# Preface

For many workers in the field, *Neuroinformatics* is the use of databases, the World Wide Web, and visualization in the storage and analysis of neuroscience data. However, in this book we see the structuring of masses of data by a variety of computational models as essential to the future of neuroscience, and thus broaden this definition to include *Computational Neuroscience*, the use of computational techniques and metaphors to investigate relations between neural structure and function.

In recent years, the Human Genome Project has become widely known for its sequencing of the complete human genome, and its placing of the results in comprehensive databases such as GenBank. This has been made possible by advances in gene sequencing machinery that have transformed the sequencing of a gene from being a major research contribution publishable in *Science* to an automated process costing a few cents per base pair. The resultant data are of immense importance but are rather simple, since the key data are in the form of annotated base-pair sequences of DNA. By contrast, the Human Brain Project (HBP)—a consortium of U.S. federal agencies funding work in neuroinformatics—has the problem of building databases for immensely heterogeneous sets of data. The brain is studied at multiple levels, from the behavior of the overall organism through the diversity of brain regions down through specific neural circuits and beyond. The human brain contains on the order of $10^{11}$ neurons, such neurons may have tens of thousands of synapses (connections) from other neurons, and these synapses are themselves complex neurochemical structures containing many macromolecular channels or receptors. Not only do we have to contend with the many orders of magnitude linking the finest details of neurochemistry to the overall behavior of the organism, but we also have to integrate data gathered by many different specialists. Neuroanatomists characterize the brain's connectivity patterns. Neurophysiologists characterize neural activity and the "learning rules" which summarize the conditions for, and dynamics of, change. Neurochemists seek the molecular mechanisms which yield these "rules", while computational neuroscientists seek to place all these within a systems perspective.

The first "map" of neuroinformatics was provided in the edited volume *Neuroinformatics: An Overview of the Human Brain Project* (Koslow and Huerta, 1997). The present volume is both broader than its predecessor—it gives a much fuller view of the computational neuroscience components of neuroinformatics and of underlying issues in research on databases—and narrower in that it has rather little to say on human brain imaging, which is a major thrust of the HBP consortium. Indeed, the book focuses on the work of the University of Southern California Brain Project (USCBP), funded in part by a Program Project (P20) grant from the Human Brain Project (P01MH52194), with contributions from NIMH, NIDA, and NASA. At first this focus might seem a weakness. However, what has distinguished USCBP from other HBP efforts is its emphasis on *integration,* and we are thus able to offer an integrated overview of neuroinformatics which was missing in the previous volume, which gathered contributions from a number of laboratories with very different foci.

We do not claim that our work subsumes the many contributions made by other laboratories engaged in neuroinformatics. Much research has been conducted on neuroinformatics, with HBP funding and under other auspices, both in the U.S. and elsewhere. To get a sense of what is being done beyond the material presented in this book, the reader should start with the HBP Website (http://www.nimh.nih.gov/neuroinformatics/index.cfm), and follow the links from there.

What we do claim is that the present volume offers a unified perspective that is available nowhere else, a perspective in which the diverse contributions of many laboratories can be better appreciated and evaluated than would otherwise be possible. Indeed, the material in this book grows not only from our own research but also from our experience in teaching, three times in five years, a graduate course in neuroinformatics to a total of 80 students in biomedical engineering, computer science, neuroscience, and other departments. We have thus kept the needs of graduate students coming to neuroinformatics research from diverse disciplines, as well as the needs of neuroscientists seeking a comprehensive introduction to neuroinformatics, very much in mind. In this spirit, this book aims to show how to approach "Computing the Brain," integrating database, visualization, and simulation technology to gain a deeper, more integrated view of the data of neuroscience, assisting the conversion of data into knowledge.

The book is divided into 6 parts:

**Part 1. Introduction:** The first chapter, "Neuroinformatics: The Issues," both sets the stage for the study of neuroinformatics in general, and also introduces the feature that makes USCBP unique among all other HBP projects, namely that we have created the **NeuroInformatics Workbench**, a *unified architecture for neuroinformatics*. This is a suite of tools to aid the neuroscientist in constructing and using databases, and in visualizing and linking models and data. At present, the Workbench contains three main components: **NSLJ**, a modular, Java-based language and environment for neural simulation; **NeuroCore**, a system for constructing and using neuroscience databases; and **NeuARt**, a viewer for atlas-based neural data (the **Neu**ro**A**natomical **R**egis**t**ration Viewer). The second chapter, "Introduction to Databases," provides the expository role of its title. Our approach to databases exploits object-relational database management and is adaptable to any database management of this kind. The specific implementation of our database uses the Informix Universal Server which provides the ability to construct new data types as Datablades (a new base type along with its associated functions) which can be "plugged in" to the Informix architecture. These facilities are described in the appendices.

**Part 2. Modeling and Simulation:** We start with a chapter, "Modeling the Brain" which provides an overview of work in computational neuroscience, providing both a general perspective and a brief sampling of models constructed at USCBP. For a variety of behaviors, we seek to understand what must be added to the available databases on neural responsiveness and connectivity to explain the time course of cellular activity, and the way in which such activity mediates between sensory data, the animal's intention, and the animal's movement. The attention paid by neuroscience experimentalists to computational models is increasing, as modeling occurs at many levels, such as (i) the systems analysis of circuits using the NSL Neural Simulation Language developed at USC; (ii) the use of the GENESIS language developed at Caltech and the NEURON language from the University of North Carolina and Yale to relate the detailed morphology of single cells to their response to patterns of input stimulation; and (iii) the EONS library of "Essential Objects of Nervous Systems" developed at USC to model activity in individual synapses in great detail. The next two chapters introduce the USC contributions, "NSL Neural Simulation Language" and "EONS: A Multi-Level Modeling System and Its Applications." Since the neuroinformatics of human brain imaging is so well covered by many research groups with and without HBP funding, this has not been a focus of USCBP research. However, we have been concerned with the following question: "How can the data from animal neurophysiology be integrated with data from human imaging studies?" The chapter "Brain Imaging and Synthetic PET" presents our answer.

**Part 3. Databases For Neuroscience Time Series:** The first chapter provides our general view of how to build "Repositories for the Storage of Experimental Neuroscience Data." We see the key to be the notion of the *experimental protocol* which defines a class of experiments by specifying a set of experimental manipulations and observations. When linking empirical data to models, we translate such a protocol into a simulation interface for "stimulating" a simulation of the empirical system under analysis and displaying the results in a form which eases comparison with the results of biological experiments conducted using the given protocol. The chapter "Design Concepts for NeuroCore and NeuroScience Databases" introduces NeuroCore, a novel extendible object-relational database schema implemented in Informix. The schema (structure of data tables, etc.) for each NeuroCore database is an extension of our core database schema which is readily adaptable to meet the needs of a wide variety of neuroscience databases. In particular, we have constructed a new Datablade which allows neurophysiological data to be stored and manipulated readily in the database. (See the appendix "NeuroCore TimeSeries Datablade.") The final chapter of Part 3, "User Interaction with NeuroCore," describes the various components we have developed of an *on-line notebook* that provides a laboratory independent "standard" for viewing, storing, and retrieving data across the Internet. We also present our view that the article will continue to be a basic unit of scientific communication, but envision ways in which articles can be enriched by manifold links to the federated databases of neuroscience.

**Part 4. Atlas-Based Databases:** How are data from diverse experiments on the brains of a given species to be integrated? Our answer is to register the data—whether the locations of cells recorded neurophysiologically, the tract tracings of an anatomical experiment, or the receptor densities revealed on a slice of brain in a neurochemical study—against a standard brain atlas for the given species. The chapter "Interactive Brain Maps and Atlases" provides a general view of such atlases, while "Perspective: Geographical Information Systems" notes the similarities and differences between maps of Earth and brain. The key chapter of Part 4 is "The Neuroanatomical Rat Brain Viewer (NeuARt)" The chapter "Neuro Slicer: A Tool for Registering 2–D Slice Data to 3–D Surface Atlases" addresses the problem of registering data against an atlas when the plane of section for the data is different from that of a plate in the atlas. The key is to reconstitute a 3–D atlas from a set of 2–D plates, and then reslice this representation to find a plane of section against which the empirical data can be registered with minimal distortion. Part 4 closes with the presentation of "An Atlas-Based Database of Neurochemical Data."

**Part 5. Data Management:** "Federating Neuroscience Databases" addresses the important issue that there will not be a single monolithic database which will store all neuroscience data. Rather, there will be a federation of databases throughout the neuroscience community. Each database has its own "ontology," the set of objects which create the "universe of discourse" for the database. However, different databases may use different ontologies to describe related material, and the chapter on "Dynamic Classification Ontologies" discusses strategies for dynamically linking the ontologies of the databases of a database federation. We then present "Annotator: Annotation Technology for the WWW" as a means to expand scientific (and other) collaboration by constructing databases of annotations linked to documents on the Web, whether they be for personal use, or for the shared use of a community. Part 5 closes with "Management of Space in Hierarchical Storage Systems," an example of our database research addressing the issue of how to support a user community that needs timely access to increasingly massive datasets.

**Part 6. Summary Databases:** "Summary Databases and Model Repositories" describes the essential role of databases which summarize key hypotheses gleaned from a wide variety of empirical and modeling studies in attempting to maintain a coherent view of a nearly overwhelming body of data, and how such summary database may be linked with model repositories both to ground model assumptions and to test model predictions. "Brain Models on the Web and the Need for Summary Data" describes the construction of a database which not only provides access to a wide range of neural models but also supports links to empirical databases, and tools for model revision. "Knowledge Mechanics and the NeuroScholar Project: A New Approach to Neuroscientific Theory" offers both a general philosophy of the construction of summary databases, and a specific database for analyzing connections of the rat brain exemplifying this philosophy. Finally, "The NeuroHomology Database" presents a database design which supports the analysis of homologies between the brain regions of different species, returning us to the issue of how best to integrate the findings of animal studies into our increasing understanding of the human brain.

The majority of chapters end with a section on "Available Resources" which describes the availability of our software and databases as this book goes to press. Much of the material is available for downloading; in other cases the prototypes are not yet robust enough for export, but in many cases may nonetheless be viewed online through demonstrations. The USCBP Website may be found at http:\\www-hbp.usc.edu, and will be continually updated to give the reader expanding access to currently available materials.

**Michael A. Arbib**
Los Angeles, California
**Jeffrey S. Grethe**
Hanover, New Hampshire