

Sistema de Archivos EDF/EDF+ utilizando MATLAB

Álvarez Surci Américo
aalvarez6@umsa.bo

ABSTRACT

Se analiza el sistema de datos Europeo de intercambio de información fisiológica humana (EEG, EMG, MEG, etc.) además se realiza una herramienta en Matlab que nos pueda permitir utilizar las bases de datos de institutos de investigación internacional como el MIT, para poder desarrollar investigación local en temas de potenciales eléctricos biológicos.

I. INTRODUCCIÓN

El formato de datos Europeos (EDF - European Data Format) es uno de los más utilizados en equipos de registro de potenciales eléctricos biológicos. El formato EDF inicialmente fue diseñado para intercambiar registros electro-encefalograficos (EEG) de sueño [1] entre diferentes institutos y centros de investigación.

Desde 1992 con la aparición del formato EDF, equipos EEG-digitales presentan en sus características este formato para almacenar registros EEG.

En 2003 con la mejora a EDF+ [2] se puede soportar registro de señales (EDF), anotaciones, estímulos realizados, acontecimientos, análisis de resultados. Con estos agregados los archivos EDF+ pueden almacenar cualquier grabación médica como EMG, ECG, EOG¹.

Registros EEG de: sueño, patologías, o algún otro registro EEG es compartido por instituciones médicas o institutos de investigación de todo el mundo en formato EDF/EDF+ para que pueda apoyar en estudios y desarrollo de nuevas tecnologías con BCI.

En este trabajo se presenta en primera parte el estudio del formato EDF/EDF+, seguidamente se analiza la estructura de este protocolo, se continuara con el desarrollo de una herramienta que nos permita trabajar con este tipo de archivos extrayendo toda la información útil en el Workspace de MatLab y para terminar se utilizará la base de datos del CHB-MIT [3] para verificar el correcto funcionamiento de la herramienta desarrollada.

¹Electromiografía (EMG), Electrocardiograma (ECG), Electrooculograma (EOG)

II. FORMATO EDF

El formato EDF fue considerado por primera vez en Copenhague en el "5th International Congress of Sleep Research" en 1987, la motivación principal de un grupo de ingenieros biomédicos en dicho congreso fue la de tener un formato abierto para almacenar e intercambiar registros EEG de larga duración.

El desarrollo del formato EDF fue entre 1989-1992 en el programa "Methodology for the Analysis of the Sleep-Wakefulness Continuum" financiado por la Comunidad Europea a través de su "Comité d'Action Concertee" (COMAC-BME).

La finalidad era que laboratorios e institutos de investigación puedan aplicar sus algoritmos de análisis de sueño y poder comparar resultados obtenidos con sus similares geográficamente distantes, en Leiden (Marzo de 1990) se acordó un formato simple para intercambiar registros EEG de sueño, este formato se hizo conocido como el **Formato de Datos Europeo EDF**, en agosto de 1990 todos los laboratorios participantes ya habían aportado una grabación de sueño EDF al proyecto.

El formato EDF se publicó en 1992 [1] desde entonces se convierte en el estándar para grabaciones de EEG y PSG² en equipos comerciales y proyectos de investigación.

Puesto que se utiliza ampliamente el formato EDF para registros de EEG y grabaciones de sueño, este formato no ofrecía cabida a otras señales utilizadas en laboratorios de Neurofisiología Clínica, teniendo este objetivo pendiente. Esta fue la razón para crear el formato EDF+.

El formato EDF+ fue desarrollado en 2002, y publicado en 2003 en Elsevier. Teniendo grandes mejoras a comparación del formato EDF, pudiendo soportar registros biológicos eléctricos y otros varios obtenidos por transductores. Las señales guardadas en EDF+ generalmente son: EEG, ECG, EOG, ERG, EMG, NC, MEG, MCG, EP, temperatura, Luz, Sonido.

EDF+ también puede almacenar eventos y anotaciones únicamente sin ninguna señal. Esta flexibilidad permite al usuario poder registrar varios acontecimientos y anotaciones en un único archivo EDF+, esta es la razón de la diferencia entre archivos EDF y EDF+. El proto-

²Polisomnografía (PSG)

colo EDF+ permite el almacenamiento de varias grabaciones No Contiguas en un solo archivo mientras que un archivo EDF almacena solo grabaciones ininterrumpidas.

Debido a que EDF+ y EDF son muy similares, el software para EDF+ que se pudiera utilizar en lectura, escritura y edición puede ser desarrollado con relativa facilidad basado en software EDF.

Software y utilidades para EDF y EDF+ están disponibles libremente y sin costo alguno en Internet, por ser un formato de codificación de información para intercambio y transporte.

III. ESTRUCTURA EDF

El fichero EDF puede contener un registro poligráfico ininterrumpido multicanal. Los ficheros en la primera parte poseen un registro cabecera de longitud variable que identifica al paciente y las características técnicas de las señales registradas, en segunda parte están los registros de las señales EEG captadas por el equipo médico [4] [5].

A. Cabecera

La cabecera esta en formato ASCII y proporciona información general sobre la configuración de la adquisición y el registro de los datos figura 1. La información proporcionada por esta cabecera es la identificación del paciente, fecha y hora de comienzo, el número y duración de los datos registrados, el número de señales y la etiqueta individual de cada canal.

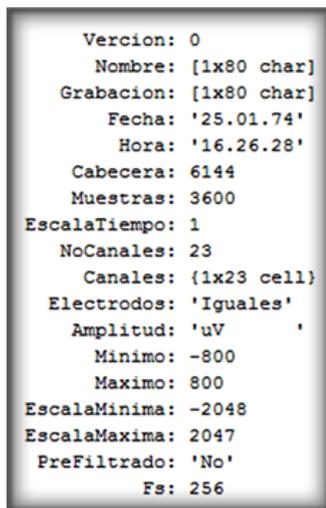


Figura 1: Datos presentados en la cabecera

El tamaño de la cabecera es de 256 bytes seguido de otro bloque o encabezado variable de 256 bytes para cada canal del registro figura 2. En cada cabecera de bloque de cada canal se especifican atributos como el tipo de señal, calibración, y el número de muestras por registro. Por lo

tanto el tamaño final del encabezado es de $256+(ns*256)$ bytes.



Figura 2: Estructura del encabezado y datos del formato EDF

La información de las cadenas de caracteres ASCII empieza a la izquierda y es rellenado con espacios si no hubiera carga útil. Se recomienda además que la duración de cada registro de datos no exceda 61440 segundos (17.1 horas), de esta manera será correcta la visualización de la cabecera.

En la cabecera podemos observar dos tipos de información:

1) Encabezado global:

Presenta toda la información referente al equipo y paciente atendido.

| Nombre | Función | Tamaño(bytes) | Pauta |
|--------------|---|---------------|----------|
| Versión | Versión del formato de datos | 8 | 0 |
| Nombre | Nombre del paciente al cual pertenece el EEG | 80 | 0 |
| Grabación | Código o nombre de identificación de la grabación | 80 | |
| Fecha | Presenta la fecha de la grabación | 8 | dd.mm.aa |
| Hora | Presenta la hora de inicio de grabación | 8 | hh.mm.ss |
| Encabezado | Número de bytes en registro de Cabecera | 8 | |
| Reservado | Espacio reservado | 44 | |
| Muestras | El número de datos o muestras tomadas(-1 si no se conoce) | 8 | |
| Duración | Muestra la duración y/o escala de tiempo | 8 | |
| Nro. Canales | Número de canales guardados(ns) | 4 | |

2) Cabecera de la(s) señal(es):

Presenta la información referente a cada señal adquirida con el equipo poligráfico.

| Nombre | Función | Tamaño(bytes) | Ejemplo |
|------------------------|--|---------------|-------------------------------|
| Canales | Presenta la identificación (nombre) de todos los canales | 16*ns | Fpz-Cz, temperatura |
| Electrodos | Presenta el tipo del transductor utilizado en cada canal | 80*ns | electrodos AgCl, electrodo Au |
| Escala Amplitud | Escala de datos de cada señal. | 8*ns | $\mu V, mV, V, degree C$ |
| Mínima Medición | Mínima Medición física posible | 8*ns | -1200, -500, 37 |
| Máxima Medición | Máxima Medición física posible | 8*ns | 1200, 500, 40 |
| Mínima Escala | Amplitud Mínima Digital de Medición | 8*ns | -5500, -2500, 0 |
| Máxima Escala | Amplitud Máxima Digital de Medición | 8*ns | 5500, 2500, 100 |
| Pre Filtrado | Si se utilizo un filtrado al tomar la Medición | 80*ns | HP=0,1 Hz ; LP=75 Hz |
| Muestreo | Presenta la Frecuencia de Muestreo | 8*ns | |
| Reservado | Espacio reservado | 32*ns | |

B. Datos

Los registros poligráficos que muestran la actividad eléctrica del encéfalo se almacena en este segmento, tiene la característica de que cada registro de cada canal se almacena en forma intercalada, teniendo una longitud variable dependiendo del número de canales que almacena el archivo EDF.

Para realizar la correcta lectura del registro poligráfico almacenado, se debe de considerar el factor de escala aplicado considerando la mínima y máxima amplificación física o digital, para realizar un correcto escalado se debe de acudir a datos del encabezado como las medidas máxima y mínima además de la amplificación digital. Además se considera el *Offset* incluido, siguiendo para estos puntos las siguientes fórmulas:

$$\text{Factor de Escala} = \frac{\text{Máxima Medición} - \text{Mínima Medición}}{\text{Escalado Máximo} - \text{Escalado Mínimo}}$$

$$\text{Offset} = \text{Máxima Medición} - \text{Factor de Escala} * \text{Escalado Máximo}$$

Considerando el factor de escala y el nivel DC añadido a cada medición, los datos resultantes serán adaptados siguiendo esta fórmula:

$$\text{Muestra Real} = \text{Muestra Almacenada} * \text{Factor de Escala} + \text{Offset}$$

IV. PROGRAMA EN MATLAB

Se presenta el código que puede extraer la carga útil de los archivos EDF, es una rutina necesaria cuando se desea trabajar con bases de datos de grabaciones EEG [3] de centros de investigación disponibles en Internet.

El programa realizado muestra el encabezado; figura 1, con la información necesaria del paciente y de las señales

almacenadas en el archivo EDF, además se obtiene la matriz que contiene todas las muestras almacenadas en el archivo EDF, siendo las dimensiones de la matriz:

$$[\text{número de canales} \times (\text{Muestras} * F_s)]$$

V. PRUEBAS Y MEDICIONES EN MATLAB

Se trabaja con el archivo `chb24_01.edf` [3] para la realizar todas las pruebas presentadas en esta sección.

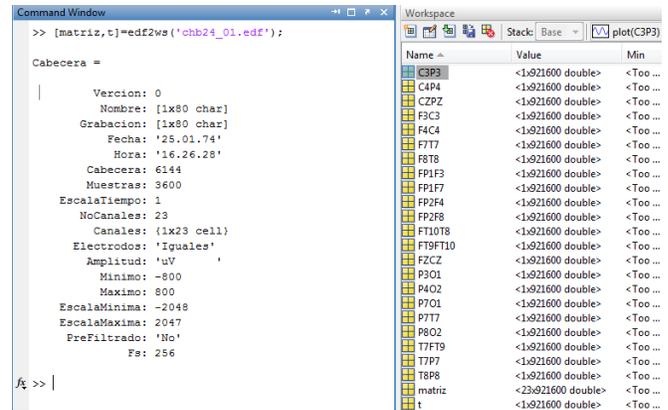


Figura 3: Cabecera y datos, obtenidos del archivo CHB24_01.EDF

Como se aprecia en la figura 3, se pueden trabajar con canales individuales o con la matriz que contiene todos los canales, teniendo estos datos en el *WorkSpace* de *MatLab* se puede realizar todo el análisis matemático que se desee.

VI. CONCLUSIONES Y MEJORAS

Como se observa tan solo se trabajo en archivos EDF, una mejora al programa sería el de poder observar las anotaciones incluidas en archivos EDF+.

Aunque los archivos EDF son un estándar actual en diferentes registros de bio-señales, el archivo presentado se lo probó tan solo en archivos EEG y no así en otra gama de señales, un siguiente paso sería mejorar el programa y que pueda dar lectura a diferentes archivos poligráficos sea cualquiera sus datos almacenados.

REFERENCIAS

- [1] Agostinho C. Rosa Kim D. Nielsen Bob Kemp Alpo Varri y John Gade. "A simple format for exchange of digitized polygraphic recordings". En: *Electroencephalography and Clinical Neurophysiology*. Ed. por Elsevier. Vol. 82. 1992, págs. 391-393 (vid. pág. 1).

- [2] Bob Kemp y Jesus Olivan. “European data format ‘plus’ (EDF+), an EDF alike standard format for the exchange of physiological data”. En: *Clinical Neurophysiology*. Vol. 114. 2003, págs. 1755-1761 (vid. pág. 1).
- [3] PhysioNet. *Children’s Hospital Boston (CHB) and the Massachusetts Institute of Technology (MIT)*. Ed. por MIT. URL: <http://www.physionet.org/physiobank/database/> (vid. págs. 1, 3).
- [4] Bob Kemp. *European Data Format*. 2003. URL: <http://www.edfplus.info> (vid. pág. 2).
- [5] Jesus Olivan Palacios. *Some introductory notes to EDF+*. 2003. URL: <http://www.neurotraces.com> (vid. pág. 2).