# ERPDB: AN INTEGRATED DATABASE OF ERP DATA FOR NEUROINFORMATICS RESEARCH

***QingHong Yan, Ning Zhang, Jie Wu, Tao Zhang*** [*]

*College of Life Sciences, Nankai University, Tianjin, 300071, China*
**Email*: zhangtao@nankai.edu.cn
*Email:mailto:yan_quinhong@yahoo.com;zhni@eyou.com;sohutu@mail.nankai.edu.cn;*
*zhangtao@nankai.edu.cn*

## *ABSTRACT*

*Event-related potential (ERP) is the measurement of the brain's electrical activity in response to different types of events, such as attention, words, thinking, or sounds. By measuring the brain's response to such events, we can learn how different types of information are processed. As the mass of recorded ERP data explodes, an automatic and accurate tool to store, manage, and retrieve data readily is of increasing concern in neuroinformatics. In this paper, we describe a relational ERP database that has been constructed using the SQL server 2000 database management system and an IIS web server that has been setup for data retrieval through a custom web interface (http://202.113.232.103:8088/erpdb/index.asp). A novel database structure has been used to store ERP data of different activity channels, which provides a rapid and accurate way for data retrieval within any given range on the time zone with various searching options. The database is divided into: (1) subjects' information and record information and (2) ERP data, which has been structured and standardized in a database table supplemented with unrestricted text files. It can integrate or exchange data with other clinical databases or computer-based information systems through a program based on ADO techniques. Users are able to readily retrieve ERP data through the user-friendly web page interface. All online resources of the database are freely available to the scientific community. As the database develops further, we anticipate it will become a valuable tool that will make a great contribution to everyday clinical practice, teaching, and research work inneuroscience and psychology in the future.*

**Keywords:** ERP, Database, SQL server, Neuroinformatics

## 1 INTRODUCTION

### 1.1 Event-Related Potentials (ERPs)

ERPs, one of the most important kinds of electrophysiological data, are the measurement of the brain's electrical activity (EEG signals) in response to different types of events, such as attention, words, thinking, or sounds - usually external stimuli or subject's movements (Picton & Hillyard, 1988). ERPs may either precede an event or follow it. They consist of a sequence of positive and negative voltage-fluctuation-labeled components, which are useful as measures of covert information processing (differences between conditions can be obtained through comparison to ERP in the absence of behavioral responding) (David & Ray, 2000).

One of the main reasons for using ERPs in psychology is that they present an intriguing possibility to obtain information about how the intact human brain processes signals and prepares for actions. Cognitive ERP has been used as a marker of cognitive function in patients with psychiatric and neurological disorders (Katada, Sato, et al., 2004). In contrast to overt behavioral measures and slow brain imaging techniques such as functional magnetic resonance, ERP reflects the information processing in the cortex in real time. Although little is known about neural correlation and functional meaning of ERP components, even this limited knowledge has led to inferences that could never be made if only the final results of many processes were recorded.

ERP signals are small (1-30 microvolt) and are embedded in electroencephalographic (EEG) activity unrelated to the eliciting event. However, they can be recorded non-invasively from the scalp by means of signal averaging techniques. The size (signal-to-noise ratio) and period (the component's frequency) of the ERP component can decide the number of trials required to produce a noise-free ERP.

There are three measurable aspects of the ERP waveform: amplitude, latency, and scalp distribution (Johnson, 1998). Component amplitude provides an index of neural activation extent (how the component responds functionally to experimental variables); component latency (the point in time at which the peak occurs) reveals the timing of this activation; and a component's scalp distribution (the pattern of voltage gradient over the scalp at any point in time) provides information on the overall pattern of activated brain areas.

## 1.2 ERP Research—The requirement of database

The neuroscience research on ERP results in substantial amounts of digital signals and textual information. As clinical techniques develop rapidly, the accumulating ERP data have created one of the largest datasets in biological science. Functional capability and flexibility in data retrieval depend on how they are structured and standardized. Usually, the raw data, together with additional information or interpretations, are stored in non-structured and non-standardized free text. In addition, the ERP tests are only designed for retrieving one by one (Aurlien, Gjerde, Gilhus, et al., 1999). As a result, an automatic and accurate tool (database) to store, manage and retrieve data readily is of increasing concern in neuroinformatics.

Databases offer accurate and timely data and clinical details for neuroscience research (Finnerup, Fuglsang-Frederiksen, Rossel, et al., 1999). Several neuroscience databases and computer-based information systems have been established previously. One example is Aurlien et al, who report an EEG description system facilitating categorization of EEG data in a computerized database (Aurlien, et al., 1999). Another example is the epilepsy and EEG information system, developed as part of a European multicentre project PRESTIGE. The aim of PRESTIGE is to bridge the gap between research and everyday clinical practice by deploying healthcare telematics technology for guideline generation, dissemination, and routine application (research-based and consensus-based) (Finnerup, et al., 1999) [6]. In addition, in the field of neural disease, Babiloni has set up a database online with respect to the Alzheimer disease (Babiloni, Babiloni, Carducci, et al., 2001). The Brain Resource International Database, supported by the Brain Resource Company in Australia, is the world's first integrated brain function database. It contains 33 dimensions of brain function through using standardized positions to identify the best biomarkers in a number of Brain diseases (http://www.brainresource.com/).

Many research efforts drew data from the database (Hermens, Michael, & Simon, 2005; Brickman, Paul, & Cohen, 2005; Clark, Veltmeyer, & Hamilton, 2004; Hermens, Eleonore, & Simon, 2005), but databases for ERP are still scarce. Even aspects such as how ERP raw data as well as the textual information can be structured and standardized are lacking. The purpose of this project is to develop an online standardized relational database of ERP data along with textual information about descriptions and interpretations. This database will also be used for collection of ERP data.

## 2 DESIGN AND IMPLEMENTATION

The ERPDB Database has been developed and implemented using the SQL Server 2000 relational database system. Having specially extended full support to XML (Extensible Markup Language), SQL Server 2000 has the flexibility, accessibility, and security functions required by data storage components of web sites. Because of its unparalleled advantage, SQL Server 2000 has already established a leading position as an integrated solution for database and data analysis (Yang, 2004).

An IIS web server has been set up for data retrieval through a custom web interface using ASP and VBScript language. Our server is running on the Windows 2000 Advanced Server operating system rather than other operating systems such as Unix or Linux. First, the SQL Server 2000 must be based on a Windows Server Series operating system. In addition, any other platforms to store and analyze data are based on Windows, such as Microsoft Excel, Compaq Visual Fortran, Matlab, Microsoft Visual Studio, and so on. The tools and integrated development environment operated on Windows are more available, compared to similar products operated on Unix or Linux. Thus it is more convenient and efficient for biology or medicine researchers not experienced in information technology.

The ERP datasets used to populate this database are in usual textual style, e.g. part of the data is provided by Dr. Sue Surguy, King’s College, University of London. A program has been developed to read those textual data and convert them into the database with the standardized relational structure. The program is developed with the VB6 language and connects to the database by ADO techniques. As an object-oriented visualization-programming tool, Microsoft Visual Basic is easy to learn, flexible, and convenient. Microsoft ActiveX Data Objects (ADO), as the COM interface based on OLE DB, is one of the most important application interfaces (API). Through the encapsulation to OLE DB API, ADO provides consistent and rapid data access, supporting various developmental needs, including the establishment of applications, tools, phylum, database clients, and business objects of the Internet. The main advantage of ADO is that it is simple and fast, requiring relatively less memory and less disk space (Xia & Fanxing, 2005). The program can be used for any other ERP raw data to convert them into the database. Since ADO can be used not only in connecting SQL Server database, but also in connecting databases based on many other database management systems, this program can retrieve data from many other databases and convert them into one database. This is important for data transferring or data sharing between different information centers.

## 3 DATA CONTENT

The ERPDB database is divided into two parts: (1) subjects' primary and record information and (2) ERP data, which has been structured and standardized in a database table, supplemented with unrestricted text files. Three tables are designed for the data.

Table *SubjectsInfo* stores the information of every subject in ERP recording experiments including person’s name, type of the operation (e.g. words, thoughts, sounds, or other stimuli), date and time of the experiment, sampling frequency, total number of channels, data length, and so on. The database has assigned a unique number, called *SubjectID*, to identify every subject in a recording experiment. The *SubjectID* is set as the primary key. With the help of *SubjectID*, it won't be a problem even when different persons have the same name, or one person has done several different records, as the Name field is not a primary key so that it can be the same for different records.

Table *ChannelInfo* is used to store the information about each channel. Every kind of ERP channel has also been assigned a unique number, called *ChannelID*. The records in the table are used globally for all subjects in the database, including channel name, channel label, and other information on different channels. Tables for raw ERP data storage never use a text label to identify different channels but use the numerical *ChannelID* instead. The *ChannelID* is set as the primary key of the table.

Table *EpochData* is the one to store ERP raw data. A novel database structure rather than a free-text style has been used. It is for storing ERP raw data only, without any information on the subjects or the channels. It has only four fields: *SubjectID*, *ChannelID*, Position and Data. The *SubjectID* and *ChannelID* are foreign keys referring to the primary key of the table *SubjectsInfo* and the table *ChannelInfo*, respectively. The field Position is the position on the time zone in channels and the data field is the data value. Here *SampleID*, *ChannelID* and Position altogether are set as Primary Key. Each record indicates a point located at the [*Position*] of the channel [*ChannelID*] of the sample [*SubjectID*], with the value [*Data*]. Users can get the points [*Records*] of any given sample or channel within any given position range they wanted ordered by the field [*Position*]. Then simply traversing the recordset from the first to the end will retrieve the specific range of EEG raw data.

Such a structure has transformed the curve diagram of an ERP signal into a standard table in a relational database and has related the data value to the standard tables of subjects' information. Because each point of ERP raw data is set as the smallest unit, it has enhanced the flexibility of data retrieval and data analysis from the database to the most. This structure enables users to retrieve ERP data readily within any range wanted. In contrast to the free-text style, in which one must get a whole dataset and separate a range by oneself, this database requires less data transfer, providing a faster way and more searching options. Furthermore, it avoids potential free-text errors or uncertain factors, as the database uses a binary style for data storage (while there may be some invalid characters in a long unrestricted text).

The ERPDB database also provides unrestricted text files as supplements to Table *EpochData*.

## 4 QUERY INTERFACE

A query interface to the data has been implemented using an IIS web server and ASP and VBScript langrage (http://202.113.232.103:8088/erpdb/index.asp). All online resources of the database are freely available to the scientific community. Simple searches may be performed on all of the data.

The database provides an interface for users to retrieve the information of each subject in the database, such as the *SubjectID*, task type, channel number, sampling frequencies, epoch dots, repeat times, and so on. All the fields can be empty as to retrieve all the subjects in the database. To protect the privacy of the subjects, only special users can search for their personal name. There are text-files links in the result form; one can get the ERP raw data in textual style as supplements by simply clicking the link.

The web server also provides a web page for data retrieving based on the *EpochData* table. It enables users to get a specific range of any ERP raw data given the *SubjectID* and *ChannelID* rather than only download the entire unrestricted textual file mentioned above. One can input the *SubjectID* or *ChannelID* or select those items from the result list form and give a position range on the time zone wanted. A precise raw ERP data will be returned. It is useful to quickly retrieve a specific part of the data for study and analysis, as the whole data file is usually big and time consuming to transfer through the Internet.

In addition, some basic information and knowledge about ERP is provided for ERP beginners online.

## 5 CONCLUSIONS

The ERPDB database is an Internet-accessible relational database that stores the samples of ERP data with the subject and record information. It has structured and standardized the mass of ERP data. It also has a program interface to recruit new ERP data into the database and exchange data with other databases. As mentioned before, because of the utilization of a novel structure to organize the ERP raw data and the creation of relations between tables, the database is rigorous, precise, and accurate, with more search options, rather than if in common free-text style.

Our lab has collected some original data into ERPDB, but there are still open issues ahead. More and more ERP data will be collected into the database, and the program and the website will be improved. With the database developing, we anticipate it become a valuable tool that would make contribution to the everyday clinical practice, teaching, and research work for neuroscience and psychology in the future.

## 6 REFERENCES

Aurlien, H, Gjerde, I., Gilhus, N., et al. (1999) A new way of building a database of EEG findings. *Clinical Neurophysiology 110*: 986-995.

Babiloni, C., Babiloni, F., Carducci, F., et al. (2001) Quantitative EEG/MEG Analysis for Objective Assessment of Alzheimer's Disease: The Project 'Alzheimer Database On-line.' *NemoImage 13(6):* S770.

Brickman, A., Paul, R., & Cohen, R. (2005) Category and letter verbal fluency across the adult lifespan: relationship to EEG theta power. *Archives of Clinical Neuropsychology 20*: 561–573.

Clark, C., Veltmeyer, M., & Hamilton, R. (2004) Spontaneous alpha peak frequency predicts working memory performance across the age span. *International Journal of Psychophysiology 53*: 1–9.

David, F. & Ray, J. (2000) Event-Related Potential (ERP) Studies of Memory Encoding and Retrieval: A Selective Review. *Microscopy Research and Technique 51*: 6-28.

Finnerup, N., Fuglsang-Frederiksen, A, Rossel, P., et al. (1999) A computer-based information system for epilepsy and electroencephalography. *International Journal of Medical Informatics 55*: 127-134.

Hermens, D., Eleonore, X., & Simon, D. (2005) Resting EEG Theta Activity Predicts Cognitive Performance in Attention-Deficit Hyperactivity Disorder. *Pediatric Neurology 32(4):* 248-256.

Hermens, D., Michael, R., & Simon, D. (2005) Sex differences in adolescent ADHD: findings from concurrent EEG and EDA. *Clinical Neurophysiology 116*: 1455–1463.

Johnson, R. (1998) Event-related brain potentials. In: Litvan, I & Agid, Y. editors. *Progressive Supranuclear Palsy: Clinical and Research Approaches*. New York: Oxford University Press.

Katada, E., Sato, K., Ojika, K., et al. (2004) Cognitive event-related potentials: useful clinical information in Alzheimer's disease. *Curr. Alzheimer's Res 1(1)*, 63-69.

Picton, T. & Hillyard, S. (1988) Endogenous event-related potentials. In: Picton, T., editor. *Handbook of electroencephalography and clinical neurophysiology, Vol. 3. Human event-related potentials.* Amsterdam: Elsevier.

The Brain Resource Company: First Global Standard for Cognitive and Brain Testing. Retrieved from the WWW October 9, 2007: http://www.brainresource.com/

Xia, B. & Fanxing, L. (2005) *SQL Server Database Development Training Guide in Chinese Version*. Beijing: People's Posts and Telecommunications Publishing House.

Yang, D. (2004) *SQL Server Database System*. Beijing: Publishing House of Mechanic Industry.