CHAPTER 3.1

Repositories for the Storage of Experimental Neuroscience Data

Richard F. Thompson,[1] Jeffrey S. Grethe,[2] Theodore W. Berger,[3] Xiaping Xie[3]

[1]*Biological Sciences/LAS, University of Southern California, Los Angeles, California*
[2]*Center for Cognitive Neuroscience, Dartmouth College, Hanover, New Hampshire*
[3]*Biomedical Engineering Department, University of Southern California, Los Angeles, California*

3.1.1 Introduction

The advancement of science depends heavily upon the ability of researchers to collect, store, analyze, and share their data efficiently. In order to help achieve these goals, informatics research has been playing an increasingly important role in various scientific communities. For example, molecular biologists have benefited greatly from the development of large-scale genome databases and new advanced genetic sequence search tools (Chen and Markowitz, 1994a,b; Frenkel, 1991). These new tools have allowed researchers, in this scientific community as well as other communities, to efficiently store, analyze, and share their sequence data. In the last decade attention has also focused on the development of informatics tools for the neuroscience community.

One of the problems encountered by many neuroscientists is the coherent storage, organization, and retrieval of extremely diverse and large datasets. Current technology allows neuroscientists to collect massive amounts of data. In order to help the neuroscientist manage these complex datasets, we have developed a novel database schema and implementation that allow datasets from diverse labs to be stored coherently in a single framework. The formal structure of that framework, called NeuroCore, will be detailed in Chapter 3.2, while various applications devised to help users interact with NeuroCore databases will be described in Chapter 3.3. In this chapter, we present the key ideas behind the development of NeuroCore and illustrate its application in the design of two databases, one for "Cerebellum and Classical Conditioning" (*in vivo* studies, Fig.1), and the other for "Long-Term Potentiation (LTP) in the Hippocampal Slice" (*in vitro* studies, Fig.2).

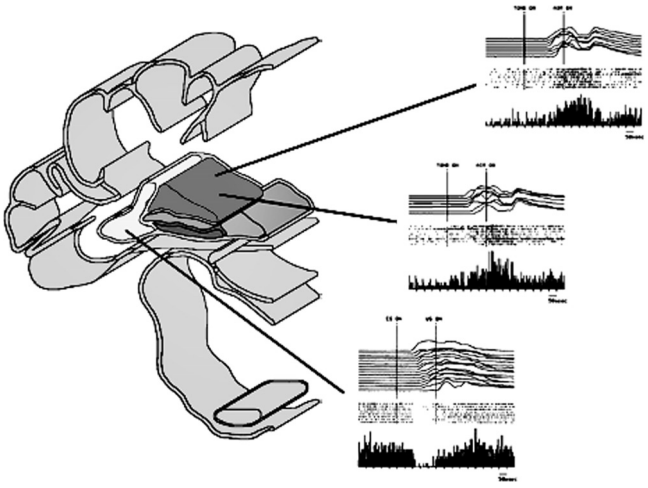

Figure 1 Unit recordings from a behavioral experiment: classical conditioning of the eye-blink response in the rabbit. In this experiment recordings are made from cerebellar cortex and nuclei and other structures while the animal performs certain behaviors as various external stimuli are presented. The data consist of unit recordings from various brain regions as well as recordings of the subject's behavioral response.

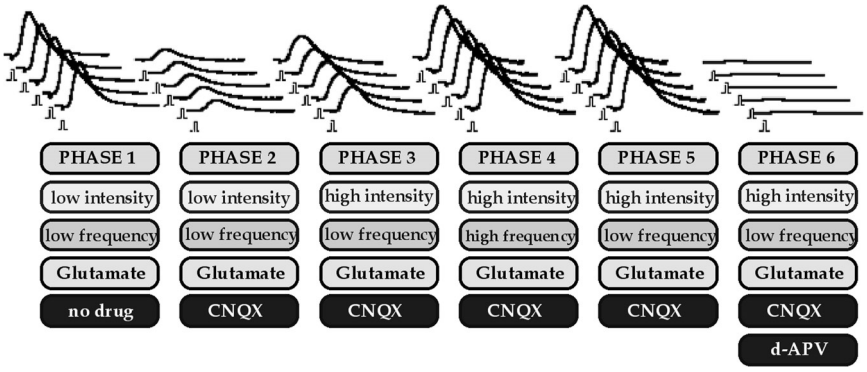

Figure 2 Recordings from a hippocampal slice. In this experiment various chemical manipulations and various types of electrical stimulation are applied to a brain slice at different instances. The data consist of field potential recordings from various regions of the hippocampus.

During experiments on learning (note that the current considerations are not restricted to learning studies), neural activity can be monitored, recorded, and then later analyzed to examine the contribution of varying brain structures to performance or learning in the intact animal or some more abstract markers of neural activity in a slice preparation. For each research subject used in an experiment (whether intact animal or neural slice), data can be collected from dozens of single neurons or even synapses. The data collected from each of these neurons or synapses typically consists of data from multiple individual training trials.

Over the years, hundreds of thousands of cells have been recorded from across a variety of paradigms, subjects, and conditions. Some interesting issues that point to the need for an experimental database are as follows:

1. How does one efficiently store data from neurons which have been recorded from multiple research subjects under a variety of conditions?
2. How can one analyze datasets of neurons that have been collected across subjects and even from different experiments?
3. How can researchers re-analyze large sets of data quickly and efficiently?
4. How can future researchers effectively use all the data from a variety of experiments even if they use different paradigms from current studies?
5. How can a group of researchers effectively collaborate on a research project using databases and Internet technology?

The goal in building a neuroscience database is to aid the researcher in the storage, retrieval, and analysis of electrophysiological and behavioral data, as well as anatomical and neurochemical data, from the critical brain regions involved in some behavior, memory phenomenon, or other function of interest. This database should also be a tool that is used by the researcher at all stages of the scientific process, from the development of a new experiment through the collection and analysis of the data generated by this experiment.

3.1.2 Protocols: A Data Model To Address Schema Complexity in Neuroscience

Probably the most important consideration in the design of any neuroscientific database is the complexity and richness of the data that need to be stored. Neuroscientists collect data comprised of many types, from the storage of simple tabular data in both numeric and textual form to more complex data types such as images, videos, and time-series data. One of the problems with many of these types of data is that there is no standard internal structure for representing them with most common commercial database systems. For example, the standard neurophysiological recording (time-series data) that consists of hundreds to thousands of individual measurements taken at regular or irregular time intervals cannot be easily or efficiently stored in standard relational database systems. However, a more daunting problem is how to represent and store the immense variety of experimental preparations that generate these types of data.

For example, we might want to store data in our database that are collected from unit recordings during a behavioral experiment as well as field potential recordings done in the hippocampal slice (Figs. 1 and 2). These data are actually quite similar in that they consist mainly of time-series recordings, even though the recordings will quite likely have been made by different recording electrodes which have their own unique properties. However, the methods used to generate the data in these two paradigms are quite different.

Our solution to coherently storing these datasets is to "associate" the experimental data with the information regarding the experimental protocol that was used in collecting the data. This protocol combines information on the preparation used, the types of experimental manipulations, and stimuli applied, as well as the kinds of observations and measurements made during the experiments. More specifically, the protocol provides a generic, parametric description of a class of experiments, while a specific experiment must be described by "filling

in the parameters" to show, for example, how the preparation was stimulated and what observations were recorded as a result.

In essence, data without the definition of the experimental protocol used to generate them are useless. One of the problems in specifying and storing protocols and data is their heterogeneity. Each researcher and/or laboratory works with protocols that might be specific to their research focus. The data they collect and analyze will also contain attributes that are important to their specific research goals.

This is quite clear by examining the very simple case outlined above (Figs. 1 and 2), where one wishes to store both *in vivo* and *in vitro* time-series recordings. Both of these situations require that time-series data be stored; however, each of these data records must contain additional information. For single-unit recordings that are made in a behaving animal, one must record the precise three3-dimensional coordinate of the location of the recording electrode so that one can later reconstruct (through the use of marking lesions and general histology) the exact anatomical region in which the electrode was located when the data were being recorded. When recording from a hippocampal slice, one does not have a similar coordinate system, but one is able to directly note exactly in what region and even in what cell layer the recordings are being made. Therefore, in designing the database architecture, a very important principle has to be taken into account: One cannot hope to describe *a priori* all the experimental protocols and research data that neuroscientists will want to incorporate in their own databases. It would be foolish to believe that one could. Due to this constraint, any database designed for the neuroscience community needs to be easily modified to meet a specific researcher's needs without requiring major modifications in the database's overall structure.

An Extendible Database

As we shall show in detail in Chapter 3.2, the key idea of NeuroCore is to provide a core database schema (i.e., a set of tables to be used in the database) comprising tables whose entries will be needed in almost all neuroscience experiments, together with machinery that makes it easy to add extensions to the core that specify data entries specific to the needs of a particular laboratory or protocol (as the latter may be shared by several laboratories). The key notion in being able to relate extensions to a core database structure is that of inheritance: Object-relational databases (see Chapter 1.2) allow types and tables to be placed in inheritance hierarchies. For example, in Fig. 3, a hierarchy of people is pictured, where PERSON is the top-level "parent" table containing the two columns *Name* and *SSN* (Social Security Number). In inheritance hierarchies, the children of a parent table inherit all the columns, primary key relations, and rules of the parent. In this case, both FACULTY and STUDENT inherit the *Name* and *SSN* columns from PERSON. However, both tables define new columns that are to contain data specific to their table (e.g., STUDENT defines the column *Major*). When selecting data from a parent table, the select function will also return data from the children. All information concerning the relations that exist in the inheritance hierarchy is stored in the system tables of the database. This allows easy reconstruction of the inheritance hierarchies through the system metadata, thereby allowing users to easily view the extensions added to the core database.

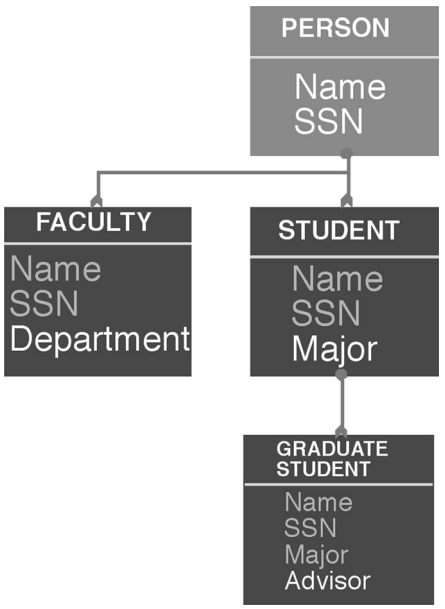

Figure 3 Example of an inheritance hierarchy. In this example, the PERSON table is a parent table. Both FACULTY and STUDENT are children of the PERSON table and inherit all the columns and constraints from the PERSON table. The STUDENT and FACULTY tables also define extra columns specific to themselves.

3.1.3 Considerations of the User Community

One of the major considerations that needs to be taken into account when designing any database system is the user community that will be interacting with the system. This is especially true of the neuroscience community where researchers from diverse fields and backgrounds (e.g., biology, psychology, engineering, and computer science) need to store, analyze, and share their data. By examining the neuroscience user community, the scope of a database for neuroscience experiments falls into one of three categories:

1. A database for a single experimenter or laboratory
2. A database for a collaboratory or research community
3. A publicly accessible database

Each of these database's users has different needs. The single experimenter needs to be able to securely store his work in progress as well as have sophisticated analysis and statistical tools to examine his data. A group of researchers collaborating on a project needs these tools as well as methods for the group to securely share the data being collected and analyzed at remote sites. Once the data collection and analysis is complete the data is then published and made publicly accessible to other researchers. The researchers who access this public data can be categorized as follows:

1. A researcher working on the same or a similar problem
2. A researcher interested in the work
3. A user who might just be browsing

For the researcher working in the same field, tools need to be available to allow researchers to examine and re-analyze the data using their own statistical methods and techniques. For both researchers, complete reference to the protocols and methods used to collect and analyze the data must be available. For all users, interfaces must be available that seamlessly allow users to gain access to the data stored in these databases from the familiar environment of an on-line journal article (for more discussion of this topic please see Chapter 5.3). In order to develop a database for the neuroscientific community all aspects of this varied community need to be addressed. The development of NeuroCore began with the task to build a database for the storage of *in vivo* neurophysiological recordings (Grethe *et al.*, 1996; Arbib *et al.*, 1996).

3.1.4 Building a Time-Series Database for *In Vivo* Neurophysiology: Cerebellum and Classical Conditioning

For many years, psychologists and neurobiologists have been searching for the substrates underlying learning and memory. One paradigm that has been extremely useful in examining these substrates is that of classical conditioning. From his research on this form of learning, Thompson (1986, 1990; Thompson and Krupa, 1994; Thompson *et al.*, 1997) has proposed an essential brain circuit, based on the cerebellum, that is responsible for this form of associative learning. This circuit now allows one to examine in detail the processes involved during classical conditioning of the nictitating membrane response. We here describe the issues involved in constructing a NeuroCore database to meet the needs of the Thompson laboratory.

Current evidence argues very strongly that the essential memory trace for the classically conditioned response is formed and stored in the cerebellum (Clark and Lavond, 1993; Clark *et al.*, 1992; Krupa *et al.*, 1993). There are actually two locations in the cerebellum where these memory traces appear to be formed: the anterior interpositus nucleus and lobule HVI of the cerebellar cortex itself. In examining the role that these two structures play in the acquisition and performance of the conditioned response, researchers rely heavily on neurophysiological data from neurons within these two regions. This section will use an experiment performed by Joanne Tracy, as part of her Ph.D. dissertation (Tracy, 1995), as a case study to examine how the NeuroCore database can be extended for a specific experiment. The dataset from this experiment concerns single-unit recordings from neurons in the deep cerebellar nuclei in rabbits trained using the classical conditioning paradigm.

Experimental Preparation

When storing data from any experiment, the data must be embedded in a framework that is based on information from the protocol which generated the data. During a typical classical conditioning experiment, a puff of air (the unconditioned stimulus, US) is paired with some neutral stimulus such as a light or tone (the conditioned stimulus, CS) (Fig. 4). The US by itself is able to elicit an eyeblink (the unconditioned response, UR) from the beginning. Over training the animal is able

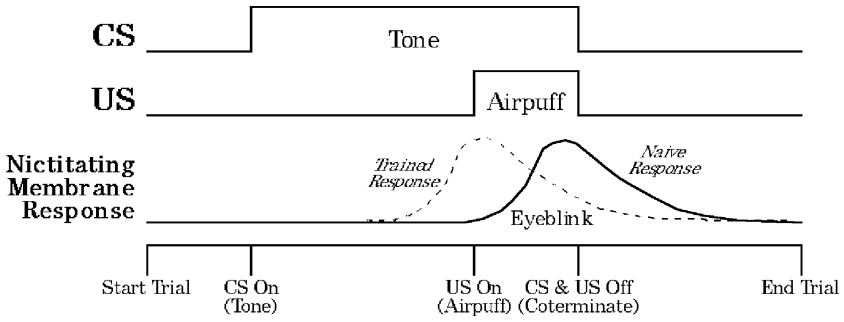

Figure 4 Depiction of a typical training trial where the subject receives two stimuli: a tone (CS) and an airpuff (US). Notice that the trained response (CR) predicts the onset of the airpuff (US), whereas the naïve response (UR) is just a reflexive response to the airpuff on the cornea.

Figure 5 Single-unit electrode marking lesions in the interpositus from a stained rabbit cerebellar brain section. The two marking lesions can be seen to the left of the arrow.

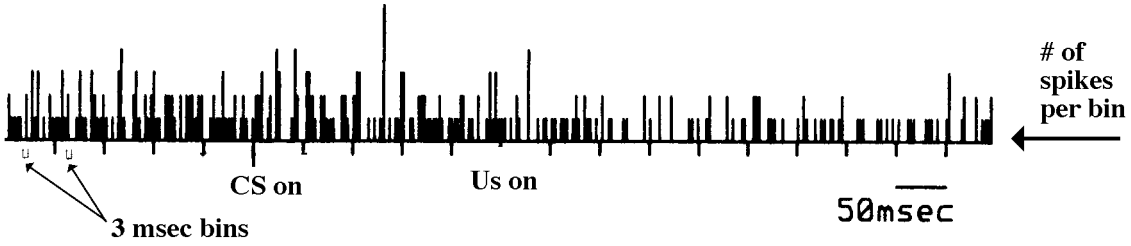

Figure 6 Sample of a single trial of single-unit data in binned format. Each bin records the number of times a neuron fired within that time period.

to associate the CS with the blinking of the eye and is able to blink in anticipation of the airpuff. This response is called the conditioned response (CR). Each training day consists of a number of these conditioning trials split into several training blocks.

When neuronal activity will be recorded, recording electrodes are lowered into the cerebellar cortex or into one of three deep cerebellar nuclei. These recordings are made under a number of conditions. Prior to removal of an electrode from a recording site, a marking lesion is made which is easily seen in a stained tissue section (Fig. 5). The actual recording sites are then determined post mortem via brain section and mapped onto a standard brain atlas diagram.

Data

A typical classical conditioning experiment generates a variety of data that must be stored. These data can consist of raw unit and behavioral data, data from post mortem histological sections, and statistical data generated from the raw data.

Electrophysiological and Behavioral Data

The electrophysiological and behavioral data collected during these experiments are typically digitized ("binned") data, including both single-unit data (record of activity from one neuron), multiple-unit activity (record of activity from many neurons near the recording electrode), and the movement of the nictitating membrane (Figs. 6 and 7).

Once the raw data have been gathered, researchers will want to analyze the data with various statistical methods. A summary over many cells and/or animals is then usually entered in a spreadsheet (Fig. 8). One of the goals, then, in developing this database is to give the researcher one location where all these data can be stored, indexed, and cross referenced; however, all these

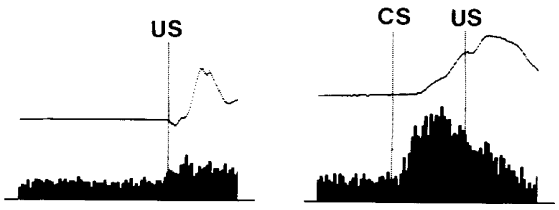

Figure 7 Sample of multiple-unit data from the interpositus. The top trace in each figure represents the behavior, whereas the bottom trace displays the unit activity from a group of neurons in the interpositus.

data must be stored in relation to from where these recordings where taken.

HISTOLOGICAL DATA

The most important histological data for an electrophysiological experiment is an accurate record of recording sites as determined from examination of electrode marking lesions (Fig. 5). Researchers also normally note the exact coordinate where a recorded cell is located. For example, Tracy (1995) stored all relevant electrode placement and cell information in a spreadsheet that was used for further analysis (Fig. 8).

Extensions to the NeuroCore Database

Each set of extensions to the base NeuroCore database described in the following sections is accompanied by a schema diagram which shows all the tables and relations involved in that particular aspect of the database. A more comprehensive discussion of NeuroCore itself can be found in Chapter 3.2. The legend for all these diagrams can be found in Fig. 9. In the figures that follow, the shaded boxes with a solid border will indicate super tables from the core schema that can be extended. The shaded boxes with the dashed border will show a variety of extension tables for the current paradigm, while other boxes will indicate various core and support tables within the NeuroCore database structure. It should be noted that a group of experimentalists may choose a set of extensions as defining an "extended core" for their investigations so that the extensions required for a specific protocol become very simple indeed. In order to develop the extensions necessary for NeuroCore to be

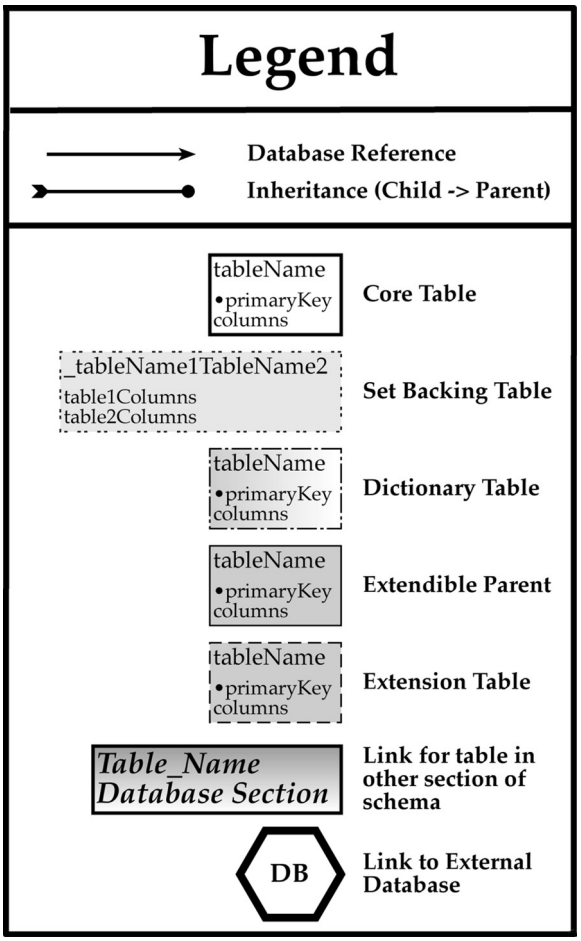

Figure 9 Legend for all schema diagrams. These diagrams represent the database schema by displaying the actual tables and their contents with the relations (primary key, foreign key, or inheritance) that exist between the tables. For a more detailed description of the table classes contained within NeuroCore see Chapter 3.2.

able to store the above described experiment, four distinct extensions need to be developed:

1. Definition of the research subject
2. Definitions of the protocol used in the experiment
3. Definitions of the research data being collected
4. Definitions of the statistical data being generated

The first step in creating the extensions to the NeuroCore database is to define the research subject that will

rabbit #	cell #	anterior	lateral	ventral	latency	duration	baseline freq.	response
91-289	1	1.0A	6.0L	19				
91-289	2	1.0A	6.0L	19.5	177	543	33.2	CR ++
91-323	3	1.0A	6.0L	19.7	69	636	30.4	CR ++
91-323	4	0.5	5	20.8	111	93	32.4	CR + UR +
91-323	5	0.5	5	20.6				UR+

Figure 8 Sample Summary Spreadsheet – Data from a few interpositus neurons. Each neuron is indentified by its ID and anatomical location. Various statistical measures are associated with each neuron.

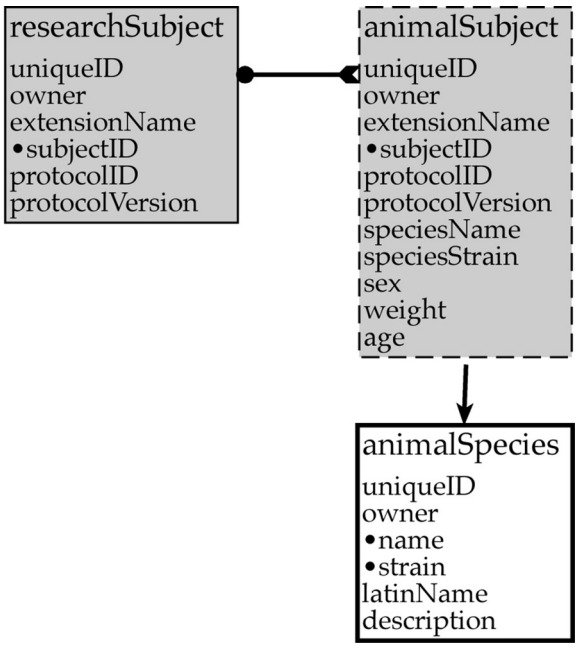

Figure 10 Research subject extensions. In order to store data concerning animal research subjects, an extension table (*animalSubject*) was added to the *researchSubject* table that allows the researcher to store the necessary information concerning the research subjects.

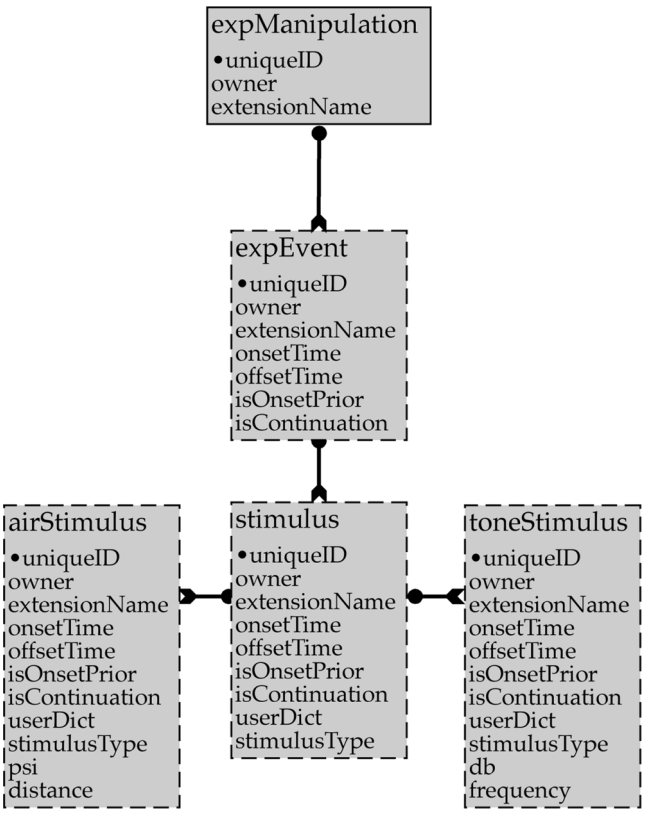

Figure 11 Experimental manipulation extensions. For the classical conditioning experiment, the tone and airpuff stimuli must be represented in the database as timed experimental events; however, each stimulus has its own parameters.

be used during the experiments. This extension (*animalSubject*) is created under the *researchSubject* table (Fig. 10). In the conditioning experiment we wish to store, the typical research subject is a New Zealand white albino rabbit. For our purposes, the only additional data necessary to be stored for each research subject are its species, sex, age, and weight.

To be able to store the experiment described above in a coherent fashion that will be easy to retrieve and analyze at a later date, the data collected must be "associated" with an experimental protocol. The protocol for the experiment contains a set of manipulations, the actual stimuli and conditions that comprise the specific experiment. These manipulations are created as children of the *expManipulation* table (Fig. 11). In defining the protocol for the standard classical conditioning paradigm, information regarding two distinct stimuli, the tone and airpuff, needs to be stored. In developing the extension for these two stimuli, it must first be noted that each of these stimuli contain explicit timing information (i.e., the onset and offset times) that is defined in the *expEvent* table. The *stimulus* table (a child of the *expEvent* table) is then able to define the general qualities of what constitutes a stimulus (e.g., what type of stimulus is it—a CS or US). The specific data for each stimulus are then stored in either the *airStimulus* or *toneStimulus* tables. This set of extensions illustrates a very nice property of inheritance: the ability to group tables into hierarchical collections where each child table further refines the information stored. This allows one to then query the database at intermediate levels of complexity. For example, one could search for all experiments that used two conditioned stimuli and one unconditioned stimulus, even though one does not know the exact nature of each stimulus.

Once we have defined the manipulations that were used to generate the data we need to describe the tables that will hold the actual research data (Fig. 12). These tables are created as children of the *researchData* table. For this experiment we need to store time-series data (defined by the *timeSeriesData* table) related to neuronal activity as well as the behavior being performed. For the unit data (defined by the *unitData* table), we need to record the actual anatomical location and coordinates where the electrode was located. However, when storing data from single cells (defined by the *singleUnitData* table), one also needs to store more information regarding the cell being recorded from (e.g., the ID of the cell and also what type of spikes were being recorded).

The last information that must be stored in the database is the statistical data generated from the analysis of the raw data discussed above. Various statistical methods are used to analyze single-unit data. The tables that define these statistics are created as children of the *expMetadata* table (Fig. 13). For the experiments to be stored in this database, two distinct classes of statistical data were generated. The first class of statistics analyzes

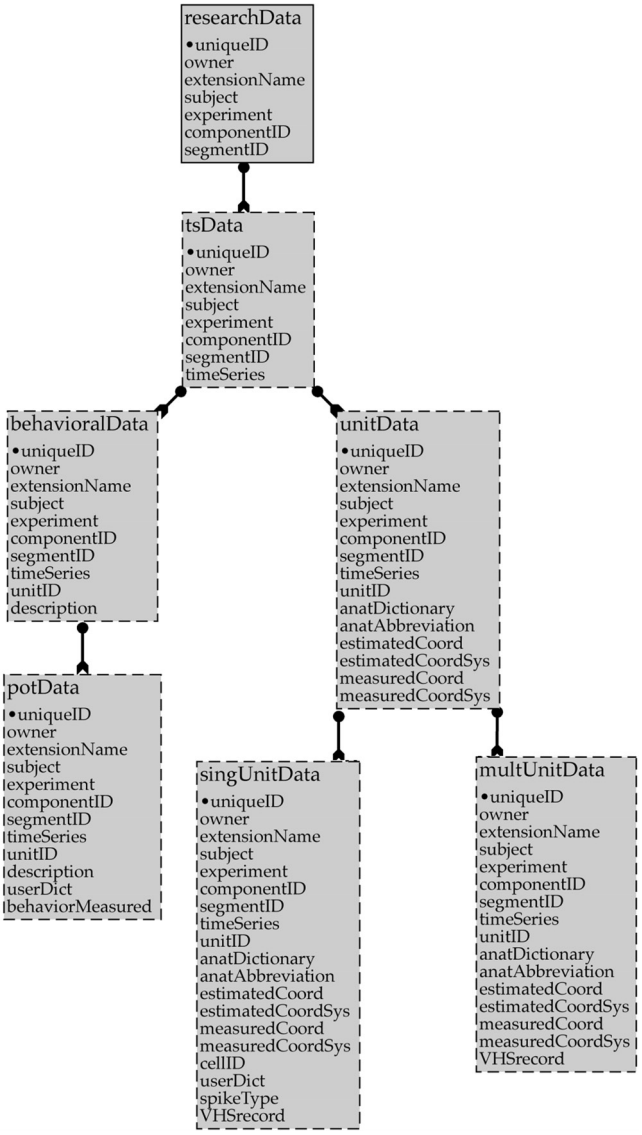

Figure 12 Research data extensions. For many neurophysiological experiments, researchers must store the neuronal unit data, single- as well as multiple-unit data, along with the behavioral data acquired from the research subject during the execution of the experimental task.

the general response properties of a neuron (defined in the *cellStatistics* table) and includes such measures as the baseline firing frequency and the significance level of the behavioral and unit response during various portions of the training trial. The second class of statistics (defined in the *cellResponses* table) analyzes the exact details of all the onset and offset responses produced by a given cell. It should be noted that all statistical data stored in the database are linked to the raw data through the NeuroCore core tables that were used to generate these statistics.

Data Input

Once the definition of the database is complete, one must begin to populate the database. Most of the data to

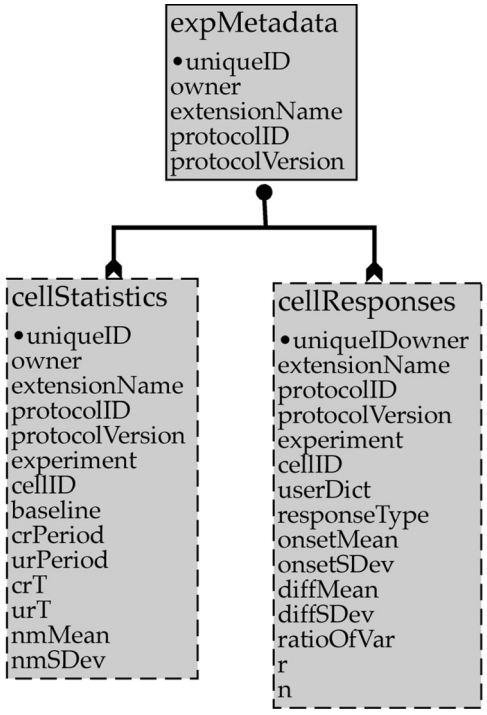

Figure 13 Metadata Extensions. For the experiment outlined above, two forms of statistical data were generated for each cell recorded. Cell statistics describe the general responsive properties of the cell to the task, whereas, cell responses characterizes the relation of the cell's responses to the behavioral response.

be entered in this database consist of legacy data; data already collected and stored as individual data files. In the case of the Thompson laboratory, much legacy data had been collected using routines written in the language "forth." In order to facilitate the entry of this legacy data into the database, a special conversion program was written that allows the conversion of forth datafiles (that contain the raw data) into SQL code (standard query language for databases; introduced in Chapter 1.2) that can be read by the database, so that the data contained in these files are stored in the database. This utility uses a datafile descriptor to read in the data from the forth datafile. All necessary variables and information regarding the experimental protocol are defined in the datafile descriptor. Using this information, the utility extracts the time-series data segments from the forth datafile which are then formatted into SQL statements that can be inserted into the database. The first step in the design of a user interface for data input was the implementation of a Web-based interface (Fig. 14) to aid the experimenter in constructing a descriptor file for a specific datafile.

After the legacy data have been stored, data for a new experiment can be entered using JADE (Java Applet for Data Entry; see Chapter 3.3). JADE is a Web-enabled interface that allows a researcher to enter experiment information (such as found in the datafile descriptor) directly into the database. After the experiment has

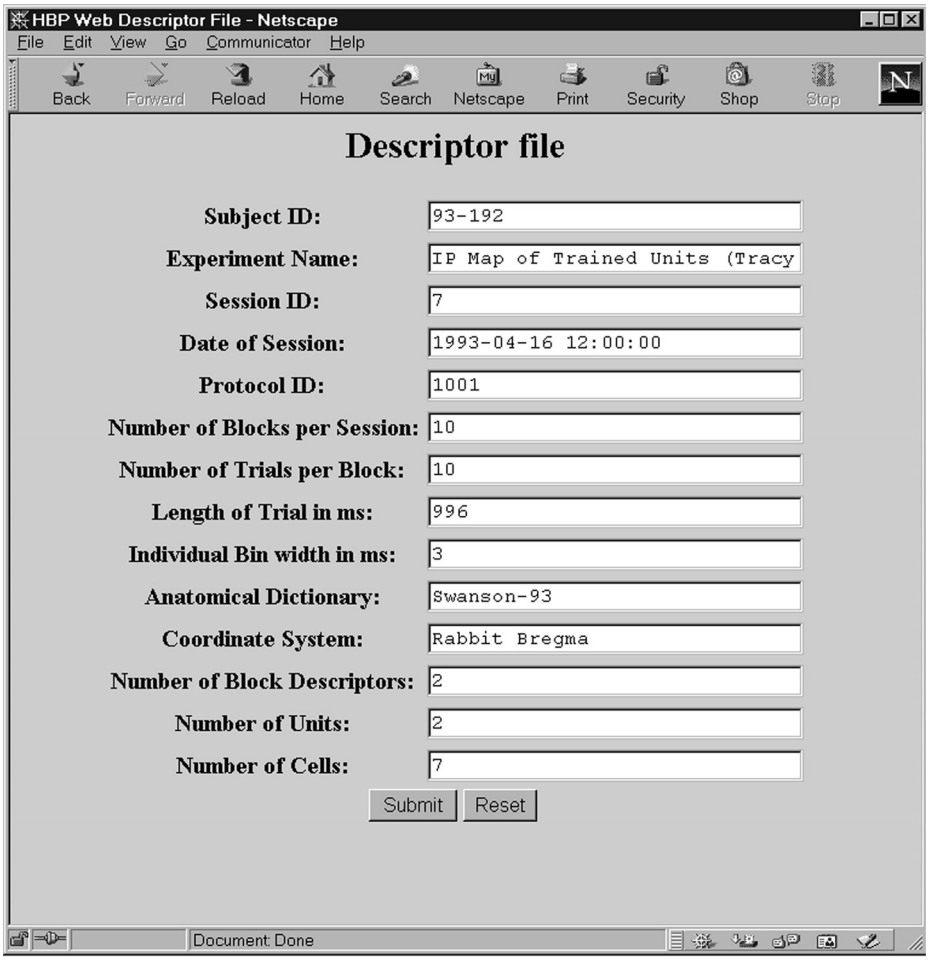

Figure 14 HTML-based interface used to generate descriptor files for data files that need to be entered into the database. The example above illustrates the top-level input required to format all subsequent data input forms. The descriptor file is used by a bulk loader to parse the data file and convert it into SQL statements that can be sent directly to the database.

been conducted, the datafile can be uploaded to the database server and the conversion program will enter the data into the database. JADE has the flexibility to adapt to any NeuroCore database associated with a laboratory. JADE prompts the user for experiment information based upon the tables in the database; therefore, new tables created to extend the needs of a laboratory can be recognized by JADE and used to enter experiment information without creating a new data entry tool. This exemplifies the benefits of having a standard core structure that is shared among various databases.

3.1.5 Building a Time-Series Database for *In Vitro* Neurophysiology: Long-Term Potentiation (LTP) in the Hippocampal Slice

In the current section we will use a series of experiments performed in Dr. T. W. Berger's lab as a case study to examine how the NeuroCore database can be extended for *in vitro* experiments. Brief trains of high-frequency stimulation to monosynaptic excitatory pathways in the hippocampus cause an abrupt and sustained increase in the efficiency of synaptic transmission. This effect, first described in detail in 1973 (Bliss and Lomo, 1973), is called long-term potentiation (LTP). LTP is a widely studied form of use-dependent synaptic plasticity expressed robustly by glutamatergic synapses of the hippocampus. The initial stage of LTP expression, typically identified as short-term potentiation (STP), is characterized by a rapid decay in the magnitude of potentiation to an asymptotic, steady-state level. Although there is a convergence of evidence concerning the cellular/molecular mechanisms mediating the induction of *N*-methyl-D-aspartate (NMDA) receptor-dependent STP and LTP, there remains substantial debate as to whether the expression of potentiation reflects change in presynaptic release mechanisms or postsynaptic receptor-channel function. Because of the synaptic co-existence of AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole proprionic acid) and NMDA glutamatergic receptor sub-

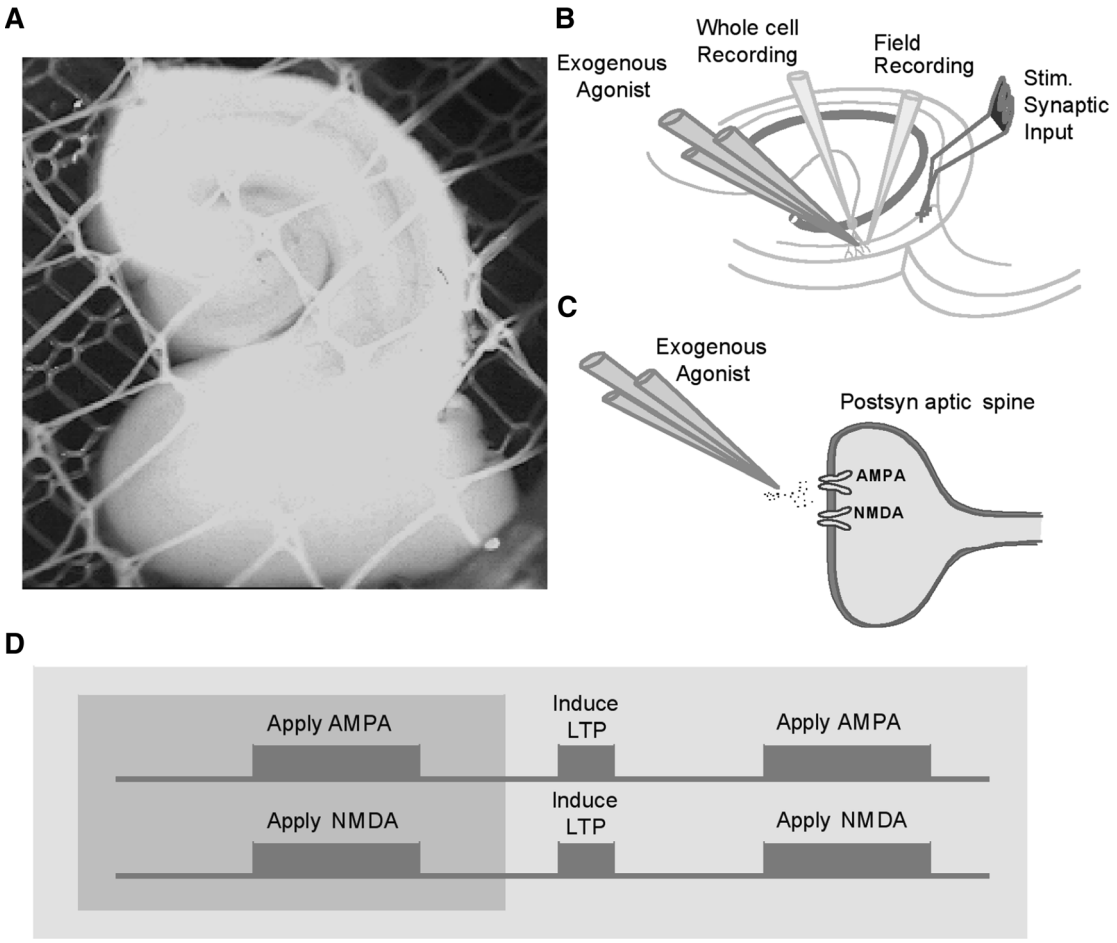

Figure 15 Demonstration of experimental paradigm. **(A)** Transverse slice of the hippocampus of a rabbit. **(B)** Placement of electrodes. **(C)** A multi-barrel pippette is used to focally apply agonists, such as AMPA and NMDA, to glutamate receptors on the postsynaptic spine. **(D)** One of the experimental protocols used in this study.

types, substantial differences in the magnitude of LTP expressed by AMPA and NMDA receptors would favor a mechanism that is postsynaptic in origin.

Experimental Preparation

Unlike the previous experiment, where neural activity was being analyzed with respect to the subject's behavioral performance, the current study aims to examine the detailed functioning of individual neurons. This is accomplished through the use of a slice preparation (Fig. 15), where a slice of tissue is extracted and all stimulation and recording is done directly on this tissue. This allows the experimenter much greater control of the conditions in the tissue slice (e.g., the balance of chemicals in the surrounding medium) as well as much finer control of the stimulation and recording electrodes used in the preparation. Fig. 15a demonstrates the placement of various electrodes in dentate gyrus of the hippocampal slice. In these experiments, we investigated the potential differential expression of STP by AMPA and NMDA receptors and found that the decay time course of STP is markedly different for AMPA and NMDA receptor-mediated excitatory postsynaptic potentials (EPSPs). Furthermore, during both STP and LTP, we found evidence of a differential responsivity of AMPA and NMDA receptors to focal application of their respective agonists. These results strongly support a postsynaptic expression mechanism of STP and LTP.

Data

Electrophysiological Data

A typical LTP experiment generates time-series data that consist of hundreds to thousands of individual EPSPs or excitatory postsynaptic currents (EPSCs) (Fig. 16a) taken at regular or irregular time intervals. The individual responses with duration of 200 msec, at a sampling rate of 10k Hz each, are then digitized and stored on the hard drive of a PC. Each of them also is parameterized and expressed in terms of its amplitude, onset slope, and area (Fig. 16b) in another data file. Data can also be gathered from cells after focal application of various agonists (Fig. 17). In this case, changes in

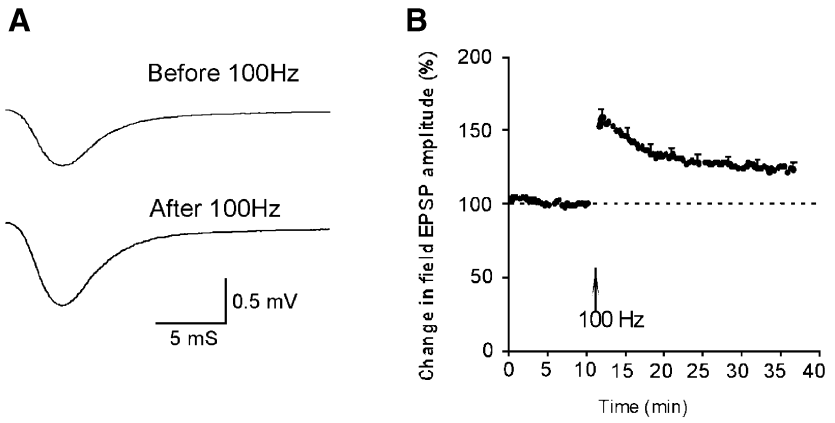

Figure 16 **(A)** Individual field EPSPs. **(B)** The amplitude of each EPSP was measured and represented as a point in this graph. A complete experiment lasts tens of minutes and consists of hundreds of individual EPSPs. Data from several experiments with the same protocol were grouped and statistically analyzed.

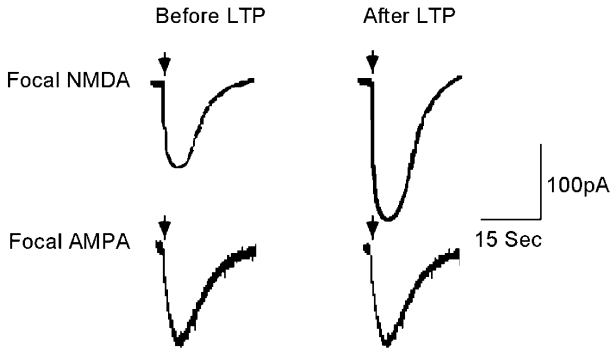

Figure 17 A representative graph showing the change in membrane conductance of a granule cell in response to focal application of glutamate receptor agonists, NMDA and AMPA. Membrane conductance in response to focal application of NMDA, but not AMPA, is increased during STP.

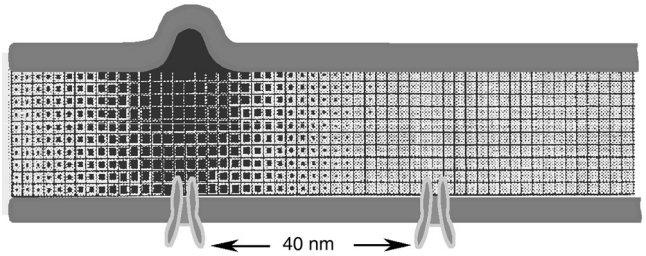

Figure 18 The effect of redistribution of receptors. Neurotransmitter release from a vesicle into the synaptic cleft. Because the cleft is extremely narrow, the concentration of glutamate (dark blue) at the postsynaptic membrane is highly localized.

membrane conductance of a granule cell in response to focal application of glutamate receptor agonists, NMDA and AMPA, were observed.

In this study, we also gathered continuous data containing information on the membrane conductance change during focal drug application. The duration of each of such record usually lasts for more than 20 seconds (Fig. 17). Raw data from this experiment are stored in a customized binary format. A conversion program has been developed to allow conversion of the experimental protocol and data to an SQL-ready format using a similar data input program described for the previous classical conditioning experiment.

Imaging or Graphic Data

More electrophysiological experiments are now seen combined with imaging techniques to reveal electrically triggered biochemical cascades, such as intracellular calcium redistribution, or morphological changes of the synapse following LTP induction. We routinely use graphic illustration (Fig. 15a,b) to record the accurate sites of recording and stimulating electrodes, drug applying pipettes, and surgical cuts in the brain slice. Working hypotheses underlying our specific research are often expressed graphically for clarity and simplicity (Fig. 18).

Extensions to the NeuroCore Database

In this section, we will discuss three distinct extensions that were developed for the storage of data from experiments performed on a hippocampal slice preparation:

1. The definition of the research subject
2. The definitions of the protocol used in the experiment
3. The definitions of the research data being collected

The first step in creating the extensions to the NeuroCore database is to define the research subject that will be used during the experiments. This extension (*animalSubject*) is created under the *researchSubject* table (Fig. 19). As was the case with the conditioning experiment, the typical research subject is a New Zealand white

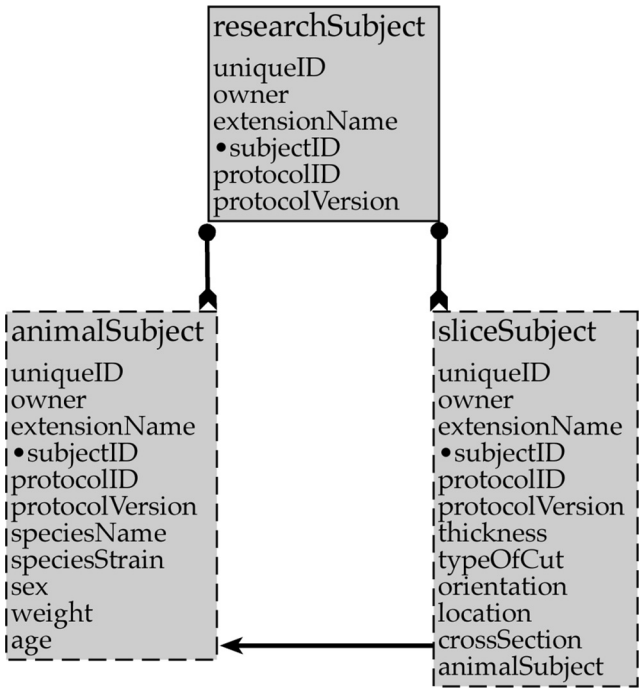

Figure 19 Database extensions for the research subject. In the case of a slice preparation, one must define a table to store information regarding the actual slice as well as the animal that was used to obtain the slice.

albino rabbit; however, in the LTP experiment, we are not dealing with a behaving animal, we are only dealing with a hippocampal slice. Hippocampal slices were prepared from male New Zealand white rabbits. Animals first were anesthetized with 5% halothane, and the skull overlying the parietal cortex was then removed bilaterally. The hippocampal formation and overlying neocortex from each hemisphere were extracted and gently separated. Both hippocampi were sectioned into blocks while being washed with cold, oxygenated medium, and slices of tissue (500 μm thick) then were cut perpendicular to the longitudinal axis using a vibratome. During the experiment, slices were superfused with medium consisting of (in m*M*): 126 NaCl; 5 KCl; 1.25 NaH_2PO_4; 26 $NaHCO_3$; 10 glucose; 2 $CaCl_2$; 0.1 or 1.0 $MgSO_4$, aerated with 95% O_2/5% CO_2 and maintained at 32°C. As one can see, the preparation and maintenance of a slice preparation is a complicated protocol by itself. The first step in being able to handle this data is to create the table that will store information regarding our various slices (*sliceSubject*). It is important to note that this table only contains the basic description of the slice as well as a reference to the animal subject that was used to obtain the slice. The critical information regarding the actual protocol used in the preparation of the slice is stored as a set of manipulations (defined as extensions of the *expManipulation* table) that belong to a specific protocol referenced by the *sliceSubject* table.

To be able to store the experiment described above in a coherent fashion that will be easy to retrieve and analyze at a later date, the data collected must be "associated" with an experimental protocol. The protocol for the experiment contains a set of manipulations (i.e., the actual stimuli) and conditions that comprise the specific experiment. These manipulations are created as children of the *expManipulation* table (Fig. 20). Unlike the previous classical conditioning experiment where two rather simple stimuli had to be represented, the current experiment requires a more complex representation of the stimuli, which in this case is either focal application of chemicals or electrical stimulation through an electrode. Focal application of NMDA (500 μ*M*) was obtained by microinjection into the middle 1/3 dendritic region of the granule cell (single pulse, 5 psi, 100 msec in duration, pH 7.3) via a multi-barrel pipette (Fig. 15a–c). Chemical stimulation is stored as part of the *chemicalFocal* table. This table defines the general parameters of the chemical application; however, the specifics of which chemicals were actually involved is referenced by *chemicalFocal's* parent table, *chemicalManip*. It is through this table that all chemical manipulations are related (through the *chemicalSet* table) to the specific chemicals (in the *chemicalDict*) and their concentration that comprise a particular chemical manipulation. The information regarding the chemical bath in which the slice remains during the entire experiment is also stored in a similar fashion through the *chemicalBath* table. Electrical stimulation to the perforant path input to dentate granule cells was activated using a bipolar nichrome stimulating electrode placed in the medial 1/3 of the *s. moleculare* to evoke field EPSPs. The information regarding the stimulation electrode and the pattern of electrical stimulation is stored as part of the *pulseElectrode* table.

Once we have defined the manipulations that were used to generate the data we need to describe the tables that will hold the actual research data (Fig. 21). These tables are created as children of the *researchData* table. As with the classical conditioning experiment, we need to store time-series data (defined by the *timeSeriesData* table) related to cellular activity in the hippocampal slice. In this experiment, data could be collected using either a whole cell electrode or an extracellular microelectrode. Whole cell electrodes (glass 7052 1.65-mm OD) were filled with (in *mM*): 120 cesium gluconate; 5 KCl; 2 $MgSO_4$; 10 N-2-hydroxy-ethyl-piperazine-N-2-ethanesulfonic acid (HEPES); 0.1 $CaCl_2$; 1.0 BAPTA; and 3.0 ATP-Mg (resistance: 6–9 MΩ;). Field EPSPs were recorded in the *s. moleculare* of the dentate gyrus using microelectrodes (glass thin-wall 1.5-mm OD) filled with 2.0 *M* NaCl (resistance: 1–2 MΩ). When field EPSPs and whole cell EPSCs were recorded simultaneously, the extracellular recording electrode was placed in the medial 1/3 of the *s. moleculare*, 200 μm from the whole cell pipette. Fig. 15a demonstrates the placement of various electrodes in dentate gyrus of the hippocampal slice. As with our experimental protocol, the exact specification of the internal chemical makeup of the electrodes is of

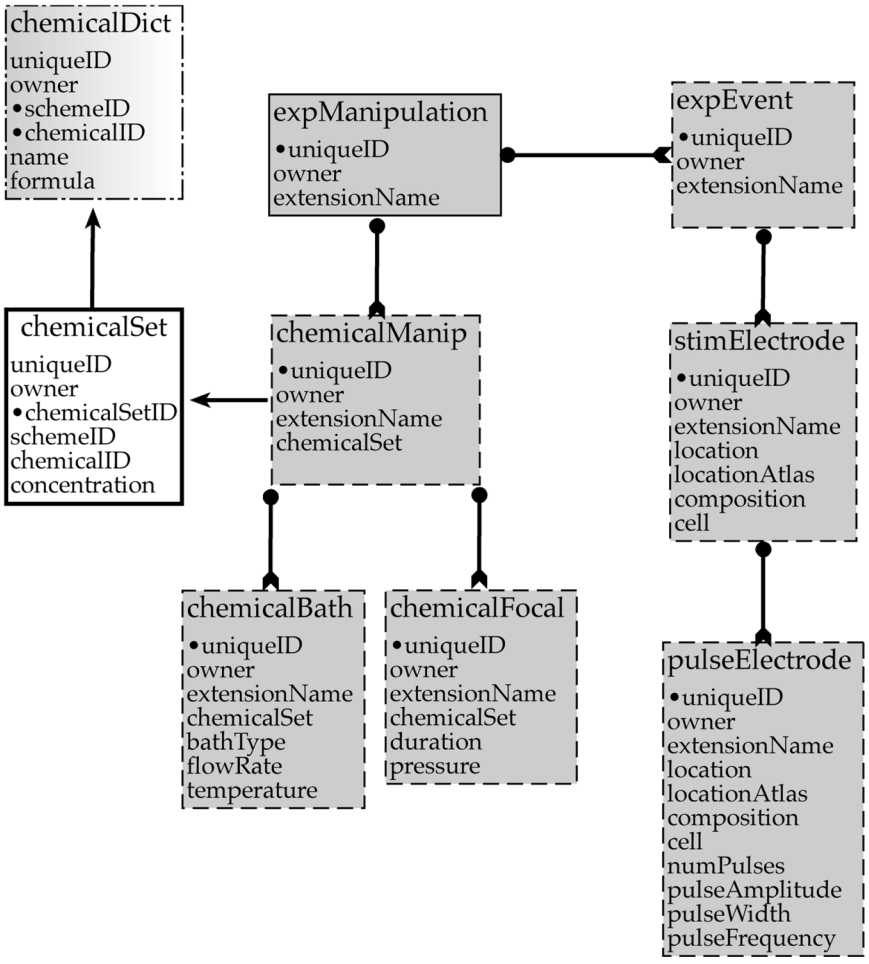

Figure 20 Database extensions for the experimental manipulations. Of great importance in the specification of any slice paradigm is the specification of the chemical manipulations that are an integral part of the protocol.

importance. Therefore, the *recordingElectrode* table used to store information regarding the various electrodes must also reference the *chemicalSet* table in order to define the chemical concentration within the electrode.

3.1.6 Building a Database for Human Neuroimaging Data

The previous two experiments have shown how NeuroCore can be extended to accommodate neurophysiological experiments. However, during the design of NeuroCore, it was extremely important to design a database framework in which a neuroscience researcher could store any type of neuroscience experiment. Recently, NeuroCore has been adopted as the foundation for an experimental database concerned with storing neuroimaging data (Grethe and Grafton, 1999). Neuroimaging has seen an explosive growth over the past decade with the proliferation of positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) paradigms.

Data

Neuroimaging data consist of three-dimensional images of the human brain. During an experiment, researchers collect both anatomical reference scans (Fig. 22) and a number of image series associated with the experimental task. In PET experiments, each trial is usually represented by a single image volume. However, in fMRI experiments, each trial consists of a large number of image volumes (Fig. 23) that are similar to an electrophysiological time-series except that the data being stored at each time point are now an image volume instead of a numerical value. Most neuroscientific databases being developed to store neuroimaging data tend to store these data as either an external file or an internal binary object. In both cases, the database is unable to manipulate the data by itself. They must first be exported to an application that can then process the data. For users to be able to automatically analyze and efficiently mine through large volumes of imaging data located at various data sites, the data must be in a format that the database can access and manipulate itself. Just

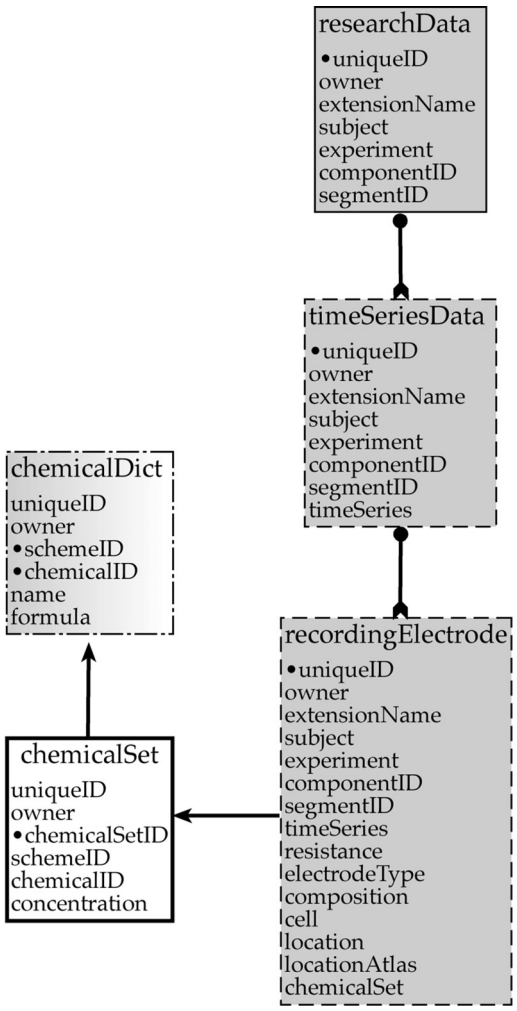

Figure 21 Database extensions for the research data. The recording electrodes used in the experiment must also reference their internal chemical composition in a similar fashion as the chemical manipulations discussed in Fig. 20.

as a TimeSeries Datablade (Appendix A2) has been developed for electrophysiological data, a new datatype is being developed to store neuroimaging time-series in an efficient format. In addition to the data associated with the large number of experiments being performed by researchers in this field, each experiment consists of a massive quantity of image data. For example, a typical imaging experiment might be comprised of thousands of image volumes from tens of subjects.

The analysis of neuroimaging data is a lengthy process involving many steps and intermediate results. In order to examine how a researcher arrived at his conclusions, the neuroimaging database must be able to store all versions of the data with information regarding the protocol or procedure that was used to transform the data at each step in the analysis process (Fig. 24). This allows for users to examine how archival imaging data was processed and analyzed to obtain certain results. Furthermore, the database structure allows users to reprocess and reanalyze datasets using new methods without alter

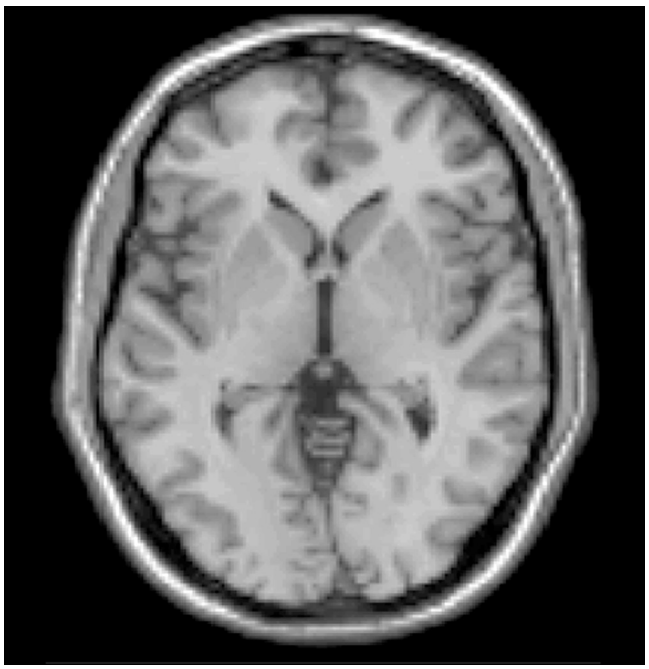

Figure 22 A single slice from an MRI anatomical volume.

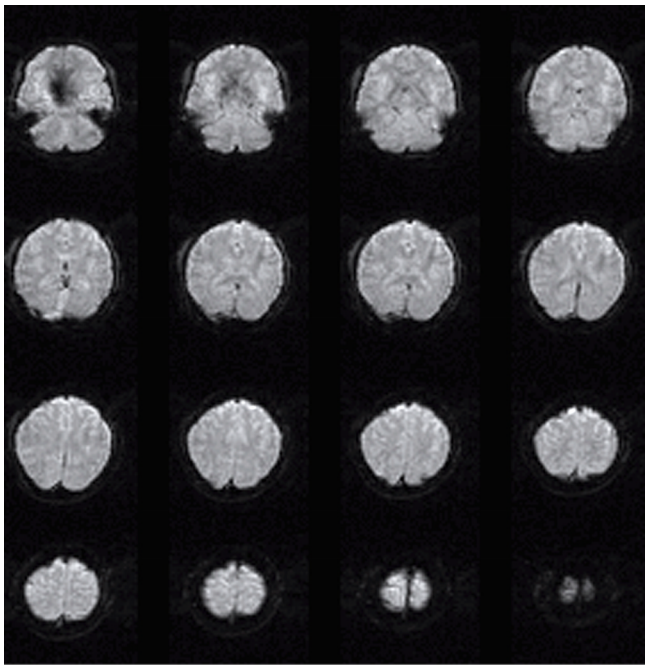

Figure 23 A single volume from a set of echo planar images (EPI) taken during an fMRI scan. This represents one time point in the functional scan.

ing past analyses. The neuroimaging database is currently under development and is an important new avenue in the development of the NeuroCore system as a whole. An interesting connection between the imaging data and physiological data can be found in the discussion of synthetic PET (see Chapter 2.4). Such technologies can hopefully provide a bridge between

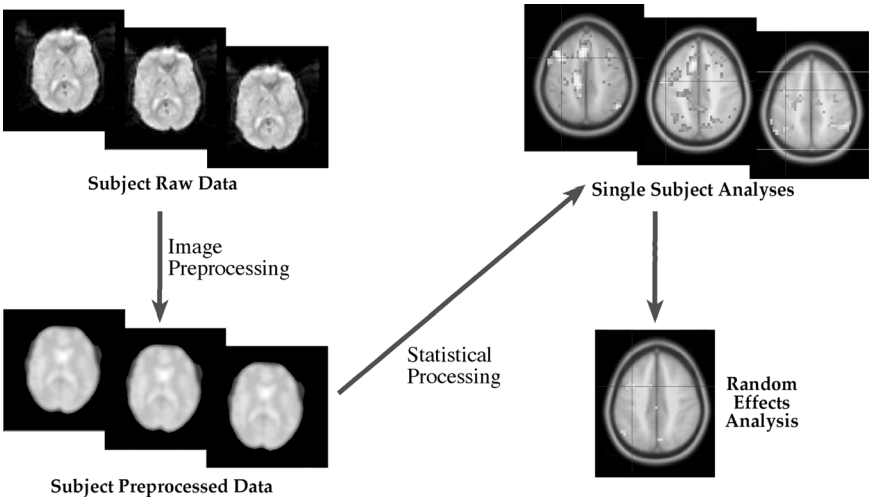

Figure 24 Statistical data flow in the neuroimaging database. All raw imaging data are stored in the database, so any analyses can be tracked back to the original data that generated them. One also has the option of storing any data generated during the processing and analyzing of the data. For example, when analyzing a set of fMRI data, one must first preprocess the data (alignment, normalization, smoothing). After the data have been preprocessed, one can apply various statistical analyses.

neurophysiological and imaging data through the use of computational models (see Chapter 2.1 for more discussion of computational models).

3.1.7 Discussion

The previous section introduced a database schema for the storage of neuroscientific data, which will be further described in Chapter 3.2, and specific extensions to this structure from various experiments. Currently only a few such database systems other than NeuroCore are under development (Payne *et al*., 1995; Gardner *et al*., 1997). The three major contributions from the development of NeuroCore are as follows:

1. *Tools for building extendible databases around a structured core*. This allows for the easy extension of the database by individual laboratories without the need to develop a complete database from scratch, as well as data sharing between NeuroCore databases as the core from each of the databases is identical. The common core also allows tools and extension sets (a collection of extensions developed for a specific experimental paradigm) to be easily shared between databases.

2. *Importance of the scientific protocol in storing experimental data*. Most neuroscientific databases only give a cursory description of the protocol for a given data record in the database. In order for data in a database to be useful to all users, they must contain all the specifics regarding the protocol used in the generation of the data.

3. *Complexity of neuroscientific protocols*. Most scientific databases developed currently view scientific protocols as being process-oriented (i.e., an input is processed by a manipulation that generates an output—in the molecular biology community, for example, an electrophoretic gel is labeled and generates a labeled separation; see Chen and Markowitz, 1995). Protocols in the neurosciences are much more complicated and consist of a hierarchy of manipulations that can be ordered, as well as specifically timed in relation to one another.

4. *Storage of complex data in a usable format*. Neuroscientific databases being developed to store neurophysiological time-series data tend to store these data as either an external file or an internal binary object. In both cases, the database is unable to manipulate the data; they must first be exported to an application that can then process the data. For users to be able to analyze and mine their data, the data must be in a format that the database can access.

Future Work

The major effort in developing NeuroCore to date has been in the construction of the core schema (Chapter 3.2) and the implementation of lab-specific extensions and interfaces that were discussed in this chapter. Work has also begun on general purpose user interfaces for the researcher to interact with the database (Chapter 3.3); however, for this database to become accepted by the neuroscience community as a whole, various developments must occur.

Handling of Heterogeneous Data

With the database having been tested on a few select lab datasets, the database and extension capabilities need to be tested on a larger scale. The Thompson database

currently consists of data from 312 cells from the paired training experiment in Tracy's thesis (Tracy, 1995). However, this database now has now been tested with data from other laboratories that use different experimental protocols and collect their data in varying ways and store them in differing formats. NeuroCore has also been used for the storage of neurochemical data (Chapter 4.5) and neuroimaging data (current chapter) and as a basis for the NeuARt system (Chapter 4.3)

Extend the Database To Handle New Datatypes

Neuroscientists collect various datasets that are highly structured and complex (e.g., time-series data, video, MRI and PET images). The current version of the database has been extended to handle time-series data through the development of a time-series datablade (a datablade is a new datatype and associated functions added to the database system; see Chapter 1.2 and Appendix A2). In order to be able to handle a larger variety of neuroscience data, new datablades and core support tables need to be developed to handle some of the other complex datatypes collected by neuroscience researchers (e.g., three-dimensional images obtained through MRI or PET). Such development is accompanying the work being done at Emory University and Dartmouth College on the neuroimaging database discussed earlier in this chapter. One could store these objects as simple binary large objects; however, this would not allow the objects to be manipulated within the database itself. Therefore, one could not query the data specifically but only the metadata associated with it. It is important to construct these various datablades so that users can interact directly with their datasets stored in the database.

Development of "Commercial Grade" User interfaces

Most of the user interfaces developed to date are lab specific and perform a specific task or query. For example, in order to load more of the raw data into the database, various tools were constructed: a bulk loader, which takes as input a definition file and a data file and produces the SQL code necessary to load that data into the database, and an HTML interface coupled to a cgi-bin application that produces the proper definition files for the bulk loader. This specific interface has been developed for our specific laboratory data and can be extended for similar types of data; however, it is not meant as a tool for the general neuroscience community. Work has begun on various general user interfaces (e.g., a protocol viewer/editor and a generic input applet), but these interfaces are in their very early stages of development. It is imperative that a consistent look and feel for all user interfaces are adopted so that users can work with all the tools available without having to learn a new interface for each tool.

Data Mining

There is no point in storing all these data in a database unless one begins to develop the statistical and analytical methods necessary to explore the data in a useful fashion. Just being able to retrieve the data is not extremely useful; however, being able to retrieve the data due to some statistical or analytical measure and then display the results is of great value. For example, many neuroscientists have developed their own techniques for analyzing single unit data (Berthier and Moore, 1986,1990; King and Tracy, 1998; Tracy, 1995). Being able to include these methods within a database would allow researchers to compare methods across the same dataset. One such tool in development at USC is DataMunch (see Chapter 3.3). The types of tools required for data mining depend on the scientific community that will be accessing the database. With the addition of significant amounts of raw unit data over the next few months, serious investigation of data-mining techniques related to electrophysiological unit recording can take place.

References

Arbib, M. A., Grethe, J. S., Wehrer, G. L., Mureika, J. R., Tracy, J., Xie, X., Thompson, R. F. and Berger, T. W. (1996). An on-line neurophysiological and behavioral database for the neuroscientist. *Soc. Neuroscie. Abstr*. **22**, 359:316.

Berthier, N. W., and Moore, J. W. (1990). Activity of deep cerebellar nuclear cells during classical conditioning of nictitating membrane extension in rabbit. *Exp. Brain Res*. **83**, 44–54.

Berthier, N. W., and Moore, J. W. (1986). Cerebellar Purkinje cell activity related to the classically conditioned nictitating membrane response. *Exp. Brain Res*. **63**, 341–350.

Bliss T. V., and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol. (London)* **232(2)**, 331–356.

Chen, I. A., and Markowitz, V. M. (1995). An overview of the object-protocol model (OPM) and the OPM data management tools. *Inf. Syst*. **20(5)**, 393–417.

Chen, I. A., and Markowitz, V. M. (1994a). Mapping Object-Protocol Schemas into Relational Database Schemas and Queries (OPM version 2.4). Lawrence Berkeley Laboratory Technical Report, LBL-33048.

Chen, I. A., and Markowitz, V. M. (1994b). The Object-Protocol Model (version 2.4). Lawrence Berkeley Laboratory Technical Report, LBL-32738.

Clark, R. E., and Lavond, D. G. (1993). Reversible lesions of the red nucleus during acquisition and retention of a classically conditioned behavior in rabbits. *Behav. Neurosci*. **107(2)**, 264–270.

Clark, R. E., Zhang, A. A. and Lavond, D. G. (1992). Reversible lesions of the cerebellar interpositus nucleus during acquisition and retention of a classically conditioned behavior. *Behav. Neurosci*. **106(6)**, 879–888.

Frenkel, K. A. (1991). The human genome project and informatics. *Communications of ACM*. **34**, 11.

Grethe, J. S., and Grafton, S. T. (1999). An on-line experimental database for the storage and retrieval of neuroimaging data. *Soc. Neurosci. Abstr*. **25**, 104.52.

Grethe, J. S., Wehrer, G. L., Thompson, R. F., Berger, T. W. and Arbib, M. A. (1996). An extendible object-relational database schema for neurophysiological and behavioral data. *Soc. Neurosci. Abstr*. **22**, 359.17.

King, D. A. T., and Tracy, J. (1998). DataMunch: a MATLAB-based open source code analysis tool for behavioral and spike-train data, http://www.novl.indiana.edu/~dmunch/.

Krupa, D. J., Thompson, J. K. and Thompson, R. F. (1993). Localization of a memory trace in the mammalian brain. *Science* **260(5110)**, 989–991.

Payne, J. R., Quinn, S. J., Wolske, M., Gabriel, M., Nelson, M. E. (1995). An information system for neuronal pattern analysis. *Soc. Neurosci. Abstr*. **21**, 376.4.

Thompson, R. F. (1990). Neural mechanisms of classical conditioning in mammals. *Phil. Trans. R. Soc. London B*. **329**, 161–170.

Thompson, R. F. (1986). The neurobiology of learning and memory. *Science* **233**, 941–947.

Thompson, R. F., and Krupa, D. J. (1994). Organization of memory traces in the mammalian brain. *Ann. Rev. Neurosci*. **17**, 519–549.

Thompson, R. F., Bao, S., Berg, M. S., Chen, L., Cipriano, B. D., Grethe, J. S., Kim, J. J., Thompson, J. K., Tracy, J. and Krupa, D. J. (1997). Associative learning. In *The Cerebellum and Cognition, International Review of Neurobiology*. Vol. 41 (Schmahmann, J., Ed.). Academic Press, New York, pp. 151–189.

Tracy, J. (1995). Brain and Behavior Correlates in Classical Conditioning of the Rabbit Eyeblink Response, Ph.D. thesis, University of Southern California, Los Angeles.