PROJECT DESCRIPTION

I. BACKGROUND
A. NEURONAL TIME SERIES ANALYSIS: POSSIBILITIES AND CHALLENGES

The analysis of neuronal firing patterns is responsible for a great deal of current knowledge concerning how nervous systems process information and organize behavior. This approach can be viewed as a brain "imaging" technology insofar as it reveals the dynamic changes of neuronal activity accompanying information-processing, storage, recall and behavioral performance. It is currently the only technology which measures neuronal activity directly, with spatial resolution in the micrometer range and temporal resolution in the millisecond range.

The emergence of neuronal time-series analysis (NTSA) as a cornerstone of neuroscience is based on the enormous success of the single-unit recording technique. Yet, the potential of neuronal recording continues to expand as new computer-based data acquisition systems and multi-channel recording technologies develop, enabling the simultaneous observation of spike trains of many neurons. Similarly, new approaches to neural modeling generate very large simulated spike train arrays that must be processed and compared to experimental data. These expansive data sets pose a formidable challenge to data management: how to efficiently store, retrieve, transfer and visualize the data.

Fortunately, the new computational technology that is responsible in part for the explosion of data is also the key to the problem of how to manage the data. The solution lies in the ongoing stream of improvements in processing speed, memory, information storage technology and new developments in visual data analysis. This proposal describes an effort to harness the emerging computational technology.

B. THE NTSA WORKBENCH: BASIC ATTRIBUTES

Overview. This project aims to develop a neuronal time series analysis (NTSA) workbench. The system being developed has four basic components: 1. a **database** for organized storage and retrieval of neuronal records meeting specific criteria of interest; 2. **high-resolution digital brain atlases**, to provide a neuroanatomical viewing context for data entry, query development, data analysis and visualization; 3. the **time series data protocol (TSDP),** a generalized standard for data representation and transfer that allows any time-series record to be processed regardless of how it was originally acquired and stored; 4. a **suite of powerful tools** for data analysis and visualization, as well as a modular programming library for new tool development and conversion of existing tools to support TSDP. The proposed system can serve as a "personal" database management system within individual laboratories, as a "workgroup" tool for exchanging data among groups of researchers spanning multiple laboratories, and as a means to implement a "public" database for wide access by neuroscientists via the Internet.

Coherence. No single data handling system can serve all of neuroscience. A principled restriction of scope is necessary to achieve conceptual and formal *coherence*. Formal coherence refers to the basic structure of the data. Neuronal data are time series data, i.e., sequences of numerical values indicating the presence or amount of neuronal activity and behavior across time. This project's focus on time series data confers formal coherence, which is essential if the range of tools for accessing and analyzing the data is to be practically limited. Conceptual coherence refers to the subject matter. This coherence is important if potential users are to have reasonable *a-priori* expectations about information obtainable from the database. The NTSA Workbench achieves conceptual coherence in that it is focussed on analyses of neural processes and systems in relation to the behavioral phenomena of sensorimotor integration, attention, learning and memory.

Generality. A major project goal is to develop a system that is generally applicable to a wide spectrum of community approaches and needs. Generality is being incorporated into all NTSA Workbench components. At the database level the system allows each laboratory to create a customized query interface using familiar, laboratory-specific terminology. At the level of data representation, the TSDP allows analysis of data files of any format by system tools. The digital atlas thrust integrates neuroanatomical information with metadata entry and retrieval, as well as data analysis and visualization functions. Digital atlas tools developed for the particular species studied by NPA scientists can be easily transferred to digital atlases for other species. Finally a suite of generally useful time series analysis tools as well as a dynamic modular programming library for creation of TSDP-compliant tools is being developed.

Researchers in the Beckman Institute NPA Group employ a variety of approaches ranging from *in-vitro* analyses of synaptic plasticity and network activity of small groups of neurons in culture to empirical studies and modeling of sensorimotor networks in acute and behaving animal preparations and *in-vivo* multi-channel analyses of neuronal activity during learning in behaving animals (see Section III, F). This diversity is quite representative of the broader neuroscience community. If the NTSA Workbench is constructed to meet the diverse needs of the

Beckman NPA Group it will be, de-facto, widely generalizable to the community at large. In addition, consultation and interaction with non-Beckman scientists is underway (see Appendix). These external collaborators will have access to the NTSA Workbench to allow them to evaluate and provide feedback on functionality and ease-of-use.

The NTSA Workbench provides a standard for exchange of data and tools that can serve the national and world-wide communities. Collaboration of scientists will be facilitated and new, powerful tools developed at particular sites will be immediately available for use at other sites. Investigators throughout the community will have access to all NTSA Workbench data sets, to test network models against empirical data and to exploit the power of data aggregation and *meta-analysis* for detection of patterns and relationships not found (or searched for) in the original analyses of smaller data sets.

C. DESCRIPTION OF NTSA WORKBENCH COMPONENTS
1. Design of the Database.
The core of the database system is the "card catalog" of the NTSA Workbench which describes the characteristics of the time series neuronal data that can be searched. This descriptive information is referred to as *metadata*, whereas the neuronal and behavioral time series data records themselves are referred to as *raw data*. The database *Table Schema* determines how the metadata are organized in the database. A hierarchical organization was adopted for the NTSA Workbench database Table Schema. This design provides a natural fit to the structure of neuroscience experiments, which can in most instances be described with reference to the following hierarchy: *laboratory, experiment, subject, session, series and trial*. Here, *laboratory* refers to a specific research group (e.g., the research of Co-PI David Clayton), *experiment* refers to a specific study in that laboratory (e.g., of songbird auditory thalamic neuron responses to natural and scrambled species-typical songs), *subject* refers to the experimental animal (e.g., zebra finch1255), *session* denotes the specific time and setting of neuronal data collection, *series* refers to sets of consecutive data collection episodes within a session having similar characteristics and *trial* refers to the individual data collection episodes, often defined by presentation of a single stimulus. Unlimited metadata descriptors can be used at each level to specify the experimental treatments.

A major problem confronting multi-user databases is that each user/laboratory will have unique laboratory-specific labels for the metadata. How can a given database incorporate a wide variety of labeling conventions? One solution is to require all users to conform to an arbitrary set of generic labels for the various dimensions of their metadata. A major disadvantage (among many) of this approach is that it requires all users to recode the standard laboratory terminology that they are used to, into arbitrary "databasese".This becomes particularly troublesome, for example, at the level of the experimental *session*, where lab-specific labels that are highly informative will have to be replaced by a non-informative label, and the mappings of the non-informative to the informative labels will have to be memorized or maintained and displayed separately. The approach adopted here avoids these difficulties. A mapping constructed for each laboratory associates the general database tables with laboratory-specific labels. Queries are constructed in the language of particular laboratories. The laboratory-specific names of metadata attributes are coded in the Graphical User Interface (GUI) used to generate queries to the database. Each scientist assigns labels to the GUI, thus designing an interface that is tailored to her/his specific application.

2. Digital Brain Atlases.
A critical requirement of many projects is the specification of the brain locations of recording electrodes, experimental brain lesions, assayed neurochemical, metabolic and MRI-imaged dynamic changes correlated with neuronal firing and other attributes of the brain areas in which data have been recorded. In order to meet this need, on-line 3-dimensional digital brain atlases are being developed for species employed by the NPA scientists. Sequential images of thin coronal brain sections are scanned at high resolution and stored. Researchers access brain sections in various planes to perform functions such as entry of metadata (e.g. the location of recording sites, the 3-D perimeter of a brain lesion, or metabolic activation), query formulation and visual data analysis in a neuroanatomical context. For *metadata entry*, researchers click on the atlas position corresponding to a recording site as determined by matching atlas sections with microscope slides or photographs. For designation of 3-D volumes, structures can be outlined in a series of 2-D planes to build a 3-D model of the volume. The user carrying out data entry need not specify the name of the brain region being clicked on or outlined. Association of the parts of the atlas with appropriate anatomical labels will be done only once, by an expert. For query construction, users can extract all instances of recordings made in particular brain areas encircled or highlighted on the atlas display. For visual data analysis, spatio-temporal patterns of brain activation based on experimental data, and patterns of related biochemical and structural change can be superimposed on relevant atlas representations or on images derived from the atlases. The patterns displayed can be extended to include animation and schematic diagrams of dynamic information flows in behavior-, and cognition-relevant neural circuitry.

3. The Time Series Data Protocol (TSDP).

The TSDP addresses the inevitable multiplicity of ways in which different researchers store their data. Typically, in order to submit data to a particular analysis or display tool, it is necessary first to translate it into a format acceptable to the tool. The challenge is to devise arrangements whereby diversely structured data files of different laboratories can be submitted to analysis and display tools without requiring users to reformat their raw data. A further problem stems from the use of different hardware platforms (e.g., PCs, MACs, UNIX workstations) for acquisition of neuronal time-series data. Each platform has idiosyncratic characteristics for binary data, including alternate word length, byte order, and so forth. Thus the acquisition of data on one platform and analysis of the data on another requires various manipulations, solely to account for the specific platform being used.

TSDP addresses these problems by providing a standard representation for time-series data, and transparent facilities for moving data between different analysis tools and platforms. Translation filters are created which accept raw time series data formats as input and produce TSDP data as output. Any analysis tool written using TSDP standard can operate on any time series data regardless of its original raw data representation. Using TSDP streams, analysis pipelines can be established with each tool in the chain performing its specific task, and outputting the resultant data to the next tool in the chain. By concatenating individual analysis tools, customized composite tools can be constructed.

It is important to note that TSDP is not "yet another file format". Rather, TSDP data representation will often be transient, existing only within analysis tools and along the pipelines between them. Researchers need not reformat their raw data files. As long as the form of the original data can be specified, translation to TSDP can be achieved. Once an appropriate translation filter has been developed, all existing and future TSDP-compliant tools will work with and "understand" the data, which can remain stored in its original format.

4. Analysis and Visualization Tools.

The *formal coherence* of neuronal data as time series renders them amenable to a common class of analytic approaches. A major goal of this project is to incorporate and support a full array of tools for *transformation, analysis, visualization and simulation* of time-series data. The tools will address both continuous time waveforms (e.g., field potentials) and point process signals (e.g., spike times and behavioral event markers). *Transformation functions* include classical frequency domain filters (highpass, lowpass, bandpass and notch) as well as Gaussian and lowpass filters for estimation of spike rates from point-process data. Other useful classes include matched filters and wavelet filters. A different class of filter performs pattern-classification, including spike sorting and tetrode analysis (see Wilson and McNaughton, 1994) which enables neuroscientists to extract single neuron spikes from multi-unit records. Other analysis tools include peri-stimulus histograms, auto-, and cross-correlograms, principal component analysis, power and cross-power spectra, spike-triggered and event-triggered average histograms, histograms with shuffled trial data (to normalize for within-trial dependence) and noise estimates.

High-performance computing affords technology for the creation of powerful tools for *display of data and visual data analysis*, enabling inspection of data from micro to macro resolution, and revelation of complex, dynamic relationships not apprehendable from simple plots and statistics. Examples include raster-like topographic plots of multichannel spike rates; exploratory functions such as coding of spike rate by color in different brain sites or a gray-scale (brightness) code of cross correlation among simultaneously recorded spike trains, techniques such as gravitational clustering (Gerstein et al., 1985) to facilitate recognition of activity patterns across many neurons; 3-D displays of brain network structure with superposed dynamically changing images of time series data.

Biologically-detailed neural simulations of the type supported by software packages such as GENESIS (developed in Jim Bower's laboratory at Cal Tech) and NEURON (developed by Michael Hines and John Moore at Duke) are typically used to generate time-series data of the same general form as data collected in neurophysiological experiments. The volume of time-series data produced in a typical simulation study is often comparable to, or in some cases much greater than, the amount of data yielded in physiological experiments. The uniform interface that NTSA Workbench presents to both experimental and simulated data greatly facilitates comparison of modeling results with experimental data.

This project's tool development effort is centered on *TSviewer*, a general purpose browser for multichannel time series data. TSviewer is the epicenter of a modular programming library which makes use of dynamically linked *plugins*. The goal is to provide a flexible working environment that neuroscientists throughout the world can use to develop custom time series display and analysis tools. Tools developed in the NTSA Workbench environment will support the TSDP convention and will thus be *interoperable* for many raw data formats.

HERE INSERT FIGURES 1 AND 2

5. NTSA Workbench on the Internet.

An important goal is to establish communication with other neuroscience database projects and the neuroscience community at large. To achieve this the NTSA Workbench database and tools are accessible on the Internet. Tools such as TSviewer, associated plugins, the digital atlases and the atlas viewer program are likely to be useful to community researchers and feedback from these researchers will engender important improvements. It is intended that this project's web site (**http://soma.npa.uiuc.edu/isnpa.html**) become an important focus of community-based development of neuronal time series analaysis tools that support TSDP and are thus interoperable in the community. The availability of the database and GUI will illustrate this project's concept of a generalized database schema and allow users to issue queries on selected NPA data sets.

II. RESULTS OF PRIOR NSF SUPPORT
A. AWARD: DBI-9504842, A database system for neuronal pattern analysis; $280K, 9/1/95-8/31/98
B. SUMMARY OF COMPLETED WORK:

Hardware and Software. Steve Quinn, the former head of the NCSA Database Tools Group, was hired as head Project Programmer and began project work in October of 1992. Consultation with Steve and other NCSA personnel led to adoption of the commercial program SYBASE as the relational database management system. SYBASE was installed in April of 1993. A Sun SPARC 10/SunOS 4.1 as the primary workstation, and four satellite stations (two Sun SPARC Station IPX/GX and two NCD X Terminals) were purchased with project funds and Beckman matching support. The satellites and hi-resolution monitors are located in Beckman Institute room 2313 (400 ft^2), designated as the NTSA Workbench area. The primary workstation with 224 MB RAM, 16 GB disk, a CD ROM drive and an Exabyte tape back-up system was originally placed in a nearby high-security room (2367). A Pinnacle Micro Demi Teton Magneto-Optical Juke Box system, judged to be a highly cost-effective solution to the large-scale project disk space needs, was funded by the Beckman Institute for $53K to match $28K of project funds, and was installed in October of 1994. The system can be expanded to 334 GB for the cost of media ($100 per 1.3 GB). The primary workstation was upgraded in 1997 to SYBASE SQL Server 11.0 on a 300-MHZ dual-processor UltraSparc2/Solaris 2.6, which was installed in the high security room. The former main workstation is now a dedicated project Internet web server and Jukebox data server.

Personnel. Head project programmer Steve Quinn left the project for a position as chief administrative Beckman Institute Programmer in April of 1996. Mr. Kittipong Mungnirun (vita attached) was hired as Steve's replacement, in June of 1996. Kittipong recently earned his Master's degree in the UIUC Computer Science Department, with a superior academic record and major academic honors. He brings to the project extensive experience with high-performance parallel computing, database programming and 3-D visualization. With initial guidance of Steve Quinn (who remains on-site) Kittipong came up to speed quickly and has since compiled an impressive record of accomplishment as indicated by the following description of progress. Kurt Gimbel, a recent graduate of the UIUC Computer Science program was hired in the Spring of 1997 and has done an excellent job of building the Digital Atlas Viewer Program Illustrated in Figure 1.

This project is relevant to the interests of the National Center for Supercomputing Applications (NCSA) in the management of scientific data and databases. Advice and consultation is available from resident NCSA personnel including Dr. Eric Jakobsson, who with Dr. Shankar Subramaniam developed the Biology Workbench, a highly successful Internet-based computing environment for molecular bioinformatics. Dr. Jakobsson has agreed to serve as a Co-PI to share his expertise and provide guidance to this project.

Improvements to the Database Interface. The database component of the NSTA Workbench was designed prior to this grant period (as detailed in Section I. C. 1). As mentioned, the database table schema was designed to allow many researchers to share a single relational database and GUI while retaining their distinct laboratory-specific terminology. The novel arrangement which permits this flexibility involved adding an extra level of indirection to the database table schema. First, generic tables were configured, one for each hierarchy level. In these tables, sets of data fields were created, one for each of four supported data types (integer, floating point, character string, and date/time). The fields were generically labeled according to their data type ("integer 1", "integer 2", "character 1", "character 2", etc.). A single additional table (the "mapping table") for each hierarchy level was then created to map the laboratory-specific labels to the generic field labels. For example, one laboratory might have a metadata field labeled "Rabbit ID" at the subject level which is mapped to the "integer 1" field, while another laboratory has a field called "Fish Number" which also gets mapped to "integer 1". Thus both share the same "integer 1" column in the generic table, but the column appears to carry their own laboratory-specific label.

A GUI was constructed using Motif library extensions of the X-Window System to implement user-friendly menus, buttons, text fields and dialog boxes. Researchers are presented with a different window for each level of the

experimental hierarchy, and each window presents labels and corresponding text fields for the metadata attributes at that level. Each attribute label has an associated pull-down menu. One menu option allows choice of the type of qualification (e.g. for numeric attributes, equal, not equal, less than, greater than). A dialog box can also be popped up to display information about the attribute, including data type and description. Another menu option provides "values lists" that indicate valid data values or ranges within which an attribute value must fall. The mouse can be used to select specific values from this list for formulating a query. All of these features provide on-line help for query formulation without requiring recall of detailed information about field formats and data types. On each window, menus are provided for access to other levels, so that researchers can formulate queries with qualification parameters spanning multiple levels. Each window contains buttons for query submission, clearing the screen and resetting the qualification types to their defaults. The list of matching information can be added to a "clipboard" of current results, and specific instances can be selected from the list for more detailed scrutiny.

In this grant period, tests of users' interaction with the database program revealed that some typical modes of querying can become difficult for the user to track. Queries primarily consist of forms filled out at any or all of the database levels, specifying which attributes are involved in a query and what their values ought to be. A typical query could be "find all files for which trial number is between 5 and 10, the sex is male at the subject level, and the stimulus is a 5 Hz sine wave". This query could be combined with another query: "...or find the same sort of trials but with a 10Hz stimulus". Queries like this, which combine groups of "and" and "or" criteria across multiple levels of the hierarchy, are difficult to formulate using a graphical interface. Three new approaches to this problem have been taken. The first involved enhancing the graphical interface for query formulation. While users are creating a query by opening various database level windows, a second window (which has been added) displays graphically the relationship of the various hierarchy levels that the query is addressing.   Second, an alternate means of issuing queries has been implemented, using an entirely different sort of interface, the powerful and readily-available PERL scripting language. Simple "batch mode" scripts can be written which describe a query algorithmically. In addition to allowing queries to be expressed that would be confusing in a graphical environment, this querying mode is well-suited to repeated or iterated querying, i.e., automatically interpreting the results of a query and responding with follow-up queries, perhaps many times in succession. The third new approach involves implementing query creation via simple (typed) English expressions (as in the quotation above).

Loading of raw data and metadata. Just as is data collection in the laboratories, raw data loading to the NTSA Workbench will be an ongoing process. Loading of existing raw data of the Gabriel and Nelson laboratories was largely completed during the initial grant year. In the second grant year, the database tables and GUI's for these laboratories were configured and initial steps were taken toward table and GUI configuration for the Anastasio, Feng, Greenough, Clayton and Malpeli laboratories. The basic steps are: a) implementation of the mappings which link laboratory-specific database tables to the general database tables of SYBASE; b) definition and selection of the metadata attributes for each laboratory; c) reading of the metadata into the database either manually through the database GUI or automatically.

Originally, filters for automatic extraction of the metadata into the database were static pieces of software which scanned data files for metadata and deposited the metadata into separate files, from which the metadata were loaded into the database. This method was easy to implement and helped to get things moving but proved cumbersome in requiring a new program for each new mapping from data to metadata. Currently, a graphical means of specifying the metadata to be read is being implemented which does not require programmer intervention for each new metadata attribute. Also, a cut-and-paste feature has made metadata entry more flexible by allowing portions of the database configuration to be moved from one part of the database hierarchy to another. These improvements now afford efficient metadata entering capability to the Gabriel and Nelson laboratories and they pave the way for the other NPA laboratories to use the database component of the NTSA Workbench (See "The Filter Guide", web site documentation of the new metadata extraction process).

Security. A contribution to streamlining of the database program and the GUI has been made in the form of multi-level security, which allows users to "lock" portions of the database that are considered complete in order to prevent accidental modification of metadata, and to enable laboratories to determine the level of access that different classes of users have to the data (e.g., browsing or active editing).

Digital Brain Atlases. This component of the NTSA Workbench is pivotal to several major NTSA Workbench functions (metadata entry, query specification, data retrieval and visualization of results in a neuroanatomical viewing context). The atlases are thus a central organizing principle of the NTSA Workbench. In this connection it was gratifying to note that SYBASE Open Server 11.1 has been endowed with a new functionality, a versatile programming interface allowing development of a specialized atlas database server, which can encapsulate the digital atlas information as a separate class of metadata. That is, the atlas representation can be contained as an

"object" in a separate database which can nevertheless be intimately federated with the neuronal database. With this arrangement, SYBASE confers all of the advantages of a mature and well-supported relational database program while also affording certain advantages of object-oriented databases.

Progress on the digital atlases in this project period included scanning and "cleaning" of high-contrast photographic images of 40-micrometer rabbit brain sections by undergraduate student William Holt. Similar work had already been accomplished for the electric fish. Janet Hanlon of the Beckman Visualization Laboratory, an expert in the processing of 3-D images of biological material, has concatenated, warped and smoothed the scanned sections to form realistic 3-D representations of these brains using *Analyze*. The completed rabbit atlas has 837 Coronal sections, 485 horizontal and 587 saggital sections, with each voxel representing 80 cubic micrometers. Beginnings have been made toward realization of the data entry, query and visualization functions as described in Section I, C, 2. Kurt Gimbel, a recent UIUC Computer Science graduate, has begun to work on a *Digital Atlas Viewer* which displays digitized images and allows selection and paging through 2-D section planes of the rabbit atlas (Figure 1). 2-D regions specified in successive images can be used to create 3-D regions as a series of 2-D layers. 3-D regions are built-up from a collection of horizontal voxel strips, each with dimensions 1x1xN. Currently the Atlas Viewer exists as a Java Applet and can be explored on the project web site. Ultimately, the Atlas Viewer program will develop into the multi functional tool for envisioned atlas-based data entry, retrieval and visualization functions.

The Time Series Data Protocol. From the outset, design of tools that do not force researchers to change the way they store their data has been a fundamental goal. TSDP has thus been developed as a generic convention for representing time series data in a common form, in order to eliminate the need to write a particular piece of software for each potential data format. TSDP is designed to allow the creation of streamlined and user-friendly tools that appear to be independent of the format and source of time-series data. As such, TSDP consists of a set of data structures *and* a library of code from which programmers can develop time-series analysis and visualization tools. Specification and implementation of these features of TSDP have progressed as described below.

TSDP data structures were developed to implement a self-descriptive data representation which can be passed from tool to tool much as electrical signals pass through a sequence of filters. Each tool adapts itself to the structure of the incoming data using the descriptive attributes of the data. The data structures allow TSDP to inform tools of simple things such as the number of data channels, or how to interpret event markers in the data. Four primary data structures have been defined: 1) a header record to carry information about the time-series data as a whole; 2) a descriptor record for carrying information about particular sections or regions of the time-series data stream; 3) a marker record for marking events within the time-series stream, and; 4) a data record for carrying the actual time-series data. The design of the TSDP data structures ensures that raw format data can be translated into TSDP without the loss of any descriptive or structural information.

Surrounding the TSDP data structures a layer of code has been implemented which arranges for the core TSDP data to be transported and accessed. Specifications established for the transport layer give TSDP its machine-independence. This aspect of TSDP accounts for hardware differences such as byte order and word alignment that exist between various systems. The specifications require that this machine independence be achieved in such a manner that neither the application programmer or the user is aware of the issue.

Finally an application programming interface (API) has been designed which encapsulates the transport layer of TSDP, and simplifies the programmer's interaction with TSDP data. Part of this layer of the TSDP interface library includes convenience routines for translating data into and out of TSDP format, data validation and error reporting mechanisms, and a sophisticated indexing system that allows segments of TSDP data to be accessed directly, without the need to search for the locations of regions of interest. Together these components constitute approximately 10000 lines of C source code, and have been tested on a variety of UNIX architectures to date.

An initial group of "filters" for converting raw format data into TSDP has been successfully implemented for the Gabriel lab T-file format and DataWave's common "UFF" format. Dense binary data such as UFF takes approximately 50% more storage space in its TSDP format, which results from the underlying platform-independent way in which TSDP stores information. TSDP data compresses efficiently however, since much of this extra space is "padding" --blank space that provides proper alignment of data. The basic indexing scheme, which provides automatic indexing of the location of each data record in the TSDP data "stream", and references to defined "regions" or epochs of data, constitutes up to an additional 25% of the total required storage space.

Analysis and Visualization Tools. A general-purpose software tool for time-series viewing and analysis (*TSviewer*) has been developed by Jeremy Payne. The program provides a Motif-based graphical interface for browsing and displaying multichannel neuronal time-series data, including continuous waveforms, spike events, and various markers related to stimuli and behavioral events. TSviewer allows the user to interactively select hierarchically-organized subregions of the data and to define new regions of interest for further analysis.

TSviewer is a modular program, which renders it readily extensible. In addition to basic and flexible visualization of multichannel time series data and selection of regions of interest, the TSviewer *plugin* modules support histogram construction, spike sorting, correlation analysis and data format translation. Plugins have become a common software concept of web-browser technology. These small external pieces of software add a specific function to the larger application, and allow the functionality of TSviewer to be extended to suit the needs of particular laboratories. In this design, the core TSviewer code is kept simple with well-defined state transitions, wherein plugin calls can be inserted as appropriate. Plugins can check for and call other plugins, allowing the creation of sophisticated TSDP-based data analysis pipelines.

TSviewer thus provides a framework that allows use of previously developed tools and quick implementation of data analysis routines. The plugin approach obviates the need to develop a custom user interface for each tool, which typically accounts for more than 90% of the programming effort in developing stand-alone applications. The key to making this work smoothly is a powerful application program interface (API) that allows plugins to exchange data with TSviewer using TSDP. Also, plugins manipulate TSviewer menus and add graphical elements to the screen, such as dialogs and forms. Currently the TSviewer core exists along with a basic set of analysis plugins (dot raster displays, rate histograms, peri-stimulus time histograms), a set of spike clustering tools (geometric, k-means, principal components), and a set of filters for importing and exporting time series data from several raw file formats. A sample display from the TSviewer program is shown in Figure 2.

Documentation: NTSA Workbench on the Internet. To foster communication with other neuroscience database projects, a considerable emphasis has been placed on Internet display of the concepts and products of this project. Readers are invited to visit the project home page (**http://soma.npa.uiuc.edu/isnpa.html**). A significant effort has been made to document the NTSA Workbench. Included are a project user's guide, a GUI manual and a technical white paper ready for submission (Payne, Mungnirin, Quinn, Nelson and Gabriel, 1998). Also, several project utilities and tools can be accessed and used on the web site. Recent additions to the site include: a) a working version of the database interface that illustrates the design of the Generalized Database Table Schema and the NPA Laboratory-Specific GUI, that visitors can use to inspect metadata of the Nelson and Gabriel laboratories and submit limited queries; b) a preliminary Java version of a Digital Atlas Viewer program allows selection of segments of rabbit and fish brains for rapid scanning of coronal sections; c) A Java version of TSviewer which allows display of time-series records with capability for zooming in on particular waveform segments, and links to raw data on which the display is based.

III. PROPOSED WORK
A. PROLOGUE

Past efforts of this project have given rise to a computational environment designed to harnesses state-of-the-art technology for processing of time series neuronal data and to promote interoperability of tools and sharing of data in the neuroscience community. Implementation of each of four key components (database, digital atlas, TSDP, analysis tools) is well-underway and enthusiastic responses have been received from researchers involved in similar efforts. The essential thrust of effort in the next grant period will be to complete implementation of these components and to work with on-, and off-site researchers to refine them and bring them to full usefulness.

B. DATABASE AND GRAPHICAL USER INTERFACE

An important goal in the next project period is to bring all participating NPA researchers to full utilization of the NTSA Workbench. This ample and productive group represents a rich array of approaches and needs and all participants have important research problems for which the NTSA Workbench provides good solutions (see Section F, below). Full involvement of the participants is an important key to the development of generalizable and exportable neuroscience tools. The Nelson and Gabriel Laboratories have been spearheading project development and are very nearly ready to use all NTSA Workbench components. Full utilization awaits storage of the anatomical data as metadata using the new Digital Atlas Viewer program, which allows efficient association of the literally thousands anatomical recording sites with the already stored time series data records. Bringing the other researchers to the desired level of involvement will require active project support. Entailed in this effort are: completion of transport of the raw laboratory data files to the data repository for some laboratories; preparation of extraction filters to facilitate coding of metadata; creation of mappings to customize the GUI for each laboratory, production of the TSDP extraction filters to allow laboratory time series data to be used with TSviewer and other NTSA Workbench tools; manufacture of digital atlases for species studied by the NPA researchers. Programming support will be provided to each NPA laboratory for the preparation of metadata extraction filters, GUI

customization and TSDP filters. The laboratories will develop their own digital atlases with full project guidance and existing interface software for use as a template.

Substantial effort will be devoted to establishing seamless links between the database query GUI, the digital atlas viewer interface and the time series analysis tools. Significant improvements will also be made to the existing forms-based GUI to provide users with more immediate feedback during the process of query construction, in order to improve search efficiency.

C. DIGITAL ATLAS DEVELOPMENT

Completion of scanning, concatenating and aligning of the high-resolution 3-D digital fish and rabbit brain atlas images in the present project period has set the stage for development of fully functional digital atlases for use both as stand-alone tools and by those who plan to integrate the digital atlases with database, analysis and visualization functions. This will be done by expanding the Digital Atlas Viewer program developed by Kurt Gimbel as described in Section II, B. 2. In its expanded form the program will associate a variety of data with atlas areas (voxels and regions). Association of the atlas areas with the raw neuronal records in the database will allow the anatomical metadata entry function. That is, a simple click on a particular atlas area matching a histological recording site will link that area with a particular subject (rabbit or fish), a particular recording electrode in that subject's brain and a larget set of time series records of that area. Various sets of stereotaxic coordinates and brain area names will also be linked with the atlas areas, thus allowing atlas-based queries yielding lists of data records obtained in particular areas or sets of areas. With this functionality, the precise anatomical origin of all neuronal records will be added as a queryable attribute in the database. Free-hand drawing and a polynomial spline interpolation method using mouse-specified control points will be used to "regionalize" the atlas and ultimately to label atlas regions with accepted anatomical nomenclatures. As an alternative to the predefined regions, provisions will be made to allow users to "draw" arbitrary regions at the time of query generation. This will be done either using the aforementioned methods or with "rubber banding", whereby the user clicks on a point to trace either a circle or rectangle as the mouse is moved. Interpolation between successively drawn user-defined regions will be implemented, to create a volume across depth. This interpolation will take into consideration the contours of the whole brain, as well as specific cytoarchitectonic regions, so that the volumes will "shrink/grow" and "bend" across depth. The volumes so defined will be usable as 3-D bounding regions for searches. The 3D objects will be rotatable and sectionable. Display of neuronal record dispersion and ultimately animated neuronal activation over time within 3-D regions will be implemented. Various logical qualifiers for generation of atlas-based queries will be provided, including AND, OR, NOT, XOR, etc. The OR function is particularly important as it will allow the user to specify any number of differently-shaped 2-D areas for queries. It should be noted that parallel versions of the digital atlases are being developed, one a full-scale version for extensive data entry, querying and work with 3-D images on local workstations and a reduced version tailored to provide maximum functionality consistent with the slower access times and processing speeds of the Internet.

As the basic Atlas Interface program is implemented it will be relatively easy to apply it to new digital atlases for any species, which can be made by scanning and concatenating the appropriate tissue sections. Also, the atlas interface is readily applicable for use with key pieces of brain, such as the monkey and human lateral geniculate nucleus (see Section F below on the Malpeli laboratory use of NTSA Workbench). The digital atlases of this project complement other efforts to incorporate neuronatomical representations with neuroscience databases. A case in point is the excellent rat brain atlas of Dr. Larry Swanson, of the USC Human Brain Project group. Cooperation with this group is ongoing (see Appendix) and PI Dr. Gabriel is a member of their External Advisory Board. Importation of the Swanson atlas to this project would provide many powerful functionalities for this widely used species. On the other hand, it is by no means unreasonable to undertake here construction of a digital rat atlas, for research that requires the ultra-high spatial resolution afforded by the present approach (e.g., for precise micrometer-range localization of single unit recording sites).

D. THE TIME-SERIES DATA PROTOCOL (TSDP)

While the core specification of TSDP has been completed, it remains to make this specification public and in so doing to work with other groups to ensure that TSDP can serve as a community standard for time series data. The development of the application program interface (API) library is currently underway. The API provides a framework for the development of analysis tools in the context of TSDP.   The existence of a programming interface dedicated to the creation of TSDP-based tools will greatly speed development time.

The development of the API will involve several phases. First, basic TSDP manipulation routines will be completed to enable establishment of connections between different TSDP-compliant tools for pipe-lining and

transfer of TSDP data back and forth along the connection, and to"unpack" TSDP time-series data, and parse and interpret TSDP information and event data. Next, steps will be taken to enhance TSDP, so as to permit choice of the optimal type of connection, depending on the situation. For example, connections between tools across a network might be established using Berkeley sockets or TLI, whereas connections between tools on the same host will be done with intra-host communication methods such as kernel message queues. Of course, these choices will be hidden from the application programmer, as well as the user. The TSDP code will decide the appropriate connection at runtime.

While this is being accomplished, development of TSDP filters for the various supported raw data formats will also proceed. The filters will sit between the database and the tools, taking raw data and automatically converting them to TSDP "on the fly" before handing them off to the tool. Initially, filters will be made for the data formats of NPA researchers who do not yet have them and for the outside collaborators who will interact in the development of these facilities (see Appendix). Additional filters will be made as needed, as new researchers come on board or existing researchers adopt new data file formats. Once the TSDP has been tested and used, a high-level "scripting language" will be developed for constructing tool pipelines and "composite tools" in batch mode. This language will provide mechanisms for attaching tools to one another, flow control and looping mechanisms.

E. TOOL DEVELOPMENT

Time Series Analysis Tools. This project's neuronal time series tool development effort is centered on *TSviewer*, a general purpose browser for multichannel time series data being developing over the past several years. TSviewer makes use of dynamically linked plugins to provide extended functionality as detailed in (see Section II, B, "analysis and visualization tools"). New work in this area will involve extending the TSviewer plugin tool set to include the most commonly used analysis techniques. All of the techniques will be incorporated that were identified at the Multi-Unit Recording Workshop of the 1994 Computational and Neural Systems conference (CNS*94). Workshop participants identified 19 widely used techniques (the full list is available at http://www.klab.caltech.edu/~pam/cns94_2.html) including correlation-based techniques (auto correlation, cross correlation, spike triggered averaging, event triggered averaging, peri-stimulus time histograms, etc.), information theoretic approaches, spectral analysis, wavelet analysis, principal components analysis, nonlinear topological analysis, and gravitational clustering. In addition, TSviewer plugins will be developed that provide   integrated support for MATLAB (The Math Works, Inc.), a commercially available numerical package that provides many powerful analysis tools and graphical display capabilities for neuronal time-series data. In terms of improving the TSviewer infrastructure, extensions will be added to the application program interface (API) to make it easier to chain together multiple plugins to create TSDP analysis pipelines. Also, the variety and functionality of graphical elements (e.g., sliders, push-buttons) which plugins can add to the user interface will be enhanced.

Tools Developed by Other Groups. This project establishes a community-wide repository for neuronal time series analysis tools. As indicated by the Appendix material, working relationships have been formed with the Neuronal Structure and Function Database directed by Dr. Gwen Jacobs and Dr.John Miller at Montana State University, the Modeler's Workspace project of Dr. Jim Bower at Cal Tech and the USC Human Brain Project of Dr. Richard Thompson, Dr. Michael Arbib and Dr. Larry Swanson. In addition collaboration is ongoing with the resident Biology Workbench project of Dr. Shankar Subramaniam and Dr. Eric Jakobsson. Dr. Jakobsson is a Co-PI on this project. This project's web site will contain both locally-developed tools and tools developed by other groups. Novel time series analysis techniques developed by other groups will be made TSDP-compliant (while retaining the original functionality and ability to read other data formats). With the willingness of the original authors, community access to these tools will be provided from the NTSA Workbench web site. In general, tools will be provided as TSviewer plugins which can be downloaded and used with the TSviewer application. Where appropriate, links will be provided to the original versions of the tools in their raw format (MATLAB scripts, C code, etc.). Digital Atlases and the Atlas Viewer program will be made available to many who have requested them (examples in Appendix).

F. PLANNED USE OF NTSA WORKBENCH IN BECKMAN INSTITUTE LABORATORIES

Thomas Anastasio's research involves both experimental and computational   studies of   the neural dynamics of sensorimotor transformations, with a focus on the vestibulo-ocular reflex (VOR) in goldfish (e.g., Anastasio, 1992; 1994a; 1994b, 1995). The VOR stabilizes retinal images by producing eye rotations that counterbalance head rotations.   It works by transforming head rotational velocity signals from the vestibular receptors into commands for controlling eye rotations. It is mediated by interneurons in the vestibular nuclei that receive and process information from vestibular sensory afferents, and send this modified information on to the

motor neurons of the eye muscles. This system is studied by recording eye rotations, eye muscle tension, and the single-unit responses of the various neurons, as the VOR is elicited by head rotation. In many cases, lesions have been made in the involved circuitry to test hypotheses about circuit functions.

The data are time series in the form of dynamic stimulus/response sequences, consisting of a head rotation signal to be correlated with the VOR, as indexed variously by neuronal firing, eye muscle tension, or an actual eye movement. Often, more than one of these response modalities are recorded at the same time, each from multiple sources (for example, rotation of both eyes and firing rates of neurons in several sites). NTSA Workbench tools will allow cross-correlations to be computed for all of the modalities and sources, with the stimulus and with each other, a task not readily accomplished now. Responses of the VOR and its constituent neurons are simulated using computer algorithms, and in all cases, correlations of the simulated data with the real thing are needed, again, a task that will be greatly facilitated by the proposed NTSA Workbench analysis tools. Dr. Anastasio will also incorporate a digital atlas for the goldfish, to store images of the histological material for entry and retrieval of neuronal time series in various parts of the VOR-relevant circuitry as well as the precise "shape" of the lesions in various preparations.

David Clayton's research focuses on neural plasticity and its limits. In particular he is interested in the question:"what constrains plastic change in the brain to particular circumstances, contexts or developmental ages?" The primary experimental system in which these investigations are made is the songbird, specifically the zebra finch. Finches learn to sing first by listening to tutors and then by practicing on their own, but song plasticity is sharply restricted to a critical period during adolescence. To understand why and how song learning is limited, Dr. Clayton has identified specific genes that are selectively expressed in the brain at times of learning.

Although Dr. Clayton has identified a number of genes whose expression is correlated with aspects of song system development and function (Clayton, 1997), his research has come to focus primarily on two. Synelfin and Zenk. Interestingly, Synelfyn is selectively absent in the adult song control nuclei embedded in the forebrain but found to be specifically increased in a key regulatory nucleus of the song control circuit early in the critical period for song learning (George et al., 1995). Dr. Clayton found that ZENK, an immediate early gene (IEG) is expressed in response to the sound of birdsong (Mello, Vicario & Clayton, 1992), but in a region of the brain (NCM) not previously implicated in song-related function. The gene is also induced when a bird sings, but in a different anatomical pattern (primarily involving the song control nuclei) (Jin and Clayton, 1997) . IEG expression is widely regarded to have a role in the consolidation of labile synaptic change into longer-lasting forms. Hence the pattern of ZENK activity in the brain may indicate sites where mechanisms of synaptic consolidation have been activated by a particular behavioral context. ZENK activity, in short, may be a indicator of circuit plasticity. This is supported by further observations that the ZENK response changes with stimulus repetition (Mello, Nottebohm and Clayton, 1995) , with the developmental progress of song learning (Jin and Clayton, 1997) , and in a way that is not easily explained by changes in gross electrophysiological activity (Stripling, Volman and Clayton, 1997) .

The NTSA Workbench is an ideal workgroup tool for the Clayton laboratory, and his group has made enthusiastic beginnings toward its use. One of their major efforts is to establish the anatomical and functional relationships between sites of electrophysiological activity and ZENK gene activation. To do this, simultaneous spike data are being collected from multiple electrodes in awake zebra finches. Also, comparison of response patterns in juvenile finches during the critical period for song learning are being made, and in all cases results are interpreted in the context of maps of ZENK expression established in other studies with in situ hybridization to measure ZENK mRNA in specific cells and brain regions. Clayton's physiological analyses are being done with tools already available in the NTSA Workbench library (Matlab and TSviewer). Additional NTSA Workbench uses will be: 1) construction of a digital atlas of the zebra finch brain, for use in cataloging and comparing data collected from birds of different ages and behavioral states, including measures of both electrophysiological and genomic (ZENK) activities; 2) development of analysis and visualization programs to evaluate intercorrelations among simultaneously-recorded neurons, and among neurons in sites of specific electrophysiological activity and ZENK expression; 3) to encode the laboratory metadata and organize a lab-specific database interface for data storage and access, again representing both Zenk expression (assayed by in situ hybridization, cell counting and autoradiography) and electrophysiological activity (assayed by both single-unit and multi-electrode recording).

Albert Feng studies the neural mediation of sound pattern recognition in frogs and bats. In frogs, the focus is on the mechanisms of sound localization and sound coding in the presence of background noise. This work has elements in common with "cocktail party hearing", wherein coherent perception of the location and meaning of sound is achieved in the presence of other competing sounds (Gooler et al., 1996). Currently the Feng laboratory is testing hypotheses about how masking by noise is released by the angular separation of the signal and noise sources. The work on bats concerns the neural basis of discrimination of flying insects. Different insects induce different

amplitude modulation patterns in a sequence of discrete echoes. Tthese patterns are used to identify different insects. The analysis of this phenomenon is approached by studies of central auditory neurons which encode amplitude modulations that span successive sound pulses (Condon et al., 1996a; 1996b). Primary emphasis is on how an active change in sonar emission can sharpen the neuronal response selectivities to sound amplitude.

These studies involve collection of large data sets from each neuron, ranging from the unit's basic responses to tones and noise, to such things as a unit's response to complex situations such as a tone from one direction and noise from a different direction, and to sound pulses of uniform amplitude presented at different repetition rates or sound pulses of variable amplitude at a constant repetition rate. The NTSA Workbench database will readily allow coding as metadata the wide array of stimulation conditions employed in these studies. In addition, the system is well-suited for comparing the responses of different neurons to the same experimental conditions, and for correlating response patterns to complex sounds with units' basic response properties. Dr. Feng is especially eager to set up is laboratory-specific database interface during the next project period in order to gain more organized and flexible access to present and previous neuronal records. Also, Dr. Feng will employ the analysis and visualization programs now available in the system's library (e.g., Matlab, TSviewer, BMDP) as well as other analytical tools to meet specific needs such as sequential interval analysis (Optican and Richmond, 1987). Recently, the Feng laboratory has begun to model the neural connectivity needed to account for neuronal response characteristics during sound pattern recognition and prey discrimination. Dr.Feng is enthusiastic about use of NTSA Workbench facilities (e.g. GENESIS) for model formulation and testing.

Michael Gabriel has used a "model system" strategy to analyze the neural mediation of learning. Discriminative training paradigms are used, wherein rabbits learn to make responses to simple auditory cues which signal reward or shock, and they learn to ignore other, non-signaling cues. Multichannel data acquisition systems allow recording of multi- and single-unit activity during learning in awake, behaving animals, simultaneously in six anatomically-related brain areas. Neuronal recording is combined with neurochemical or electrolytic lesions, which deprive learning- relevant neurons of input from upstream structures (e.g., Poremba and Gabriel, 1997a; 1997b). This strategy has yielded a model of the basic brain circuitry for learning, and the dynamic circuit level information processing that gives rise to memory retrieval and performance of the learned behavior. Key elements in the circuit for learning are the medial geniculate nucleus, the amygdala, cingulate cortex and the limbic (anterior and medial dorsal) thalamic nuclei (reviewed by Gabriel, 1990; 1993).

Underway since the mid-70's, this project has yielded a massive accumulation of multi-site, learning-relevant neuronal data. Training sessions (from ten to thirty per subject) have been given to approximately 1800 rabbits to date. The data of each session are stored in a 1.2 MB uncompressed "T-files". With the addition of DataWave Systems Inc. workstations, each session now yields an additional 2 - 10 MB file (a "UFF" file) which stores individual spike waveforms, sampled at high frequency for off-line spike sorting. .

Exciting prospects await the full implementation of the NTSA Workbench, particularly in regard to the database component, which will allow Dr. Gabriel to obtain views of the learning-relevant brain information flow never before achieved, by combining data from hundreds of records from separate experiments now stored on NTSA Workbench optical discs. For example, he will be able to build a picture of the neuronal activity recorded during each tone presentation, throughout the entire course of behavioral acquisition in subnuclei of the auditory thalamus, the amygdala, each of six layers of anterior and posterior cingulate cortex, subnuclei of limbic thalamus and all subfields of hippocampus. (At present, processing restrictions allow, practically, only analyses of neuronal data averaged over many trials. The averages, while obviously useful, doubtless obscure many important changes that take place during acquisition). Other exciting projects such as extensive analyses of trial sequence effects on cue-related neuronal patterns in many brain areas, analyses of within-, and between-area cross-correlations and 3-D animated visualization of the changes in these areas using the digital rabbit atlas can potentially yield unprecedented views of dynamic learning-related brain processes in a realistic neuroanatomical context.

Rhanor Gillette studies foraging behavior and learning in marine molluscs. In their foraging behavior all motil animals perform cost-benefit analyses in decision-making concerning approach and avoidance of potential food stimuli. Information from sensory sources is integrated with internal state (hunger, health, hormonal-condition) and experience to formulate least-probable-error behavioral policies in the midst of a noisy environment. Dr. Gillette studies the simple predatory marine snail *Pleurobranchaea californica* to probe the network substrates for these decisions. His laboratory is describing the neural networks mediating feeding and avoidance behaviors, how these networks interact and how their interactions are modulated by hunger/satiation and food avoidance conditioning (Gillette et al., 1997; Jing and Gillette, 1995; London and Gillette, 1986). He and his associates have recently identified nitric oxide and serotonin as factors that regulate feeding arousal and whose CNS levels are related to hunger. They have shown that these agents act in large part through a cyclic cAMP-gated ion current in the

feeding network (Cruz et al., 1997; Fuller et al., 1998; Gillette et al., 1997; Moroz et al., 1997, 1996a;1996b; Sudlow and Gillette, 1997; Sudlow et al., 1998). Immediate plans complement these studies in a parallel undertaking: modeling of these networks, their interactions and the actions of the neuromodulatory factors NO and 5-HT as they bias behavior via actions on identified neurons. Model networks for feeding, locomotion, swimming and turning will be based on documented real network structure in terms of identified neurons and connections, and updated as experimental information arrives. Time series data records from elements of the model network will be organized, stored and analyzed using the NTSA Workbench database system and a lab-specific interface that will allow us to access data with analytical and visualization programs in the system's library.

William Greenough.   The primary work for which Dr. Greenough's lab has been using the NTSA Workbench involves a novel form of plasticity in the cerebellar cortex. Previous cerebellar electrophysiological studies have identified one form of plasticity, termed long-term depression, in which conjoint activation of parallel fiber and climbing fiber input to the cerebellar cortex reduces the postsynaptic response of Purkinje cells to parallel fiber activation. This response persists for at least several hours although perhaps not for long enough to account for the persistence of learned motor skills with cerebellar components. Dr. Greenough had previously shown that both learned motor skills and training-induced increases in parallel fiber-to-Purkinje cell synapse numbers persist for at least 4 weeks following a 10 day period of motor skill training (Kleim et al., 1997a 1997b). The increase in the numbers of these excitatory connections suggests that there should be an enhancement of Purkinje cell responsiveness to parallel fiber activation. Data from certain motor learning tasks and from eyeblink conditioning studies has similarly suggested an enhancement of parallel fiber to Purkinje cells in at least a proportion of cells. Responses of cerebellar cortex to parallel fiber activation are now being studied using field potential recording. It is possible by examining frequency following (1 Hz vs. 50 Hz) to identify the presynaptic and postsynaptic components of the evoked response separately. The cerebellar cortical evoked response to intracortical parallel fiber stimulation is enhanced in animals that receive motor skill training on an elevated, difficult obstacle course, compared either to animals that are inactive except for brief daily handling or to animals subjected to a motor control procedure in which they traverse an alleyway without acquiring appreciable new motor skills (Swain et al., 1998). This result is being followed up with recordings of the activity of individual Purkinje neurons. The data demonstrate that the newly-formed parallel fiber synapses are functional and that the cerebellar cortex exhibits at least two forms of plasticity, LTD, a reduction in parallel fiber driving of Purkinje cell activity, and the behaviorally-induced plasticity described here, an apparent increase in parallel fiber driving of Purkinje cell activity. This thus represents a novel form of cerebellar plasticity that can be detected outside of the training situation. Future studies will analyze physiological plasticity associated with behaviorally-induced anatomical plasticity observed in somatosensory and visual cerebral cortex.

Eric Jakobsson's research interests center around mechanisms and functional significance of ion permeation in biological membranes. His studies are computational and theoretical, and the level of biological organization he studies is both molecular and cellular. His lab has worked on the details of ion permeation in the antibiotic gramicidin A. The group has elucidated the underlying physics of both ion and water permeation in the simple channel and laid the groundwork for doing the same types of studies on more complicated protein channels such as the voltage-gated potassium and sodium channels. Dr. Jakobsson is a co-developer of the very successful Biology Workbench, an Internet-based computing environment for molecular bioinformatics.

Joseph Malpeli will use the NTSA Workbench to make available to the scientific community two sets of data on the primate lateral geniculate nucleus (LGN): a combined anatomical and functional atlas of the rhesus macaque LGN, and an anatomical atlas of the human LGN. The LGN relays information from the eye to visual cortex. It is composed of several layers, segregating information according to functional type and eye of origin. Each layer contains a retinotopic map of visual space, and all maps are in physical register, so that a point in visual space is represented in the LGN as a "projection column" cutting through all layers. The retinotopy of the rhesus LGN was qualitatively described by   Malpeli and Baker (1975), who densely mapped a single LGN with microelectrode recordings. This appears to be the only such data set. Recently, Malpeli et al. (1996) quantitatively reanalyzed these data by creating computer representations of selected "sections" through this same LGN, in which retinotopy superimposed on the laminar morphology, along with cell density functions obtained from the original, serially-sectioned brain. Currently, these data are being interpolated throughout the volume of the LGN to create three-dimensional, space-filled representations of morphology, cytology and retinotopy. These will be combined into a single digital atlas at a 25-micron voxel resolution which will be made public on the NTSA Workbench Database. The NTSA Workbench Digital Atlas Viewer will be incorporated with these atlases to provide a convenient means for researchers to inspect and use these atlases.

The reconstruction of the monkey LGN has already provoked unique ideas about the morphogenesis of the LGN's internal laminar structure, leading to the first computational models of geniculate development (Lee and Malpeli 1994; Tzonev et al 1995; Tzonev et al, In Press). The human LGN study is being undertaken to look for comparative evidence for or against these ideas. This data set will consist of several (up to five) three- dimensional atlases with a voxel resolutions of 5x5x40 microns. Every serial section of five human nuclei has been digitized, and the images are now being individually aligned and contrast-matched. For the one atlas that is finished, the contrast and section-to-section alignment are good enough to make clearly visible the detailed laminar structure in computer "cuts" at arbitrary angles through the LGN.

The atlases should be a useful tool for understanding the function, structure, and morphogenesis of the LGN. Manipulating them in three-dimensional projections will afford insights that are difficult to grasp from verbal descriptions or static figures. For example, with the monkey atlas, one can visualize the projection columns associated with any locus of points on the retina, or obtain the number and class of cells activated from an arbitrary region of visual space. Comparisons of the human and monkey atlases may afford insight into the functional organization of the human LGN not now apparent from the limited 2-D views available in the literature. The LGN provides a well-ordered laboratory for understanding some basic rules of morphogenesis, and when these atlases are public, others may be encouraged to extend or supersede the initial theoretical models of morphogenesis (Lee and Malpeli 1994; Tzonev et al 1995; Tzonev et al, In Press) that these data have evoked. The NTSA Workbench provides an excellent vehicle for widely disseminating these data.

Mark Nelson's laboratory is exploring principles involved in active sensory acquisition, namely how animals acquire, enhance and represent information about their surroundings. Using a combination of behavioral, physiological and computer modeling approaches, his laboratory studies sensory acquisition in the electrosensory system of weakly electric fish. The electric sense allows these nocturnal animals to hunt and navigate in total darkness. Experimental work includes behavioral studies of prey capture (MacIver et al., 1997) and in vivo neurophysiological recordings from electrosensory neurons (Nelson et al., 1996; 1997). Computer modeling studies (Nelson,1996) focus on adaptive sensory processing alogrithms in biologically-detailed models of a hindbrain sensory nucleus, the electrosensory lateral line lobe (ELL). Simulations of ELL circuitry are implemented using a general purpose neural simulation package, GENESIS, which Dr. Nelson helped develop at Cal Tech.

Dr. Nelson plans to use the database system to organize the time-series data generated by his neurophysiological experiments and computer simulation studies, to facilitate comparisons between the two. The database system will be used to keep track of all parameters and conditions associated with both simulated and experimental data sets. The database system will facilitate the following data management tasks: (1) finding existing and generating new simulated data sets that match the conditions of an experimental data set, (2) generating groups of simulated data sets in which the model parameters are systematically varied (parametric studies), and (3) extracting and comparing simulated and experimental data sets to find regions of the parameter space that are consistent with the experimental data. Without a database system, these tasks require a tremendous amount of additional bookkeeping, a great deal of time, and are prone to error.

Bruce Wheeler and his co-workers use the NTSA Workbench for analyses of data yielded from 64 electrodes in a planar array. Their attention has been focused on the patterning of novel materials to the surface of the array so as to create sparsely populated yet highly defined neural networks with the neurons localized largely to the electrode sites (Wheeler and Brewer, 1994). They expect to stimulate and record individual action potentials from several neurons per channel, with initial recordings having either very short trials (tens of milliseconds) or consisting of continuous recordings over many minutes. As the work progresses they expect to archive recordings from many cultures over many trial sessions, and will want retrospectively to ask questions about the changing nature of the firing rates of individual neurons and of the collection of neurons, about the correlation of activity among pairs or groups of neurons, and how stimulation affects that correlation.

Dr. Wheeler's group will continue to use the NTSA Workbench for development of algorithms for sorting of spike waveforms and pattern recognition as applied to large numbers of channels of neural data. Currently, techniques for developing automated spike sorting routines are being emphasized (Wheeler, 1998, in press), and successful spike sorting techniques are being incorporated into the suite of capabilities of the NTSA Workbench. However, work is also being done on novel techniques for analyzing and displaying the interactivity among large numbers of neurons as inferred from their spike train activity (Wheeler, 1995).

Robert Wickesberg studies two different aspects of how the peripheral auditory system processes sounds. One project concerns auditory nerve encoding of speech sounds. A pilot study combined the peri-stimulus time (PST) histograms based on responses of individual auditory nerve fibers to a consonant-vowel syllable into one grand average histogram (GAPST). Each syllable was presented in normal and whispered voicing at several intensities. For

each consonant-vowel syllable and each method of voicing studied, this method yielded a unique GAPST, that was relatively invariant with changes in intensity. The combining of more than 70 individual PSTs from 20 different animals was a formidable task. This was accomplished using Quattro Pro, a relatively small commercial database program. The fairly simple preliminary analysis that was attemped was at the limit of the Quattro Pro's capacity. It will thus be greatly beneficial to move this project to the NTSA Workbench and to develop the necessary analysis tools. Initially standard TSDP-compliant analysis tools for the study of auditory spike trains will be developed. A model of how the auditory nerve encodes these stimuli is now under development, the output of which will be time-series data that must be analyzed as actual auditory nerve fiber activity. The NTSA Workbench provides excellent tools for this purpose. The second project involves studying the processing of echoes by the peripheral auditory system. This involves obtaining the responses of individual neurons to stimuli with and without echoes.   The responses are in the form of spike times. While a set of standard routines is currently available comparison across populations of neurons are difficult. Standard analysis routines will be introduced as TSDP-compatible tools in the NTSA Workbench library, for analysis of the aforementioned neuronal data. The NTSA Workbench database will facillitate the population comparisons and afford means for cross-species comparisons in collaboration with Dr. Feng.

G. PUBLICATIONS:
The PI and Co-PIs have published extensively since this project was previously reviewed, as shown from the text above and the Biographical Sketches. In most cases the published work was aided by NTSA Workbench facilities. The article cited below to be submitted in spring of 1998 describes the goals and basic components of the NTSA Workbench. Additional papers describing the implementations of the NTSA Workbench Database Table Schema, TSDP, TSviewer and the associated modular programming library will be submitted for publication later in 1998.

Payne J, Mungnirin K, Quinn S, Nelson M, Gabriel M (1998) An information system for neuronal pattern analysis.

H. HUMAN RESOURCE DEVELOPMENT IN SCIENCE AND ENGINEERING
The NTSA Workbench is an ideal venue for training in relation to state-of-the-art bioinformatics technology. The Beckman NPA Group faculty have a history of extensive postdoctoral and graduate training and trainees make extensive use of the NTSA Workbench components. Many laboratories also meaningfully engage undergraduate students in ongoing research. Recent history in PI Dr. Gabriel's laboratory affords a representative example. Three graduate students (Eunjoo Kang, Yasuo Kubota and Amy Poremba) received the Ph.D. in Biological Psychology since 1995 and are currently doing postdoctoral work respectively at Stanford, MIT and NIH. They extensively used NTSA Workbench analysis tools and digital atlas images for their dissertation research (see publications in the Journal of Neuroscience and Brain Research: Poremba & Gabriel, 1997a, 1997b; Kubota et al., 1996). Postdoctoral fellow Dr. John Freeman who just assumed a tenure track assistant professorship in Biological Psychology at the University of Iowa also used the NTSA workbench extensively (e.g.,   Freeman et al., 1996; 1997). Six to 12 Illinois undergraduates participate in an extended research curriculum in Dr. Gabriel's laboratory, for which course credit is earned as the students spearhead their own research projects. After program completion one of these students (William Holt) earned project funds as an hourly technician to scan and refine the coronal rabbit brain images used for the digital rabbit atlas. Mr. Holt is now a graduate student at the University of Michigan. Charles Shyu and Nick Pastorek are among 12 currently involved undergraduates. These two students are carrying out research based specifically on use of the database for meta-analysis of archival data.

APPENDIX

Letters of Support