Physiology Analysis System

Jaques Reifman[†], Jamila Gunawardena, Zhenqiu Liu

Bioinformatics Cell
Telemedicine and Advanced Technology Research Center
U.S. Army Medical Research and Materiel Command
MCMR-ZB-T, 504 Scott Street
Fort Detrick, Maryland 21702 USA
[†]reifman@tatrc.org

Abstract - *In order to address the needs of managing and analyzing large amounts of continuous physiologic data for a diversity of human and animal research studies, the U.S. Army is developing the Physiology Analysis System. This flexible, modular, and interactive research system integrates both data and analysis tools at a central server location, eliminating painstaking downloads to the researcher's workstation. The system offers, through a Web browser, a feature-rich, workstation-like environment that analysts require to perform sophisticated post-hoc data mining. For proof-of-demonstration, the Physiology Analysis System is being used to warehouse and analyze continuous pre-hospital vital-sign data collected during transport from the scene of the incident to the trauma center for over 900 civilian trauma casualties.*

Keywords - *Physiologic Monitoring, Medical Informatics Applications, Medical Informatics Computing.*

I. INTRODUCTION

The Combat Casualty Care Directorate of the U.S. Army Medical Research and Materiel Command (USAMRMC), Ft. Detrick, Maryland [1], supports various intramural and extramural research studies involving the collection and analysis of human and animal physiological data. These studies generate voluminous amounts of time-series data, e.g., electrocardiogram waveforms (EKG), oxygen saturation waveform (SpO2), and respiratory traces, that traditional relational database management systems are ill-suited to handle. Therefore, these data sets are generally stored by investigators in their individual workstations and are manually manipulated for subsequent visualization and analysis.

The problem space involved in analyzing and managing these physiologic data sets is quite large, often requiring the collaboration of domain experts in diverse fields engaged in the research process. Large data sets, often ranging between tens to hundreds of gigabytes, need to be distributed to the collaborators requiring each research team to locally maintain the tools necessary to properly handle and store the data. Facilities for querying and analyzing the data must also be developed. To establish common facilities amongst the various groups, issues of hardware and software incompatibility must be overcome. Furthermore, the lack of an integrated system for management and analysis of large sets of time-series data forces the users to develop endless data input-output interface systems for sequential data

analysis. Established tools exist for handling some, but not all, of these issues. Nevertheless, the burden is placed on the research team to construct a solution from scratch resulting in each research study requiring its own effort to combine data storage, search, and analysis.

To afford a flexible solution to warehouse and manage physiologic data sets and provide analytical tools to research scientists for *post-hoc* analysis of time-series data in a central location, the USAMRMC Bioinformatics Cell is developing the Physiology Analysis System (PAS). PAS is a system designed to support the disparate needs of storing and mining data from various studies conducted throughout the USAMRMC laboratories and extramural partners. For demonstration and evaluation purposes, it is currently being used to warehouse and analyze continuous pre-hospital physiologic vital sign data collected from trauma patients during transport form the scene of the incident to the trauma center. Here, we present this trauma dataset, collected by the University of Texas Health Science Center at Houston (UTHSC), to provide a context when discussing the capabilities of PAS.

II. TRAUMA DATA

The trauma dataset currently consists of over 900 patients collected by the UTHSC Life Flight helicopter ambulances [2]. Information is collected from patients using a Propaq 206 vital signs monitor [3] and recorded in a flash memory card via a handheld computer. The monitor is attached to the patient at the scene of the incident prior to loading the patient in flight, and data are recorded until arrival at the hospital where the monitor is disconnected. The Propaq continuously monitors several physiologic variables, including EKG, SpO2, respiratory rate, non-invasive blood pressure (NIBP), end-tidal CO_2, and heart rate, with collection time ranging from 10-min to 30-min.

Each continuous physiologic variable from the Propaq dataset is gathered at a different sampling frequency. Thus, over the same collection time period, the number of data points varies amongst each variable, presenting a challenge for analysis. The highest resolution variable, EKG waveform, is gathered at approximately 180.0 Hz, which results in approximately 324,000 data points for a 30-min data
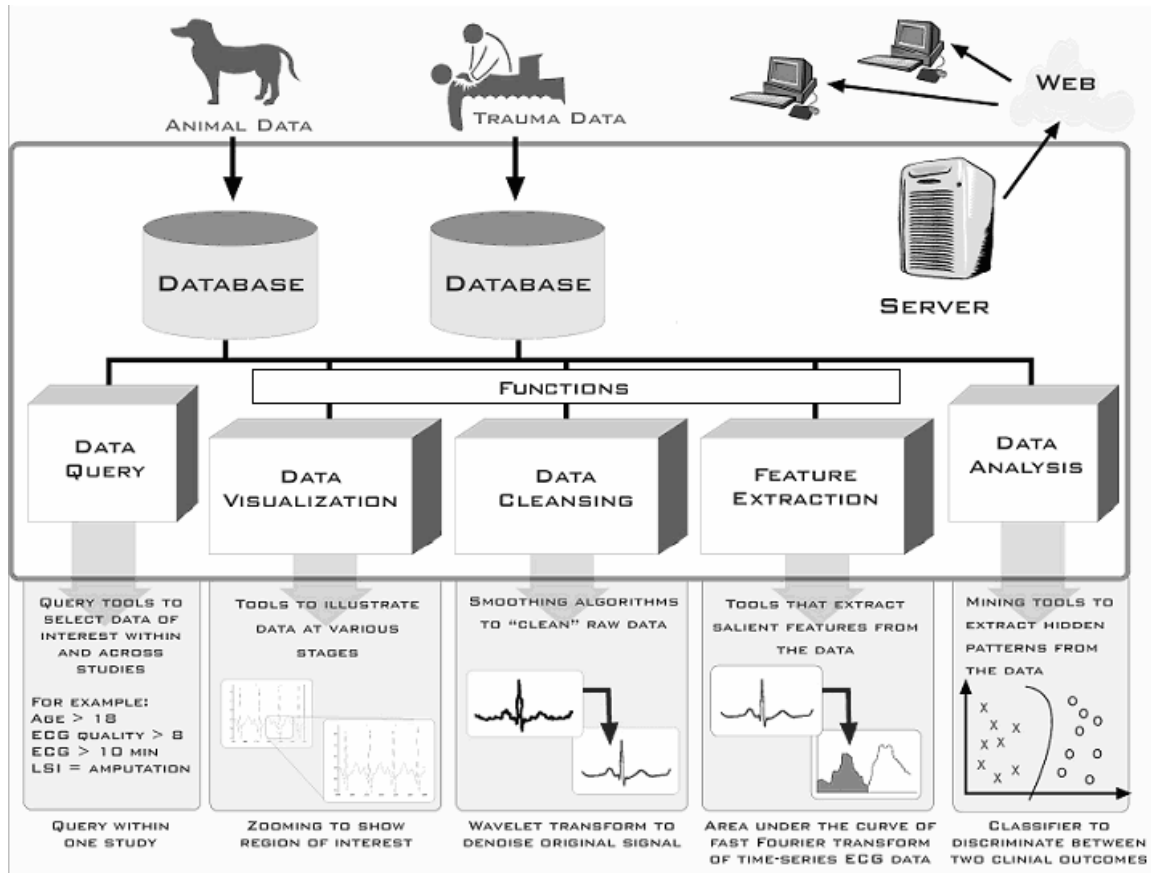
Fig. 1. Overview of the Physiology Analysis System architecture and capabilities.

acquisition. SpO2 and respiratory waveforms are gathered at approximately 91.0 Hz and 45.5 Hz, respectively. The lower frequency time-series data, including NIBP, end-tidal CO_2, respiratory rate, and heart rate, are available at 1.0 Hz. However, new values of NIBP are only collected every 2.0-3.0 minutes, depending on the Propaq setting.

In addition to the Propaq vital sign data, other pertinent information, such as demographics data, cause and nature of injuries, life saving interventions (LSIs) performed, and types and quantities of fluids delivered, are recorded and available for analysis.

The objective of the pre-hospital trauma study is to identify physiologic parameters that are early indicators of some clinical outcome, such as a need for a LSI, leading to battlefield-usable triage algorithms that help the combat medic identify the immediate need for casualty evacuation.

III. THE PHYSIOLOGY ANALYSIS SYSTEM

A. *Overview*

Built on a combination of proven computational platforms and easily accessible through the Web, PAS provides a flexible, modular, and interactive environment for rapid prototyping and the incorporation of analytical tools. Facilities and tools provide for the storage of new studies and to query, visualize, cleanse, extract features, and analyze physiologic data with emphasis on tools to analyze time-series data. The innovative concept of PAS is an architecture where both the data and the analysis tools reside at the server while offering—through a Web browser—a feature-rich, workstation-like environment that researchers require to perform sophisticated data mining (Figure 1). This architecture facilitates on-line collaboration among different research teams, allows for inter-study comparisons, and eliminates the need to download tools and data from the server to the desktop, eliminating the often-painstaking process of installing the downloaded tools.

Users connect to PAS using a standard Web browser and authenticate into their individual accounts by typing a user name and password. Inside the system, user interaction is segmented into three distinct components: *Query*, *Functions*, and *Analysis*. Albeit separate and independently customized by the user, these components are integrated by an automatic tracking system, which keeps track of the sequence of operations performed in a given session. These operations can be saved and stored in the server for future use. The three components are easily applied through pull down

menus designed to permit the construction of arbitrary sequence of operations that analysts perform during data mining. PAS allows researchers to transition away from the mundane yet necessary task of data and system management and focus on data mining and analysis.

During the *Query* phase, the researcher selects a subset of subjects from the study. Once those subjects have been identified, they are recorded within the system for use during Analysis. During the *Function* phase, analytical functions—in the form of data visualization, data cleansing, and feature extraction tools—available through a library of functions within the system, are assembled by the researcher to operate on variables present within the study. Any library function may be applied to any variable in any order, subject to function constraints. In the *Analysis* phase, the efforts of the previous two phases are married. The functions are applied to the subjects harvested by the queries, and the results are fed into additional mining tools for more involved analysis, such as feature selection and classification algorithms.

B. *Query Management*

Users construct queries by selecting variables from the list of potential choices within the study and limiting the ranges of the variables by selecting the corresponding constraints using the graphical user interface in Figure 2. Once the list of constraints is created, the users run the query and the system returns a list of subjects that meet the selected constraints. Figure 2 illustrates a query that shows the selection of female patients between ages 30-40 which had a pre-hospital intubation LSI. From this list, each subject can be selected for viewing in greater detail. The system records all the information related to a query, including any selected sub list, as a named object within the system, allowing the researcher to track exactly what constitutes the outcome of a particular query.

C. *Function Management*

During the Function phase, the researcher selects variables upon which to perform analysis. First, a variable from within the study is chosen. Then, individual functions are selected in sequence from a library of functions to construct a *function chain*, which defines the order of the mathematical operations to be applied to the chosen variable. These functions allow for data visualization, data cleansing, and feature extraction, spanning a variety of tools, such as data zooming, wavelet transform for noise filtering, and mathematical and statistical functions like minimum, maximum, standard deviation, and trending for feature extraction. The system allows for easy and rapid integration of user-provided and third-party-developed analytical functions into the system's library of functions [4]. Functions can be written in C, FORTRAN, Java, or Matlab, allowing algorithms to be provided by various research teams.
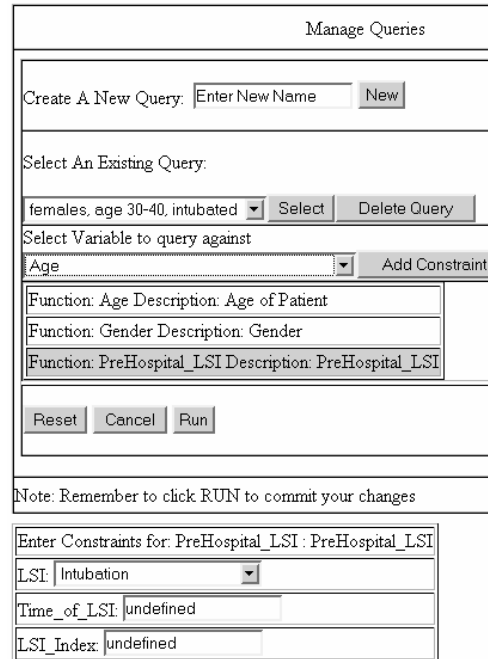

Fig. 2. Query construction, indicating the request for "female, ages 30-40, intubated."

Function chains permit the user to customize a sequence of data operations by selecting functions from the library of functions and arranging them in any desired order so that the output of one function can serve as the input to a subsequent function. Built-in mechanisms automatically construct function input-output interfaces and check for function-to-function consistency. Another advantage of function chains is that they only instantiate the functions in the sequence and generate derivative data when requested by the user, reducing the amount of data the system has to manage. Once these functions are completed, they are saved as named objects within the system and accessible for later use.

In Figure 3, a function chain is constructed to identify the peaks and compute the R-R interval distance for a small range of an EKG trace. This is done by selecting the "EKG Waveform" variable type, and subsequently selecting two functions: "Range Selection" and "RR Interval." If the researcher desires to instantiate the function and see its output in graphical or numerical format at any point along the function chain, the researcher can simply click on the link associated with that function. For instance, the graphical visualization of the R-R interval in Figure 4 is constructed by clicking on the "Graph" button to the right of the "RR Interval" function. If the researcher wishes to compute the standard deviation of the distances among the peaks, he/she simply selects the standard deviation function from the "Select Function To Add" pull down menu in Figure 3 (i.e., the library of functions), clicks the "Add Function" button to the right, and the standard deviation function is placed at the tail of the function chain.
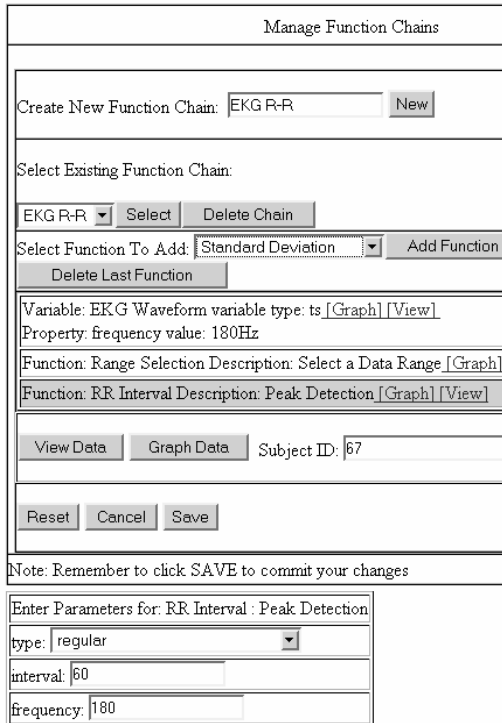
Fig. 3. Function chain construction, consisting of
the selection of "EKG Waveform" followed by
functions "Range Selection" and "RR Interval."

Fig. 4. Graphical output of the R-R peak detection.

D. *Analysis*

The Analysis component allows the user to apply a desired set of function chains to any subset of the query results. For example, Table 1 shows the results of running the EKG function chain (Figure 3, after adding standard deviation) against the subjects from the query "females, age 30-40, intubated" in Figure 2. This allows users to easily apply functions across a group of subjects in an interactive manner, permitting function chains and queries to be modified and fine tuned, leaving the system to track the mechanisms being used to produce the analysis.

The Analysis component also hosts algorithms for in-depth multivariate data mining. Tools are being developed to collect the feature variables resulting from the application of function chains to queries and to concatenate such features into input files for subsequent application of feature selection and classification algorithms. Feature selection algorithms, such as linear correlation coefficients and information entropy, and classification algorithms, such as artificial neural networks, decision trees, and support vector machines, are being integrated into PAS for sophisticated data mining.

IV. DISCUSSION

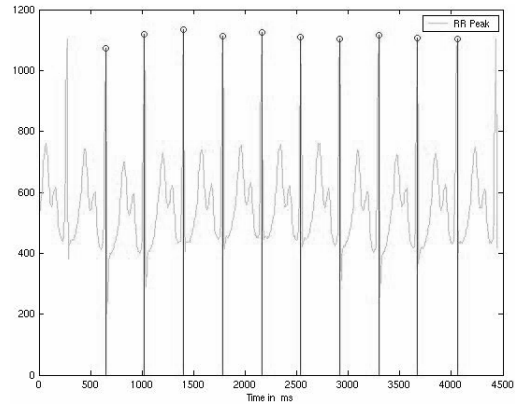The Physiology Analysis System design emphasizes performance, modularity, flexibility, and third-party software

integration. The initial test dataset, the Trauma Data, allowed fleshing out PAS with a pertinent set of functionalities to evaluate the capabilities and limitations of the system design. In the Query component, future work involves providing facilities for time-series data query, such as: "provide the subjects/data for which there are at least 5 minutes of uninterrupted EKG data within 20 minutes of the incident." At present, queries can be performed only on non time-series data. Another salient query feature is cross study query. Currently, only a single study can be queried at a given time. It would be desirable to query the subject population across several studies, in particular, between human and animal studies.

Table 1. Results of applying the "EKG R-R" function chain to
the "females, age 30-40, intubated" query.

Subject ID	Standard deviation of R-R Interval (ms)
67	2.9
94	72.2
96	208.6
107	3.2
110	26.6
127	2.4
130	279.5
398	162.7
518	519.2

In the scope of the Function and Analysis components, the next step is to allow functions to be developed within the system itself, so that they can be added directly by the user

without the need for an engineer to manually add third-party algorithms. This will allow PAS to satisfy the research needs of a sophisticated and diverse user population. Also, more interactive visualization tools need to be provided to allow graphical exploration of time-series data.

V. CONCLUSION

The initial prototype of the Physiology Analysis System successfully achieved its initial design objective: to provide an integrated, interactive, modular, and flexible computer system that can host high-resolution, continuous physiologic data and allow users to quickly and easily perform data analysis tasks. By allowing processing to occur in a centralized repository, users are able to step away from issues of data management and focus on data analysis. The system is equally useful for experienced and not experienced data miners, as it allows for the creation and storage of analysis functions for repetitive usage and, through its modular and flexible architecture, the easy incorporation of user-provided and third-party-developed analysis tools.

ACKNOWLEDGMENTS

DISCLAIMER

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or the U. S. Department of Defense. "This paper has been approved for public release; distribution is unlimited."

REFERENCES

[1] U.S. Army Medical Research and Material Command, Fort Detrick, Maryland, USA. <http://mrmc-www.army.mil>.

[2] Convertino VA, and Holcomb JB. Advanced diagnostics for the combat medic. Army Medical Department Journal 2003: PB8-03-7/8/9; 42-48.

[3] Welch Allyn Protocol, Inc., Beaverton, Oregon, USA. <http://www.monitoring.welchallyn.com/>.

[4] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000: 101(23):e215-e220. <http://circ.ahajournals.org/cgi/content/full/101/23/e215>.