

USING MATLAB FOR LARGE DATA IMPORTS OF INTRACRANIAL ELECTROENCEFALOGRAMS

P. Ježdík¹, L. Šenfeld¹, F. Lehr¹, T. Havel¹, J. Balach¹, A. Jahodová²

¹ Faculty of Electrical Engineering, CTU in Prague,

² 2nd Medical School, Charles University in Prague

Abstract

Various EEG recording systems are currently used in neurological departments. Due to the use of mostly commercial products for data acquisition and visualization by neurologists, the data storage formats or structures are not open. Some of the common software allows at least for export to text i.e. ASCII formats. The problem for further processing by tools such as Matlab is the very large data flow and memory demands for the ASCII data imports. The paper demonstrates the possibility of using a combination of a stream editor such as SED and the Matlab importer for extracting the useful information from the text export of one of the commercial EEG recordings software.

1 Introduction

The technical branch of the Intracranial Signal Analysis Research Group in Prague (ISARG) is responsible for supplying technical support to the researchers. One of the most important tasks for the technicians is not only reliable recordings exchange between the medical and technical parts of the group, but also data exchange into the technical branch. Here, the crucial step is the import, sorting and distribution of the raw records to the researchers. The technical researchers mostly use the Matlab environment for signal evaluation, though a few studies are also done using the Python platform. The one chosen for the internal data exchange format is the MATLAB data format.

The typical data recorded by the system in the neurology department is a hundred channels at 1kHz sampling frequency at 12bits resolution. The ASCII file of recordings of ten minutes, as the typical configuration, needs approximately 0.5GB of space. In comparison with the MAT format, which for the same information needs approximately one third of the original size, the ASCII is really demanding, especially when the typical recording durations used are about half an hour.

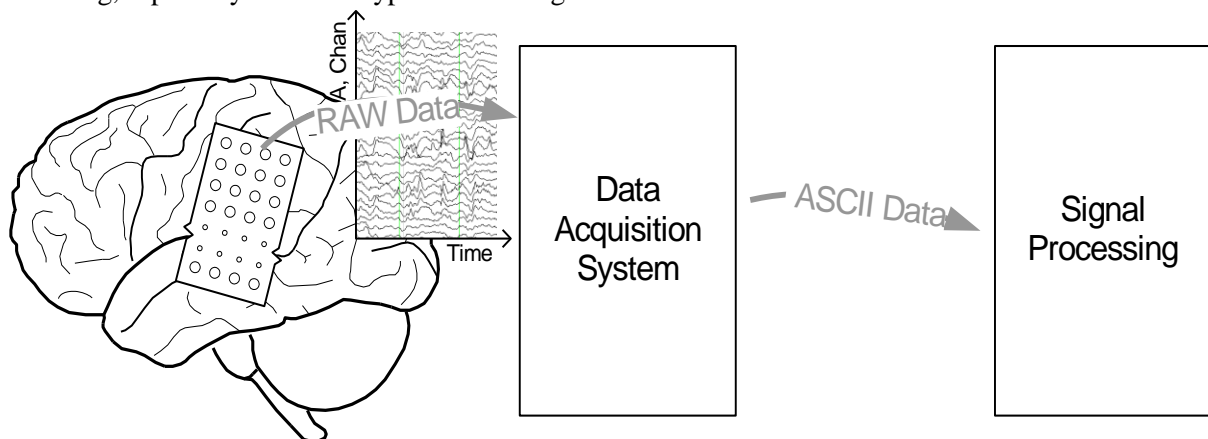


Figure 1. Data acquisition: block diagram.

2 Algorithm

The cooperative neurology center uses the commercial data acquisition, visualization and database system Harmonie. Every row in the ASCII data export file from the Harmonie system consists of a timestamp of every sample and a sequence of amplitudes of each connected channel one by one. The exporter uses colons, semicolons and dashes as separators of the data columns. The problem is that the Matlab ASCII load command expects only dashes or fixed spacing separators for

automatic imports, see the example in Table1. The necessity of colon and semicolon replacement appears as well as the header block separation.

Header block

```
"File name", "Patient signal name.sig"  
"Montage", "66 el"  
"Patient name", "XXXX"  
"Recorded at", "Neurologic Center of a Hospital"  
"Date recorded", 201x-0x-2x  
"Time recorded", 12:29:33  
"Channels information:"  
"#1 ah1,EEG,uV,200 Hz"  
"Filter1: None"  
"#2 ah2,EEG,uV,200 Hz"  
"Filter1: None"  
...etc
```

Data block

```
12:30:01.565,16.825083,57.692818,64.715439,-64.227638,99.535919,36.722500,-  
98.657974,-58.570526,51.352955,13.313773,182.246750,-32.112946,111.045212,9.314782,-  
59.253281,0.048828,21.896973 ..... etc.
```

Using conventional editors such as Windows Notepad or programmer's notepad, there emerge problems with memory demands. The file, mostly more than one GBs, can be stored into the memory but the performance, when the computer belongs to a standard class, rapidly decreases. Also, when 32-bit Windows are used, the maximum storage memory due to the address space is 3GB per application.

A stream editor can also be used, derived from Linux platforms. Using the external program SED [5] with adequate parameters, the text file can be modified to make it sufficient for the Matlab load script. The advantage of the stream editor is that such operations as finding characters, their replacement and/or adding new characters is performed during streaming of the file, which is not memory demanding compared to the standard text editors. The processing time is then sufficient and the system more stable.

The usage of the stream editor through Matlab script is as follows.

```
tmp=sprintf('sed -e \x22/d;s/:/,/g %s.in>%s.out',filename_in,filename_out);  
dos(tmp,'-echo')
```

The SED switch `-e` with their additional parameters says "delete all rows containing quotation marks, globally find colons and replace them with dashes". After this sequence, the output file is fully ready for importing by the Matlab load. The load command imports a matrix, where the columns are equal to the following sequence:

Hours, minutes, seconds, channel1, channel2, channel3, ...etc...

The rows are equal to the samples of the recordings. The *.mat structure of the data is then exported to another data format sufficient for partial research steps. The other formats used in the ISARG group are generally known or at least free biosignal data formats such as European Data Format (EDF, EDF+), Global Data Format (GDF) and so on. [3], [1].

The sole limitation for the explained procedure is the address space of the Matlab and the platform where it runs. Using the 32bit version and Win XP, the limitation is 2GB, or respectively 3GB with -3G system switch. Using the 64bit version of Matlab, the practical limit depends on the technical equipment of the system where Matlab is running.

3 Summary

The concept that we use saves time for preparation of raw data, its distribution and presentation to the technical part of the research group. We also appreciate the high stability and robustness of the

special character replacement as compared to conventional text editors. We also apply additional methods due to the advantages of open data format, for instance, conversions to other open EEG data formats such as GDF, EDF, etc. [3], [1] and their processing or presentation in relevant viewers such as SigViewer [2] or EDFViewer [4]. We also used a similar procedure for large and automatic data import from data acquisition systems in different fields of interest in our department, see [8] and [9]. The described procedure is an example for which the standard class of personal computer performance is sufficient for the partial steps of quite demanding research computing as well as our distributed computing system as reported in [7].

4 Acknowledgement

This work has been supported by the grants IGA NT11460-4/2010 Intracranial EEG signal processing; epileptogenic zone identification in non-lesional refractory epilepsy patients, SGS 10/272/OHK4/3T/13 Analysis of intracranial EEG recording, and research program MSM6840770012 Transdisciplinary Research in Biomedical Engineering.

5 References

- [1] <http://biosig.sourceforge.net/links.html>
- [2] <http://sigviewer.sourceforge.net/>
- [3] <http://www.teuniz.net/edfbrowser/edf%20format%20description.html>
- [4] <http://www.edfplus.info/downloads/downloads.html>
- [5] <http://www.grymoire.com/Unix/Sed.html>
- [6] http://www.natus.com/index.cfm?page=company_1&crd=418
- [7] Jezdik, P.; Cmejla, R.; Krsek, P. & Jahodova, A. (2010), *Distributed computing system for EEG signal processing using MATLAB as activex object in DCOM model*
- [8] Kubínyi, M. - Kreibich, O. - Neužil, J. - Šmíd, R. (2011), *Novel S-Transform Information Fusion for Filtering Ultrasonic Pulse-Echo Signals*
- [9] Fexa, P., Vedral, J., Svatos, J.(2011), *DAC Testing Using Modulated Signals*

Ing. Petr Ježdík, Ph.D.
e-mail: jezdip1@fel.cvut.cz
Lukáš Šenfeld
e-mail: senfeluk@feld.cvut.cz
František Lehr
e-mail: lehrfran@feld.cvut.cz
Tomáš Havel
e-mail: havelto3@feld.cvut.cz
Jiří Balach
e-mail: balacjir@feld.cvut.cz
MUDr. Alena Jahodová
e-mail: a.jagoda@email.cz