are *entrained*.

20.3.3 Finding Critical Sites by Quadratic Optimization Approach

An interesting application of optimization theory to the problem of determining critical cortical sites involved in the preictal transition into the seizure state is given by an analog of the Ising spin glass model [13]. The Ising model is defined via the Sherrington–Kirkpatrick Hamiltonian, which is used to introduce the mean-field theory of the spin glasses. The spin glass is represented by a regular lattice, with the elements at the vertices of the lattice. Furthermore, the magnetic interactions among the elements hold only for the nearest neighbors, and each element has only two possible states [27]. The Ising spin glass model is widely utilized to examine phase transitions in the field of statistical physics. More specifically, the ground state problem in the Ising model of finding the spin glass configurations of the minimal energy can be applied to determine phase transitions in dynamical systems.

Since the Ising model is defined on a regular lattice, it admits the following natural representation in terms of graph theory. Given a graph $G(V,E)$ with the vertex set $V = \{v_1, \cdots, v_n\}$ of size n, and the edge set E, let us assign a weight $\overline{\omega}_{ij}$ to every edge $(v_i, v_j) \in E$. Here the weights $\overline{\omega}_{ij}$ represent the interaction energy between the elements (vertices) v_i and v_j of the spin glass (graph $G(V,E)$). Let $\sigma \in \{+1, -1\}$ denote a magnetic spin variable associated with a given vertex v_i of the spin glass graph. Then the spin glass configuration $\sigma_{\min}$ with the minimum energy of magnetic interactions is found by minimizing the Hamiltonian H:

$$H(\sigma) = -\sum_{1 \le i,j \le n} \overline{\omega}_{ij}\sigma_i\sigma_j, \tag{20.4}$$

over all possible configurations $\sigma = (\sigma_1, \cdots, \sigma_n) \in \{-1, +1\}$.

This problem (20.4) is equivalent to combinatorial formulation of the quadratic bivalent programming problem [7]. Analogously to the quadratic programming formulation (20.4) for the Ising spin glass model, the problem of finding the cortical sites critical with respect to the transition of the epileptic brain into the seizure state is formulated as a quadratic 0–1 programming problem.

Let $x_\in\{0,1\}$ denote the choice between selecting $(x_i = 1)$ and disregarding $(x_i = 0)$ the information from the channel i, then by introducing a T-index that represents a statistical measure of the similarity in the STLmax values between a pair of EEG channels, the problem is formulated as finding critical electrodes that minimize the average value of T-index statistic.

Suppose that the T-index values T^{ij} between a pair (i,j) of electrode sites are given by the elements of an *nxn* real-valued matrix Q, and let the selection of the critical electrodes be represented by an *n*-dimensional vector $x = (x_1, \cdots, x_n) \in \{0,1\}^n$, where the selection of the cortical site i corresponds to $x_i = 1$, while $x_i = 0$ indicates that the channel i is not selected. By adding a linear constraint on the number $k(1 \leq k \leq n)$ of selected channels,

$$\sum_{1 \leq i,j \leq n} x_i = k, \tag{20.5}$$

the problem of determining k critical electrodes sites involved in transition into the ictal state based on the matrix of T-index values Q can be formulated as the following quadratic 0–1 knapsack problem:

$$\min x^T Q x, \ s.t \sum_{\Sigma 1 \leq i,j \leq n} x_i = k, x \in \{0,1\}^n. \tag{20.6}$$

By introducing the penalty term to guarantee that the optimal solution satisfies the constraint (20.5) the problem (20.6) can be equivalently reformulated as a quadratic 0–1 programming problem:

$$\min x^T Q x + c \left(\sum_{1 \leq i,j \leq n} x_i - k \right)^2, \ s.t \ x \in \{0,1\}^n, \tag{20.7}$$

where the penalty constant c is computed from $Q = (q_{ij})_{i,j=1}^n$ as

$$c = 2 \left(\sum_{1 \leq i,j \leq n} |q_{ij}| \right) + 1. \tag{20.8}$$

There are several computational approaches to solving problem (20.7), including a branch and bound procedure (B&B) with dynamical rule for fixing variables, a linearization approach to reformulate (20.7) as an integer programming (IP) problem by introducing additional variables to represent $x_i \times x_j$, and utilizing Karush–Khun-Tucker optimality conditions to obtain mixed integer linear programming (MILP) reformulation.

20.4 Two Main Components of the Seizure Monitoring and Alert System

The proposed seizure monitoring and alert system (SMAS) consists of two main components, the algorithm for generating automatic seizure warnings, and the seizure susceptibility index (SSI).

20.4.1 Algorithm for Generating Automatic Warnings about Impending Seizure from EEG

Based on the methodology, presented in the previous section, the algorithm for generating automatic warnings about possible seizure from multichannel EEG recording can be outlined as follows. The algorithm consists of two key phases, namely the training and the seizure detection. During the first stage, the EEG data are recorded and analyzed to determine the critical sites with respect to the brain's transition into seizure, which are individual to each patient. The sites found during the training phase are used to automatically detect the conditions signalizing of impending seizures. The algorithm for generating automatic warnings includes the following steps:

- Training phase:

  1. Collect EEG data for a given number $m \geq 1$ of the first seizures detected with manual assistance of a qualified person;
  2. For each seizure, determine a given number k of critical electrode sites by following steps:
     - estimate the STLmax values for all electrodes in a 10-min window immediately before the seizure
     - compute matrix $Q = (T^{ij})_{i,j}$ of T-indices of the STLmax values between a pair (i, j) of electrode sites
     - solve the corresponding quadratic 0–1 problem (20.7)
  3. Among m different $C_i, 1 \leq i \leq m$ sets of critical sites, select:
     - either the sites that are common to all m seizures, i.e., $\bigcap_{1 \leq i \leq m} C_i$
     - or electrode sites that can be found in most seizures

- Seizure alert phase:

  1. Compute the critical threshold value $t_{\alpha/2, d-1}$, where d is the total number of the STLmax values per channel in 10-min window
  2. Sequentially analyze the EEG from the electrode sites selected in Step 3 of the training phase as follows:
     - calculate STLmax values
     - compute corresponding T-indices $T^{ij}(t)$ between pairs of selected critical sites
     - go to the next step (Step 3 below) when the T-indices $T^{ij}(t)$ drops below the threshold $t_{\alpha/2, d-1}$, i.e., $T^{ij}(t) > t_{\alpha/2, d-1}$
     - otherwise continue sequentially analyzing the data
  3. If the threshold-drop time t lies within some fixed prediction horizon h of the previous warning, then go back to the previous step (Step 2 of the seizure alert phase); otherwise generate a warning.

The proposed algorithm is a version of the adaptive threshold seizure warning algorithm (ATSWA) introduced by Sackellares et al. in [35]. In particular, the main difference between the new version and ATSWA is that the proposed version utilizes

information from several manually identified seizures instead of a single seizure as in ATSWA.

Analyses of sensitivity, specificity, and predictive power of ATSWA with respect to various seizure warning horizons as compared to periodic and random warning schemes has shown that ATSWA performs significantly better.

One disadvantage of SMAS is that the seizure warnings only provide long-term anticipation of impending seizures in a fixed time interval (i.e., seizure warning horizon). Although this is valuable information to epileptic patients and clinicians, there is no information within the warning horizon. Therefore, we are developing seizure susceptibility indices, probability measures (between 0 and 1) that represent the likelihood of an impending seizure. Since our seizure algorithms are based on the dynamical descriptors of EEG, SSI should be generated in real time in a form of probability index by analyzing the distribution of dynamical descriptors.

20.5 Conclusions

A useful seizure monitoring and alert system is capable of not only generating automatic warnings of impending seizures from EEG recordings, but also quantifying and outputting the information on the likelihood of a seizure occurrence. The techniques described in this study appear to be potentially useful in wide range of applications for brain monitoring, including the ICU, and could potentially revolutionize the care for patients with neurological disorders.

References

1. Binnie, C. Long term EEG recording and its role in clinical practice. IEE Colloquium on Data Logging of Physiological Signals, pp. 5/1–5/2 (1995)
2. Elger, C., Lehnertz, K.: Seizure prediction by non-linear time series analysis of brain electrical activity. Eur J Neurosci **10**, 786–789 (1998)
3. Freeman, W.J. Strange attractors that govern mammalian brain dynamics shown by trajectories of EEG potentials. IEEE Trans CaS **35**, 781–784 (1988)
4. Griffiths, M., Grainger, P., Cox, M., Preece, A. Recent advances in EEG monitoring for general anesthesia, altered states of consciousness and sports performance science. In: The 3rd IEE International Seminar on Medical Applications of Signal Processing, pp. 1–5 (2005)
5. Hirsch, L. Continuous EEG monitoring in the intensive care unit: an overview. J Clin Neurophysiol **21**(5), 332–340 (2004)
6. Hoeve, M.J., Jones, R., Carroll, G., Goelz, H. Automated detection of epileptic seizures in the EEG. In: Engineering in Medicine and Biology Society, Proceedings of the 23rd Annual International Conference of the IEEE, pp. 943–946 (2001)
7. Horst, R., Pardalos, P., Thoai, N. Introduction to Global Optimization, 2nd edn. Kluwer Academic Publishers, Boston (2000)
8. Iasemidis, L. On the dynamics of the human brain in temporal lobe epilepsy. Ph.D. thesis, University of Michigan, Ann Arbor, Ph.D. Dissertation (1991)

9. Iasemidis, L., Principe, J., Sackellares, J. Measurement and quantification of spatiotemporal dynamics of human epileptogenic seizures. In: Nonlinear Signal Processing in Medicine. IEEE Press, New Jersey (2000)

10. Iasemidis, L., Sackellares, J. The evolution with time of the spatial distribution of the largest Lyapunov exponent on the human epileptic cortex. In: Measuring Chaos in the Human Brain. World Scientific, Singapore (1991)

11. Iasemidis, L., Sackellares, J., Roper, S., Gilmore, R. An automated seizure prediction paradigm. Epilepsia **39**(6), 207 (1998)

12. Iasemidis, L., Shiau, D., Chaovalitwongse, W., Sackellares, J., Pardalos, P., Principe, J., Carney, P., Prasad, A., Veeramani, B., Tsakalis, K. Adaptive epileptic seizure prediction system. IEEE Trans Biomed Eng **50**(5), 616–627 (2003)

13. Iasemidis, L., Shiau, D.S., Sackellares, J., Pardalos, P., Prasad, A. Dynamical resetting of the human brain at epileptic seizures: application of nonlinear dynamics and global optimization techniques. IEEE Trans Biomed Eng **51**(3), 493–506 (2004)

14. Iasemidis, L., Zaveri, H., Sackellares, J.C., Williams, W. Linear and nonlinear modeling of ECoG in temporal lobe epilepsy. 25th Annual Rocky Mountain Bioengineering Symposium **24**, 187–193 (1988)

15. Iasemidis, L., Zaveri, H., Sackellares, J., Williams, W. Phase space analysis of EEG data in temporal lobe epilepsy. In: IEEE Engineering and Medicine & Biology Society 10th Annual International Conference. New Orleans (1988)

16. Jordan, K. Continuous EEG and evoked potential monitoring in the neuroscience intensive care units. J Clin Neurophysiol **10**(4), 445–475 (1993)

17. Jordan, K. Continuous EEG monitoring in the neuroscience intensive care unit and emergency department. J Clin Neurophysiol **16**(1), 14–39 (1999)

18. Jordan, K. Emergency EEG and continuous EEG monitoring in acute ischemic stroke. J Clin Neurophysiol **21**(5), 341–352 (2004)

19. Kaji, Y., Akutagawa, M., Shichijo, F., Kinouchi, H.N.Y., Nagahiro, S. Analysis for brain activities during operations using measured EEG. In: Conference on Proceedings of IEEE Engineering in Medicine and Biology Society, pp. 6003–6006 (2005)

20. Kalayci, T., Ozdamar,O. Wavelet preprocessing for automated neural network detection of EEG spikes. IEEE Eng Med Biol Mag **14**(2), 160–166 (1995)

21. Kelso, J., Mandel, A., Shlesinger, M. Dynamical patterns in complex systems. World Scientific, Singapore (1988)

22. Kirlangic, M., Perez, D., Kudryavtseva, S., Griessbach, G., Henning, G., Ivanova, G. Fractal dimension as a feature for adaptive electroencephalogram segmentation in epilepsy. In: Engineering in Medicine and Biology Society. Proceedings of the 23rd Annual International Conference of the IEEE, pp. 1573–1576 (2001)

23. Kittel, W., Epstein, C., Hayes, M. EEG monitoring based on fuzzy classification. In: Circuits and Systems, Proceedings of the 35th Midwest Symposium, pp. 699–702 (1992)

24. Kraskov, A., Kreuz, T., Quiroga, R.Q., Grassberger, P., Mormann, F., Lehnertz, K., Elger, C. Phase synchronization using continuous wavelet transform of the EEG for interictal focus localization in mesial temporal lobe epilepsy. Epilepsia **42**(7), 43 (2001)

25. Liu, Q., Sun, M., Scheuer, M., Sclabassi, R. Patient tracking for video/EEG monitoring based on change detection in DCT domain. In: Proceedings of the IEEE 31st Annual Northeast, pp. 114–115 (2005)

26. Martinerie, J., Adam, C., Quyen, M.L.V., Baulac, M., Clemenceu, S., Renault, B., Varela, F. Epileptic seizures can be anticipated by non-linear analysis. Nat Med **4**, 1173–1176 (1998)

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

28. Mormann, F., Andrzejak, R., Kreuz, T., Rieke, C., David, P., Elger, C., Lehnertz, K. Automated preictal state detection based on a decrease in synchronization in intracranial electroencephalography recordings from epilepsy patients. Phys Rev E **67**, 021912 (2003)

29. Mormann, F., Kreuz, T., Andrzejak, R., David, P., Lehnertz, K., Elger, C. Epileptic seizures are preceded by a decrease in synchronization. Epilepsy Res **53**, 173 (2003)

30. Mormann, F., Lehnertz, K., David, P., Elger, C. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. Physica D **144**, 358 (2000)
31. Packard, N., Crutchfield, J., Farmer, J., Shaw, R. Geometry from a time series. Phys Rev Lett **45**(9), 712–716 (1980)
32. Park, H., Lee, Y., Lee, D.S., Kim, S., et al. Detection of epileptiformactivity using wavelet and neural network. Eng Med Biol Soc **3**, 1194 (1997)
33. Sackellares, J., Iasemidis, L., Gilmore, R., Roper, S. Epilepsy – when chaos fails.In: Chaos in the brain? World Scientific, Singapore (2002)
34. Sackellares, J., Iasemidis, L., Shiau, D.S., Pardalos, P., Carney, P. Spatio-temporal transitions in temporal lobe epilepsy. In: Quantitative Neuroscience: Models, Algorithms, Diagnostics, and Theraputic Applications, pp. 223–238. Kluwer Academic Publishers, Norwell, MA (2004)
35. Sackellares, J., Shiau, D.S., Principe, J., Yang, M., Dance, L., Sucharitdamrong, W., Chaovalitwongse, W., Pardalos, P., Iasemidis, L. Predictability analysis for an automated seizure prediction algorithm. J Clin Neurophysiol **23**(6), 509–520 (2006)
36. Varela, F.J. Resonant cell assemblies: A new approach to cognitive functions and neuronal synchrony. Biol Res **28**, 81 (1995)
37. Varela, F.J., Lachaux, J.P., Rodriguez, E., Martinerie, J. The brain web: Phase synchronization and large-scale integration. Nat Rev Neurosci **2**(229) (2001)
38. Wolf, A., Swift, J., Swinney, H., Vastano, J. Determining Lyapunov exponent from a time series. Physica D **16**, 285–317 (1985)