

Detección de crisis epilépticas partir de señales EEG mediante índices basados en el algoritmo de Lempel-Ziv

Iván Mora López-Tercero

Grado en ingeniería en electrónica industrial y automática, Escola Politècnica Superior Enginyeria Vilanova i la Geltrú, Universitat Politècnica de Catalunya.

Resumen

Hoy en día hay 50 millones de personas en el mundo afectadas por la epilepsia, en España se calcula que la padecen cerca de 400.000 personas y cada año aparecen 20.000 casos nuevos. A la mayoría de los enfermos se les puede tratar con fármacos, pero hay un 30%, es decir, 15 millones de personas en el mundo que la sufren de forma intratable, no existe ningún fármaco que pueda evitar ni atenuar las crisis.

Este proyecto desarrolla una metodología para la detección de las crisis.

Para poder realizar el estudio se ha utilizado una base de datos, la cual se ha registrado utilizando la técnica del Electroencefalograma (EEG).

La base de datos está tomada en el Hospital Infantil de Boston, se compone de registros EEG de pediatría, como se ha citado líneas anteriores, todos los pacientes que se estudian sufren epilepsia intratable.

Para detectar las crisis se van a utilizar tres técnicas distintas.

La complejidad de Lempel-Ziv (LZ) es un método que logra dimensionar estadísticamente las interacciones dinámicas neuronales a través del análisis de las señales cerebrales. La de Lempel-Ziv Multiestado tiene la misma finalidad pero introduce el parámetro de amplitud. El ARV es el valor medio rectificado, para rectificarlo se hace uso de una fórmula.

Con estos tres algoritmos se estudia la señal EEG original obtenida en la base de datos, para poder realizarlo, se requiere un proceso de filtrado. Una vez obtenidas las señales, se realizan varios acondicionamientos de la señal para obtener un umbral que determinará a partir de qué momento ocurre una crisis y cuando acaba, no obstante el fin del proyecto no es medir el tiempo de las crisis sino detectar cuantas ocurren.

Como último se obtienen los resultados de los tres estudios, en los que se obtiene la conclusión que el método más preciso es el de ARV seguido muy de cerca por el LZM.

1. Introducción

1.1 Epilepsia

La epilepsia está provocada por anomalías en la actividad eléctrica del cerebro. Este órgano es incapaz de frenar o inhibir los impulsos eléctricos entre neuronas. Cuando tiene lugar una descarga excesiva se produce una crisis o ataque epiléptico, estas descargas son las que se van a estudiar en el proyecto mediante el EEG.

A finales del siglo XIX, el neurólogo inglés John Hugling Jackson, estableció la definición de epilepsia que permanece vigente hoy en día: "una descarga súbita, rápida y excesiva de las células cerebrales".

1.2 Señal EEG

La Electroencefalografía es el registro y evaluación de los potenciales eléctricos generados por el cerebro y obtenidos por medio de electrodos situados sobre la superficie del cuero cabelludo.

Para obtener la base de datos utilizada en el presente proyecto, se han colocado los electrodos siguiendo la distribución que marca el sistema internacional.

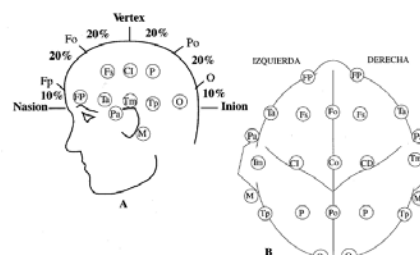


Figura1. Sistema internacional 10-20

2. Materiales y Métodos

2.1 Base de Datos

La base de datos está tomada en el Hospital Infantil de Boston, se compone de registros EEG de pediatría, los pacientes sufren una epilepsia intratable. Estos fueron monitorizados durante varios días después de retirarles la medicación, con el fin de clasificar sus crisis y evaluar su posible candidatura a la intervención quirúrgica.

Esta consta de 23 casos, tomados de 22 pacientes, 5 hombres de edades comprendidas entre 3 y 22 años y 17 mujeres de edades entre 1,5 y 19 años.

Los registros de señales tienen una duración media de una hora.

Todas las señales se muestrean a 256 muestras por segundo con una resolución de 16 bits. La mayoría de los archivos contienen 23 señales, excepto algunos pocos que constan de 24 o 26, el total de ataques registrados asciende a 198.

La adquisición de la señal se ha llevado a cabo colocando los electrodos EEG siguiendo el sistema internacional 10-20.

En la figura 2 muestra, las señales de los 23 de uno de los pacientes, tal como se obtienen de la base de datos.

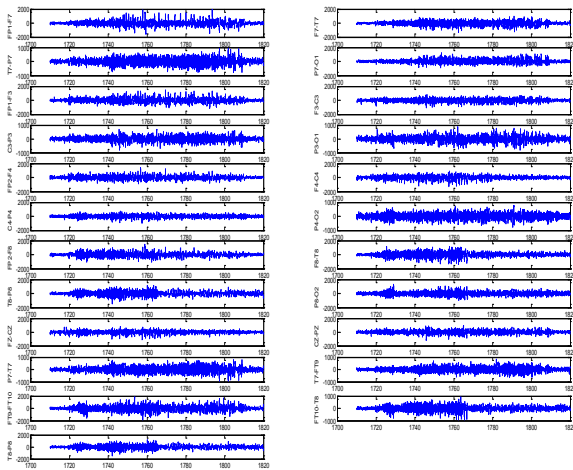


Figura2. Visualización de 23 canales con crisis

2.2 Técnicas

El primer método propuesto es análisis de la complejidad LZ, el cual se basa en la transformación de la señal en una secuencia P formada por un número finito de símbolos. Esta secuencia P se examina de izquierda a derecha, y un contador de complejidad $c(n)$ se incrementa en una unidad cada vez que se

encuentra una nueva subsecuencia de caracteres consecutivos. Dicho contador refleja la tasa de aparición de nuevos patrones a lo largo de la secuencia P. Por lo tanto, antes de calcular $c(n)$, la señal ha de convertirse en esa secuencia con un número finito de símbolos. Diversos estudios han mostrado que una conversión binaria (dos símbolos: 0- 1) puede ser suficiente para caracterizar adecuadamente la señal.

El segundo método aplicado es el LZ multiesado, el cual consiste en lo mismo pero añadiendo más líneas de cambio, es decir, en el primero los símbolos 0-1 cambian al pasar de un estado a otro, del inferior al superior se convertía en un uno y al pasar del superior al inferior, se convertía en un cero, en el LZM es lo mismo, pero añadiendo más estados, así que los símbolos cambiarán con más frecuencia con la consecuencia de la obtención datos de la amplitud de la señal.

El tercer método a poner en prácticas es el ARV, es una de las técnicas de procesamiento de señales más utilizadas, la mejor aportación que hace este método al proyecto, es que su coeficiente de variación es menor al de otras técnicas similares. Su cálculo se hace con la siguiente fórmula.

$$ARV = \frac{\sum_{i=1}^N |x_i|}{N}$$

2.3 Algoritmo de detección

El algoritmo propuesto para la detección de ataques epilépticos se basa en la determinación de la energía y la duración de los registros, por cada canal y paciente. Hay que tener en cuenta que cada canal tiene una impedancia debido a la distancia del mismo con el punto concreto en el que se desarrolla la crisis, así como la longitud del cable, el estado del electrodo.... Este factor hace variar el resultado de la energía, aunque no se deja de apreciar su incremento de valor y de complejidad, más adelante se explica cual es el método utilizado para la determinación del umbral, el cual se ve basa en el incremento del valor de la energía.

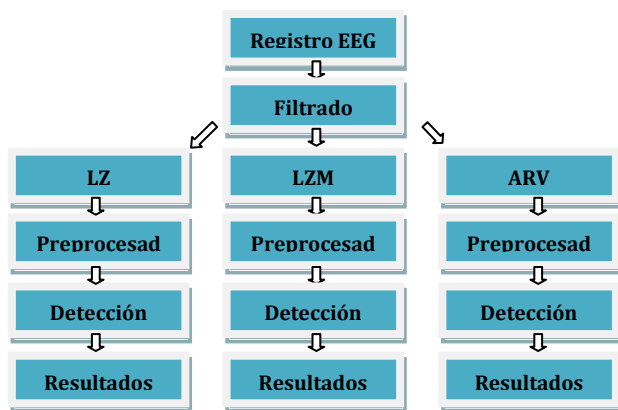


Figura 3. Diagrama de bloques del algoritmo

2.3.1 Registro EEG

Como se ha explicado en el capítulo 3, se parte de una base de datos de 22 pacientes con 23 casos, ya que el primer paciente repite la prueba. Los archivos son de extensión .m, es decir, para visualizarlos con Matlab. Así que se implementa un programa para hacerlo. Los programas utilizados se explicarán brevemente en apartados posteriores.

2.3.2. Filtrado

La mayoría de señales biomédicas son señales débiles en un ambiente lleno de interferencias procedentes de otras señales de diversa procedencia. Esas señales pueden ser interferencias procedentes de artefactos, o simplemente ruido. La fuente del ruido puede ser fisiológica, instrumental, o interferencias ambientales de la zona experimental.

Por tanto, es de gran importancia realizar un buen preprocesado de la señal, para poder extraer de esta la máxima información posible, sin que se vea corrompida por información no deseable. Dado que el contenido frecuencial de la señal EEG está contenido en la banda entre los 0.5 y 60 Hz, todos los registros EEG fueron filtrados utilizando un filtro tipo Butterworth bidireccional de orden cuatro con dicha banda frecuencial.

3.Resultados

3.1. Obtención de las señales LZ, LZM y ARV

Para obtener las señales Lempel-Ziv, Lempel-Ziv Multiestado y ARV, se ha recurrido a tres programas de Matlab, con los cuales se

realizan las operaciones necesarias explicadas en apartados anteriores, estos programas se muestran en capítulos posteriores.

Para visualizar la crisis se ha desarrollado otro programa con Matlab, con el cual estudiamos cuatro gráficas, en la figura 4 se muestra un ejemplo utilizando la señal chb01_04 en el canal 1, la cual pertenece al cuarto registro del segundo niño.

La primera gráfica es la señal EEG obtenida en la base de datos, en ella, al igual que en las tres señales restantes, hemos dibujado dos ejes verticales, el primero verde, indicando el inicio de la crisis y el rojo indicando el fin. En este caso la crisis se encuentra acotada entre los segundos 1467 y 1494, estos datos están tomados de la tabla 1.

En la segunda gráfica se ha registrado la señal llevando a cabo los cálculos para la complejidad de Lempel-Ziv explicado en capítulos anteriores. Para poder aplicar este algoritmo se ha tenido que utilizar una ventana y un paso, en el siguiente punto se explica exactamente de que se trata, cual es su importancia para el análisis y como se ha determinado su elección.

La tercera gráfica es el algoritmo Lempel-Ziv multiestado, se observa que sin estar aún tratada, esta señal aumenta suavemente de forma constante durante el tiempo de la crisis, posteriormente se observan otros picos pero no mantienen un valor constante, estos picos los vamos a eliminar cuando acondicionemos la señal efectuando cambios en el tiempo de la ventana y su paso.

La última es el ARV en el cual se aprecia una variación muy notable en el punto más álgido de la crisis, pero al igual que en el LZM, cuando acaba la crisis se aprecian otros picos, los cuales también atenuaremos mediante el acondicionamiento de la señal.

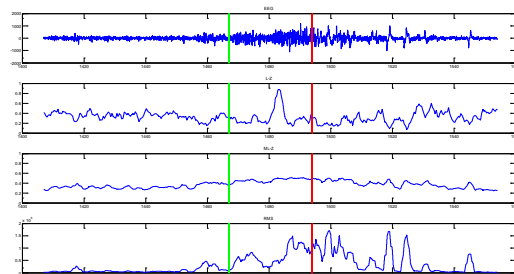


Figura 4. Análisis de la señal

3.2. Preprocesado de las señales

Para procesar la señal de forma correcta hemos utilizado un filtro de ventana móvil, el cual será muy importante para todo el proceso de preprocesado o filtrado de la señal.

Para la obtención de las primeras gráficas, como las de la figura 4, se ha utilizado una ventana y un paso pequeños, para así no modificar la señal, de forma que tome un gran número de muestras para el cálculo de las complejidades y del ARV. Los valores escogidos han sido de 2 segundos para la ventana y 0,2 para el paso. Entre cada paso de 0.2 segundos se encuentran un total de 51 muestras, estas son sometidas a interpolación lineal como se muestra en la figura 4.

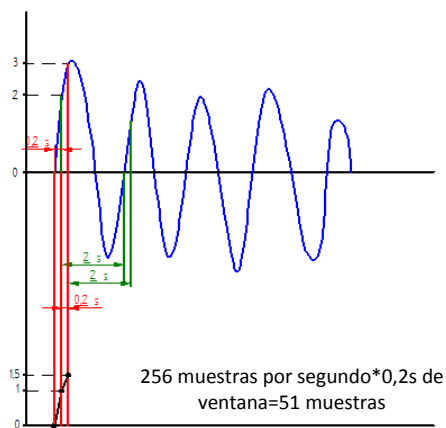


Figura 5. Análisis de la señal

En la figura 5 se muestra el procedimiento para filtrar la señal con una ventana de 2 segundos (en verde) y un paso 0,2 segundos (en rojo). La senoidal resultante se muestra en negro, donde se ven los tres primeros puntos creados, el resto de la señal seguiría el mismo procedimiento. Entre cada nuevo punto creado, se encuentra una línea de unión, esta es la que nace de la interpolación lineal de los puntos conocidos.

La parte más delicada del proyecto es decidir cuál será el umbral que determine si en la señal nos encontramos con una crisis o no. Para ello se hacen algunas pruebas, sacando los valores máximos y medios de las cuatro señales mostradas en la figura 3. Se procede para varios registros y se calculan las medias y medianas. Lo hacemos primero para señales con crisis y posteriormente para señales sin ellas. Los valores resultan muy dispares ya que son señales con muchas variaciones, tanto las que pertenecen a registros con crisis como las que no.

Para solucionar este problema se decide modificar los valores de las ventanas y el paso y así quedarán señales más suaves y con menos irregularidades.

En la siguiente figura se muestra la evolución de la señal según los cambios de la ventana móvil y el paso. En las ventanas de LZ, LZM y ARV, se encuentran tres señales, la primera es la azul, para la cual se utilizó una ventana de 2 segundos con un paso de 0,2. La segunda es la verde, para la cual se ha utilizado una ventana de 4 segundos y un paso de 0,5 segundos. Para la señal definitiva (roja) se ha utilizado una ventana de 20 segundos y un paso de 1 segundo.

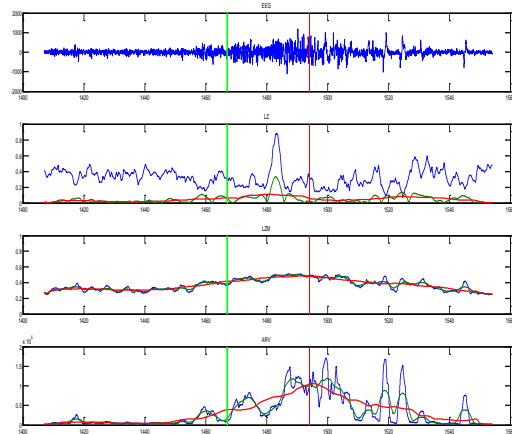


Figura 6. Preprocesado de la señal

Desde la primera señal hasta la última, el valor de la ventana se ha multiplicado por 10 y el del paso por 5, esto promedia la señal cada vez más, pero es necesario para poder hacer un análisis preciso y sin ruidos. La base del proyecto es la detección de crisis y estas se detectan por la variación de la señal, así que promediarla no es un problema, ya que la relación entre las variaciones no va a variar. De todas formas se han ido cambiando los valores con cuidado de no ponerlos demasiado elevados, ya que un exceso provocaría que la señal quedara completamente lineal y las variaciones fueran tan pequeñas que dificultara su detección.

3.3. Detección del umbral

Para llevar a cabo este proceso se ha estudiado las señales de 10 niños con crisis, los casos estudiados son los mostrados en la tabla 1:

Paciente	Registro
1	4
2	19
3	2
3	34
5	13
8	13
10	27
11	92
12	42
14	17

Tabla 1. Pacientes estudiados para decidir el umbral

Con el total de 10 escogidos al azar de la tabla 1 se procede a la determinación del umbral. Para ello, se hace la media del valor de la crisis para cada paciente por sus 23 canales, es decir, se obtiene una media por paciente. Se sigue el mismo proceso para un tramo de señal sin crisis, este tramo se selecciona siempre al principio del registro y el tiempo de duración es igual al de la crisis. Se obtienen los resultados de las 3 tablas.

Para estas tablas se procede a reenumerar a los pacientes por orden, ya que no es relevante su orden real.

Se ha obtenido una media del ratio por paciente, pero al obtener los resultados, se ve claramente que algunos resultados no son correctos, ya que el ratio queda menor de 1, esto es un problema porque la finalidad del ratio es multiplicarlo por la media sin crisis y de esta forma el análisis no resultaría correcto, ya que detectaría un gran número de crisis erróneas. Esto se debe a que los valores obtenidos son muy variables debido a que los 23 canales poseen valores muy dispares entre ellos, dados por la impedancia provocada entre otros motivos, por la distancia entre la focalización de las crisis y los electrodos que la detectan. Por este motivo, se determina una condición, diremos que el paciente está sufriendo una crisis cuando la señal supera el umbral en al menos 5 canales.

Antes de seguir con el estudio, cabe matizar que en el paciente 11_92 se ha observado una desconexión de los electrodos en un tramo de

señal y por ello se ha recurrido a extraerlo del estudio, así que ahora nos encontramos con 9 pacientes.

Los resultados obtenidos son los siguientes para los tres análisis.

	LZ	LZM	ARV
Paciente	Ratio	Ratio	Ratio
1	2,9058	1,5878	4,0369
2	2,1589	1,4632	3,6349
3	1,4915	1,82	4,1599
4	1,9302	1,8761	4,7902
5	0,6745	1,5363	4,3993
6	0,9198	1,044	1,4285
7	1,2485	1,4246	2,2425
8	0,9103	1,1272	1,1736
9	2,4407	1,3204	1,8254

Tabla 2. Ratios obtenidos

Viendo los resultados de las tres tablas, parece evidente que las mejores detecciones se harán con el método de LZM y el ARV, ya que los valores de los ratios de LZ, en muchos son menores de 1.

Una vez obtenidos los ratios correctos, el siguiente paso es multiplicarlo por los valores medios sin crisis para cada canal, los cuales se pueden ver en las tablas 3, 4 y 5. Para evitar ruidos o pequeños picos en la señal, se determina un umbral inferior y otro superior, de un 1% del valor del umbral. En este caso, el problema del LZ será que al multiplicar el valor medio por un valor menor de 1, detectará muchas crisis sin que las haya.

3.4. Resultados de la detección

A continuación se muestran las tablas de resultados de los tres análisis. Las de las columnas significan lo siguiente:

P: Paciente

C.E: Crisis Errónea

C: Canales en los que se encuentra

C.C: Crisis ciertas

LZ				
P	C.E.	C	C.C.	C
1	0	0	0 de 1	0
2	0	0	1 de 1	16
3	3	19	1 de 1	20
4	1	10	1 de 1	9
5	4	23	1 de 1	23
6	2	15	1 de 1	23
		23		
7	2	17	1 de 1	23
		15		
8	3	18	3 de 3	23
		19		23
		18		23
9	0	0	1 de 1	5

Tabla 3. Resultados LZ

LZM				
P	C.E.	C.	C.C	C
1	0	0	0 de 1	3
2	0	0	1 de 1	13
3	0	0	1 de 1	15
4	0	0	1 de 1	5
5	0	0	0	0
6	2	7	1 de 1	21
		5		
7	0	0	1 de 1	17
8	1	20	3 de 3	23
				23
				23
9	0	0	1 de 1	5

Tabla 4. Resultados LZM

ARV				
P	C.E.	C	C.C	C
1	0	0	1 de 1	10
2	0	0	1 de 1	16
3	0	0	1 de 1	21
4	1	9	1 de 1	22
5	0	0	1 de 1	8
6	0	0	1 de 1	23
7	0	0	1 de 1	21
8	1	20	3 de 3	19
				16
				12
9	0	0	1 de 1	5

Tabla 5. Resultados ARV

Los tres algoritmos de detección se han aplicado a un total de 9 señales EEG con 11 crisis, se han analizado 1800 segundos por cada paciente, comprendiendo siempre las crisis en este tiempo. Para cada paciente se han diferenciado entre las crisis detectadas en el tiempo en el que ocurren las crisis reales (crisis ciertas) y las crisis detectadas en el tiempo en el que realmente no hay crisis (crisis falsas), ambas acompañadas del número de canales en el que se detecta, cabe recordar que una crisis será detectada cuando se produzca en al menos 5 canales. A continuación, como ejemplo, se muestran las gráficas del paciente 7, donde se ve como se han obtenido los resultados.

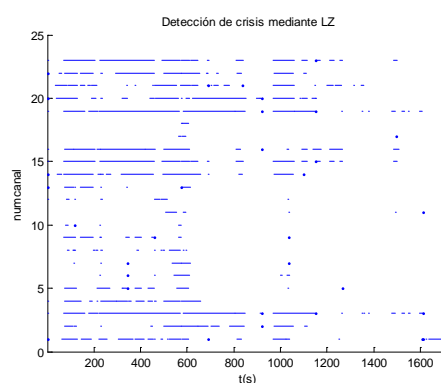


Figura 7. Resultados LZ paciente 7

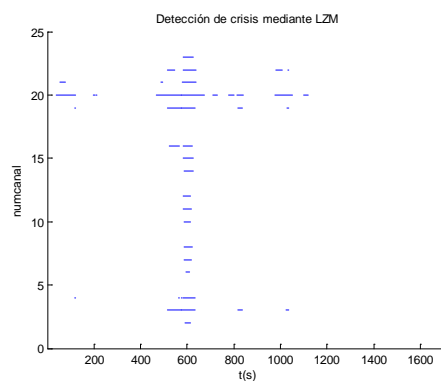


Figura 8. Resultados LZM paciente 7

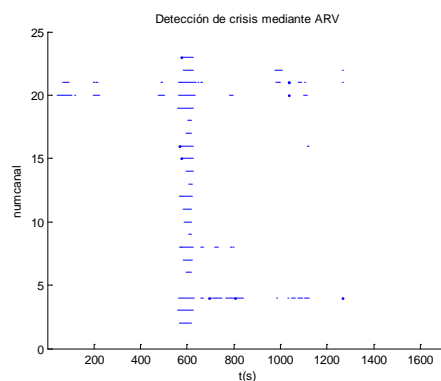


Figura 9. Resultados ARV paciente 7

En el eje de las 'y' se ha colocado el número de cada canal y en el de las 'x' el tiempo en segundos.

Cada vez que el umbral superior es rebasado se dibuja la línea azul y cuando el valor es menor que el umbral inferior deja de hacerlo. El paciente 7 corresponde al 10_27 de la base de datos, su crisis está comprendida entre los segundos 2382 y 2447, cabe recordar, que el análisis lo hacemos sobre 1800 segundos, así que la crisis debería empezar aproximadamente en $2382-1800=582$ segundos. En este proyecto no se pretende detectar la duración de las crisis sino su aparición, por eso no hemos de fijarnos en la extensión de cada crisis sobre el eje del tiempo, sino el número de canales que se detecta en el mismo tiempo.

A continuación, en las figuras 28, 29 y 30 se muestra el canal 23 de la mismo paciente (7), en las cuales se observa una detección correcta para LZM y ARV pero una detección incorrecta para el LZ, tal como reflejan las últimas tablas.

Cabe destacar la figura, ya que ella se muestra el correcto funcionamiento de los umbrales máximo y mínimo, ya que justo antes de la crisis, se aprecia un pico de energía, pero al no sobrepasar el umbral máximo, sino solo el mínimo, el sistema no lo reconoce como una crisis.

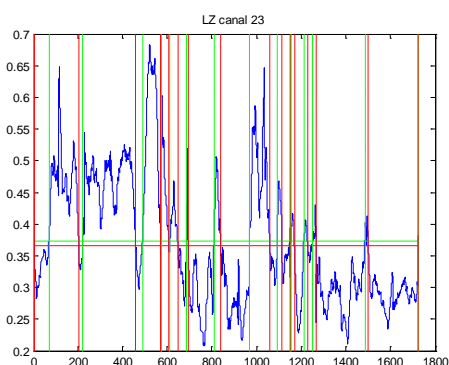


Figura 10. Resultados LZ

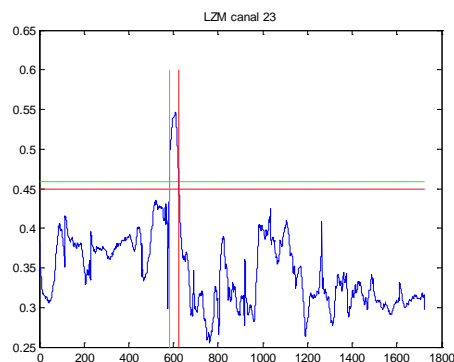


Figura 11. Resultados LZM

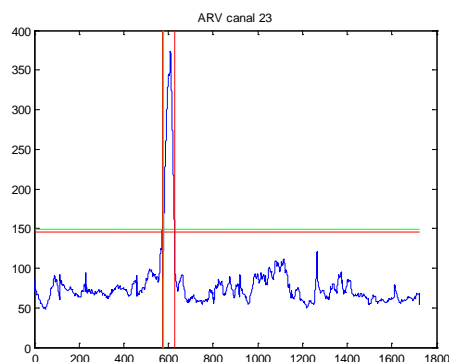


Figura 12. Resultados ARV paciente 7

4. Discusión y conclusiones

En la tabla 3, se encuentran los resultados de la detección utilizando el algoritmo de Lempel-Ziv, de las 11 crisis, se detectan 10, pero en el resto del tiempo en el que no hay crisis, se detecta de forma errónea en 6 pacientes.

Los resultados de la detección con el algoritmo de Lempel-Ziv multiestado, mostrados en la tabla 4, reflejan que con este algoritmo se reconocen 9 de las 11 crisis verdaderas, una menos que con el LZ, pero solo detecta 2 pacientes de los 9 con crisis erróneas.

Finalmente, con el ARV (tabla 5) todas las crisis ciertas son detectadas, 11 de 11, y solo en los pacientes 4 y 8 detecta crisis erróneas.

En la tabla 6 se muestra una síntesis de los resultados obtenidos.

C.V.11: Crisis verdaderas sobre 11.

%ac.: % de acierto del algoritmo.

P.E.: Pacientes con crisis erróneas sobre 9.

Total C.E.: Total de crisis erróneas.

	C.V. 11	% ac.	P.E. 9	Total C.E.
LZ	10	90,9	6	15
LZM	9	81,8	2	3
ARV	11	100	2	2

Tabla 6. Resultados finales

Al ser un detector de crisis y no de tiempo, no se puede dar un valor de sensibilidad, pero se puede afirmar con total seguridad que el método de ARV es el más fiable con un 100% de acierto sobre las crisis reales y solo 2 errores de detección en los 16200 segundos analizados en total. En segundo lugar se encuentra el LZM el cual ha detectado un 81,8% de las crisis reales y solo ha errado 3 veces en los 16200 segundos. Por último, la detección mediante el algoritmo de LZ no ha sido satisfactoria ya que aun detectando un 90,9% de las crisis, ha detectado 15 que no lo eran, así que de los tres métodos, este es el menos fiable.

5. Agradecimientos

El proyecto se ha realizado sobre la supervisión de Abel Torres Cebrián, profesor de la Universitat Politècnica de Catalunya (UPC). Agradezco su ayuda y conocimientos, sin los que no hubiera sido posible la realización de este estudio.

6. Bibliografía

- [1] GlaxoSmithKline. *Gskard Epilepsia y el cerebro*. Recuperado 30 Abril 2013, desde http://www.gskard.com/epilepsia/epilepsia_cerebro.php
- [2] The Patient Education Institute, Inc. *X-Plain*. Recuperado 30 Abril 2013, desde <http://www.nlm.nih.gov/medlineplus/spanish/tutorials/seizuresandepilepsyspanish/nr0491s3.pdf>
- [3] Diario médico. *Dmedicina Epilepsia*. Recuperado 4 Mayo 2013, desde <http://www.dmedicina.com/enfermedades/neurologicas/epilepsia>
- [4] Barea Navarro, Rafael. *Electroencefalografía, Instrumentación biomédica*, Departamento Electrónica, Universidad Alcalá. Recuperado 20 Mayo 2013, desde <http://www.bioingenieria.edu.ar/academica/catedras/bioingenieria2/archivos/apuntes/tema%205%20-%20electroencefalografia.pdf>

[5] Aboy, M., Hornero, R., Abasolo, D. & Alvarez, D. (2006). Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Trans Biomed Eng*, 53.

[6] Alejandro Gustavo Sabatini (2006). *Interacción humano máquina. Metodología del procesamiento de la bioseñal*.

[7] Leonardo Sarlabous, Abel Torres, José A. Fiz, Josep Morera, Raimon Jané, "Index for estimation of muscle force from mechanomyography based on the Lempel-Ziv algorithm" *Journal of Electromyography and Kinesiology*, 2013.