

# Detección de fases de actividad en señales electromiográficas usando un criterio basado en el número de corridas

*José Antonio Guerrero Díaz de León  
Jorge Eduardo Macías Díaz*

## **Introducción**

Las señales de electromiografía superficial, o señales EMG, capturan a nivel de piel señales eléctricas asociadas a la actividad de los músculos cercanos a la región donde se colocan los electrodos (Mills, 2005).

Una tarea que forma parte del análisis de la actividad muscular a través de señales EMG es la detección de las fases de actividad, que son partes de la señal que se distinguen de las fases de silencio por tener una amplitud relativamente mayor. En la parte izquierda de la Figura 1 se muestra una señal EMG en la que, a grandes rasgos, se observan seis fases de silencio y seis fases de actividad. El método más sencillo para determinar las fases de actividad es el método de umbralizado simple, en

el cual se calcula una envolvente de la señal EMG y luego se aplica un umbral (Ozgünen, Celik & Kurdak, 2010). De esta manera, los valores por debajo del umbral se consideran como pertenecientes a una fase de silencio y los que están por encima como pertenecientes a una fase de actividad. Los resultados obtenidos de este procedimiento dependen entonces del tipo de envolvente calculada y del valor de umbral usado.

En este trabajo se describe un criterio basado en el número de corridas para determinar el valor óptimo del umbral a usar, partiendo de que se cuenta con la envolvente (de cualquier tipo) de la señal EMG a umbralizar.

## Desarrollo

Sea  $x = \{x_1, x_2, \dots, x_n\}$  una señal EMG, donde cada  $x_i$  es un valor de voltaje obtenido al tiempo  $t_i$  y los valores  $\{t_1, t_2, \dots, t_n\}$  son puntos de muestreo equiespaciados en un intervalo de tiempo. A cada punto de la señal  $x$  lo podemos caracterizar como perteneciente a una fase de silencio o una fase de actividad mediante una secuencia binaria  $b = \{b_1, b_2, \dots, b_n\}$ , donde  $b_i$  toma el valor 1 si  $x_i$  pertenece a una fase de actividad y toma el valor 0 en cualquier otro caso. Este vector  $b$  por lo regular no es conocido y su estimación corresponde al problema de detección de fases de actividad.

Por otra parte, bajo un esquema de umbralizado simple se toma un valor de umbral  $\tau$  para formar un vector  $\beta = \{\beta_1, \beta_2, \dots, \beta_n\}$ , donde  $\beta_i$  toma el valor 1 si  $\gamma_i > \tau$  y toma el valor 0 en cualquier otro caso, donde  $\{\gamma_1, \gamma_2, \dots, \gamma_n\}$  es el resultado de aplicar un operador de cuantificación (rectificación o de cálculo de envolvente) en la señal  $x$ . El objetivo es determinar el valor del umbral  $\tau$  que permita diferenciar puntos de fases de silencio de puntos de fase de actividad, es decir, queremos obtener  $\beta = b$ .

## Número de corridas

Hay que notar que tanto en  $\beta$  como en  $b$  podemos encontrar secuencias formadas por el mismo símbolo (puros 0's o puros 1's). Cada secuencia maximal de 0's o 1's se denomina "corrida". Por ejemplo, en la secuencia *11001000011001001101110001001100* se forman 16 corridas, las cuales son *11, 00, 1, 0000, 11, 00, 1, 00, 11, 0, 111, 000, 1, 00, 11* y *00*. El vector  $\beta$  ideal (es decir, el aso-

ciado a la mejor clasificación de los puntos en  $x$ ) tiene la característica de que las corridas corresponden a cada una de las fases (de silencio y de actividad) presentes en la señal. Esta característica puede ser usada para determinar el valor de umbral adecuado para umbralizar una señal EMG (Thexton, 1996).

Para un valor particular de umbral  $\tau$ , podemos determinar el número de corridas  $R_\tau$  formadas después de umbralizar. Dado que la señal EMG está organizada en fases de actividad y de silencio, un buen valor de umbral da un número pequeño de corridas. El problema es que si solamente buscamos el mínimo en el conjunto de posibles  $R_\tau$  tenemos que siempre será 1 (cuando  $\tau = 0$  o  $\tau = \max \gamma_i$ , por ejemplo). Así, el punto no es encontrar el mínimo número de corridas, sino determinar cuando el número de corridas obtenido es mucho menor al esperado si la señal no estuviera organizada en corridas. Para calcular el valor esperado (estadísticamente) del número de corridas se necesita conocer la distribución de  $R_\tau$ .

### Distribución estadística del número de corridas para un vector aleatorio

Consideremos una variable aleatoria Bernoulli  $b$ , con probabilidad de éxito  $p$  y probabilidad de fracaso  $q = 1-p$ . Sea  $R$  el número de corridas formadas en un vector  $\{b_1, b_2, \dots, b_n\}$  de  $n$  muestras aleatorias independientes de  $b$ .  $R$  es una variable aleatoria y estamos interesados en su distribución.

Sean  $D_i = |b_i - b_{i+1}|$ , para  $i = 1, 2, \dots, n-1$ .  $D_i$  son variables indicadoras de un cambio de  $b_i$  a  $b_{i+1}$ , ya que  $D_i = 0$  si  $b_i = b_{i+1}$  y 1 si  $b_i \neq b_{i+1}$ . Estas variables indicadoras están relacionadas con el número de corridas de la siguiente manera:

$$R = 1 + \sum_{i=1}^{n-1} D_i$$

El número 1 en esta expresión es debido a la corrida presente antes del primer cambio en  $b$ . De esta manera podemos obtener la distribución de  $R$  de las propiedades estadísticas de  $D_i$ . Es fácil mostrar (a partir de las propiedades de la distribución Bernoulli) que las variables indicadoras  $D_i$  tienen las siguientes propiedades estadísticas (valor esperado, varianza y covarianza):

$$E [D_i] = 2pq$$

$$V [D_i] = 2pq (1 - 2pq)$$

$$\text{Cov} [D_i, D_j] = \begin{cases} 0 & |i-j| = 1 \\ pq (1 - 4pq) & |i-j| > 1 \end{cases}$$

Las propiedades anteriores permiten determinar las propiedades estadísticas del número de corridas  $R$  (valor esperado y varianza):

$$E [R] = 1 + 2 (n - 1)pq$$

$$V [R] = 2pq (1 - 2pq)$$

Además, dado que  $R$  proviene de la suma de variables aleatorias  $D_i$ , por el Teorema del Límite Central, la distribución de  $R$  se puede aproximar mediante una distribución normal, esto aún cuando las variables están (ligeramente) correlacionadas; de hecho, la correlación es débil, ya que ésta es 0 entre variables  $D_i$  no consecutivas y  $\text{Cov}[D_p, D_{i+p}] < 0.0625$  (que es el máximo de  $\text{Cov}[D_p, D_{i+p}]$  vista como función de  $p$ ).

### Cálculo del umbral

Sea  $p_\tau$  la probabilidad de que  $g_i$  sea mayor que un umbral dado  $\tau$  y  $q_\tau = 1 - p_\tau$ . La probabilidad  $p_\tau$  se obtiene de la distribución empírica de los datos (Silverman, 1986). Consideremos el número  $R_\tau$  de corridas formadas en  $\beta$  al aplicar dicho umbral. Por los resultados mostrados anteriormente, la distribución de  $R_\tau$  se aproxima mediante una distribución normal con parámetros (media y varianza):

$$\mu_\tau = E[R_\tau] = 1 + 2(n - 1)p_\tau q_\tau$$

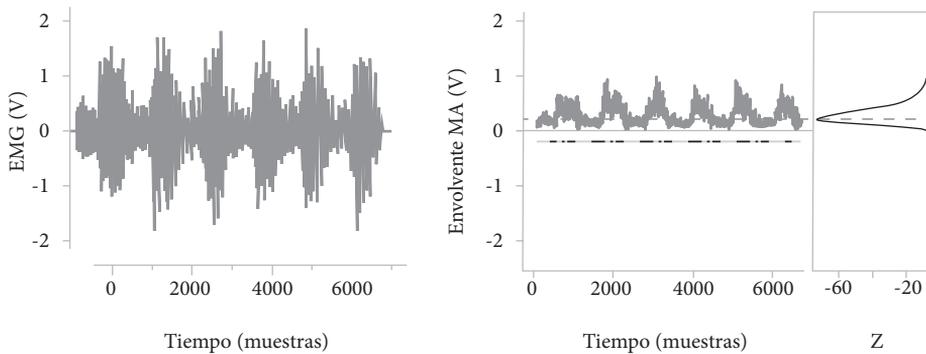
$$\sigma^2 = V[R_\tau] = 2p_\tau q_\tau (1 - 2p_\tau q_\tau)$$

Para determinar si un valor  $R_\tau$  es mucho menor al esperado (en el caso aleatorio), bajo un esquema frecuentista, calculamos la probabilidad  $P(R < R_\tau)$ . Dado que la distribución de  $R_\tau$  se puede aproximar mediante una distribución normal, podemos usar el valor estandarizado  $Z_\tau = (R_\tau - \mu_\tau)/\sigma_\tau$  y calcular las probabilidades  $P(Z < Z_\tau)$ , donde  $Z_\tau \sim N(0, 1)$ . De esta manera, el valor de umbral propuesto está dado por:

$$\tau_{Runs} = \underset{\tau}{\operatorname{argmax}} P(Z < Z_{\tau}) = \underset{\tau}{\operatorname{argmin}} Z_{\tau}$$

Esto, porque todas las variables  $Z_{\tau}$  tienen la misma distribución (están estandarizadas) y, por lo tanto, sus valores son comparables entre sí. La forma de calcular el umbral para una señal EMG se muestra en la Figura 1.

Figura 1. Esquema del cálculo de umbral  $\tau_{Runs}$  para una señal EMG. De izquierda a derecha: señal EMG original del músculo bíceps, envolvente MA de la señal EMG, y la gráfica de los valores  $Z_{\tau}$ . La línea punteada en color rojo muestra el valor calculado del umbral  $\tau_{Runs}$  y la secuencia de puntos debajo del gráfico de la envolvente representa el vector  $\beta$  resultante del umbralizado, con puntos en color azul para representar el valor 1 y el color gris para representar el valor 0.



Fuente: Elaboración propia.

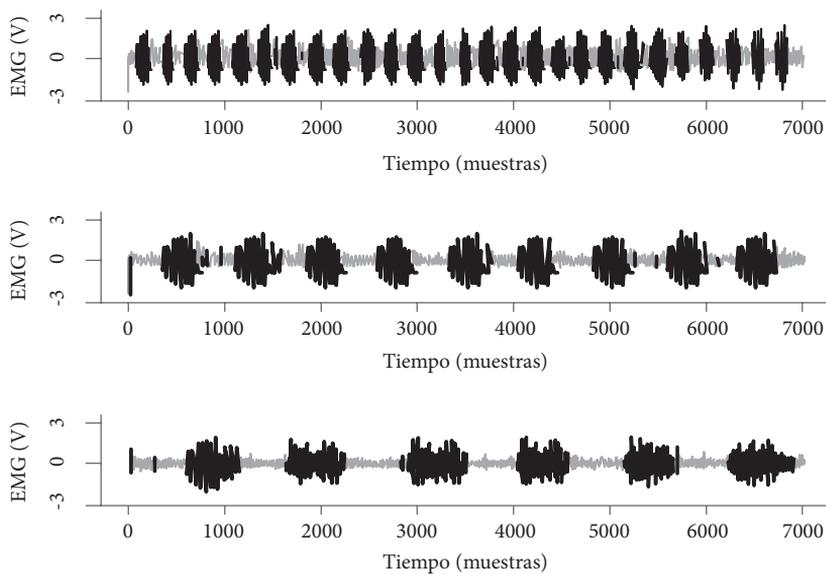
## Experimentos

En esta sección se muestra el desempeño en señales reales experimentales (de manera cualitativa), obtenidas del músculo bíceps brachii, con diferentes duraciones de las fases de actividad (Figura 2). Con el propósito de evaluar el desempeño de manera cuantitativa, también se usan señales sintéticas. La evaluación cuantitativa se realiza con dos medidas de calidad que reflejan el nivel de error, a saber la tasa de verdaderos positivos (sensibilidad) y la tasa de verdaderos negativos (especificidad).

Para obtener las señales experimentales, un sujeto de prueba se sienta frente a un escritorio, colocando el codo cerca del cuerpo y formando un án-

gulo de 90 grados con respecto al antebrazo. La palma de la mano se coloca hacia arriba y con un peso de 3.5kg. El brazo está inicialmente en reposo sobre el escritorio. Luego, el peso se levanta hasta que el brazo forma un ángulo de 45 grados con respecto al plano horizontal. Después el brazo se coloca de nuevo en la posición de reposo sobre el escritorio. Estos movimientos se repiten varias veces para obtener un patrón alternante de actividad y silencio de aproximadamente la misma duración.

Figura 2. Resultado de aplicar el algoritmo de detección propuesto a señales EMG experimentales obtenidas del músculo bíceps brachii, realizando flexiones con una duración aproximada de las fases de 0.5, 1.5 y 2.5 segundos (de arriba hacia abajo). Los puntos en azul corresponden a puntos detectados como pertenecientes a fases de actividad y los puntos en gris a fases de silencio.



Fuente: Elaboración propia.

En la Figura 2 se muestran tres señales EMG obtenidas del músculo bíceps brachii mientras realizaba flexiones con una duración aproximada de 0.5, 1.5 y 2.5 segundos. Cada señal consiste de 7,000 muestras correspondientes a 28 segundos de actividad con una frecuencia de muestreo de 250Hz. En la misma

figura se muestra el resultado de aplicar el umbral propuesto  $\tau_{\text{Runs}}$  mediante un código de color para marcar las partes de las señales detectadas como fases de silencio (en color gris) y fases de actividad (en color azul). Para el cálculo de la envolvente se usó el método MA (*moving average*), con un tamaño de ventana de 21 muestras.

Por otra parte, para cuantificar la calidad de la detección se usan señales sintéticas, de manera que se conoce para cada punto si pertenece o no a una fase de actividad (es decir, se conoce a  $b$ ). Las señales sintéticas fueron generadas mediante la yuxtaposición de fases sintéticas de silencio y de actividad generadas mediante el método propuesto por Guerrero, Castillo-Gálvan y Macías-Díaz (2018). Las fases de silencio sintéticas se conforman de muestras aleatorias independientes de una distribución normal con media 0 y desviación estándar  $\sigma_s$ .

Las fases de actividad sintéticas también son una secuencia de muestras aleatorias de una distribución normal con media 0, pero en este caso con desviación estándar  $\sigma_A$ , con  $\sigma_A > \sigma_s$ . A esta secuencia luego se le multiplica por una ventana de pesos para obtener una forma de la fase similar a las señales experimentales. Por último, a toda la secuencia formada por fases de silencio y de actividad se le aplica un filtro Butterworth pasa-altas con frecuencia de corte de 20Hz, a fin de simular la correlación intrínseca de las muestras en una señal EMG experimental. Los valores de  $\sigma_s$  y  $\sigma_A$  se eligen de manera que se obtiene una relación señal-ruido deseada, cuantificada mediante  $\text{SNR} = 20 \log_{10}(\sigma_A / \sigma_s)$ . En particular, se tomaron  $\sigma_s = 0.24$  y  $\sigma_A = 0.83$ , dando un SNR de 10.77dB, similar al observado en señales EMG experimentales. Se simularon 100 señales con duración aleatoria de fases de aproximadamente 0.5s (120 muestras a 250Hz) y 100 señales con duración aleatoria de aproximadamente 1.5s (375 muestras a 250Hz).

Para medir la calidad de una detección de fases, después de aplicar un método de detección, se categoriza a cada punto  $x_i$  como verdadero negativo ( $\beta_i=0, b_i=0$ ), falso negativo ( $\beta_i=0, b_i=1$ ), falso positivo ( $\beta_i=1, b_i=0$ ) y verdadero positivo ( $\beta_i=1, b_i=1$ ). Se calculan luego TP, FP, TN y FN, que son el total de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos, respectivamente. Con base en estas cantidades se definen la sensibilidad (tasa de verdaderos positivos) y la especificidad (tasa de verdaderos negativos) como  $\text{SEN} = 100\% \times \text{TP} / (\text{TP} + \text{FN})$  y  $\text{SPE} = 100\% \times \text{TN} / (\text{TN} + \text{FP})$ , respectivamente.

El desempeño del umbral propuesto se comparó con los umbrales  $\tau_{\text{Rand}}$ ,  $\tau_{\text{RMS}}$  y  $Dt$ , que se describen brevemente a continuación:

1. Umbral  $\tau_{\text{Rand}}$  (Thexton, 1996): Se basa en el criterio sobre el número de corridas similar al propuesto en este trabajo. Para la integral acumulada (Bak y Loeb, 1979) se utilizó un valor de reinicio de cinco muestras.
2. Umbral  $\tau_{\text{RMS}}$  (Ozgünen *et al.*, 2010): El umbralizado usa un valor de umbral prefijado sobre la envolvente RMS dado por un porcentaje del promedio de la envolvente de toda la señal. En el trabajo se recomiendan los porcentajes 25%, 30% y 40%. Se utilizó el de 25% ya que éste arrojó los mejores resultados para las señales utilizadas en el experimento. El tamaño de ventana usado para el cálculo de la envolvente RMS fue de 21 muestras.
3.  $Dt$  (Bonato, D'Alessio y Naflitz, 1998): El método se basa en un doble umbralizado. Los parámetros (ver documento original para más detalles) fueron  $m = 5$ ,  $P_{fa} = 0.01$  y un mínimo de 15 muestras por fase. Para la estimación del valor de la desviación estándar del ruido que necesita el método, se calculó la desviación estándar de las primeras 30 muestras de cada señal (todas las señales sintéticas utilizadas comienzan con una fase de silencio de mínimo 120 muestras).

Para el método propuesto, se usó la envolvente MA con un tamaño de ventana de 21 muestras.

El resumen de los resultados para las pruebas con las 200 señales (100 de cada diferente duración de fases) se muestran en la Tabla 1. Se puede observar que para las dos duraciones de fases (0.5s y 1.5s) los resultados son similares. El método propuesto  $\tau_{\text{Runs}}$  y el método  $Dt$  tuvieron un buen desempeño tanto en SEN como en SPE, teniendo mayor variabilidad en los resultados el método  $Dt$ . Los métodos  $\tau_{\text{Rand}}$  y  $\tau_{\text{RMS}}$  tuvieron un buen desempeño en sólo una de las dos medidas de calidad.

Tabla 1. Medidas de calidad de detección de fases en señales EMG

Método	0.5s		1.5s	
	SPE	SEN	SPE	SEN
$\tau_{\text{Runs}}$	96.73±0.91	96.36±0.96	98.29±0.68	98.15±0.64
$\tau_{\text{RMS}}$	56.04±3.95	99.90±0.10	59.26±3.81	99.95±0.08
$\tau_{\text{Rand}}$	93.88±2.11	62.28±4.00	94.77±2.08	62.46±4.02
$Dt$	98.49±2.13	84.56±8.85	99.6±1.17	86.77±9.30

Fuente: Elaboración propia. Nota: Los valores de SPE y de SEN están en la forma media  $\pm$  desviación estándar de las 100 señales (por cada duración).

## Conclusiones

En este trabajo se obtuvo el valor del umbral para un método de umbralizado simple, que es óptimo usando un criterio basado en el número de corridas obtenidas. El valor de umbral obtenido se adapta a las características de la distribución estadística de los valores de la envolvente usada. Con esto es posible obtener buenos resultados de detección para un amplio tipo de señales EMG.

Se realizaron pruebas del umbral propuesto en señales EMG reales, obtenidas del músculo bíceps braquial mientras se realizaban rutinas de ejercicio. Se mostró cualitativamente que la detección de fases fue satisfactoria. También se comparó cuantitativamente el desempeño del umbral propuesto contra varios métodos del estado del arte, usando señales sintéticas. Para el método propuesto se obtuvo un mejor compromiso en las dos medidas de calidad empleadas (buen desempeño en ambas medidas), así como menor variabilidad en los resultados obtenidos.

Cabe mencionar que el método propuesto puede ser mejorado aplicando algún método de posprocesamiento, para eliminar fases resultantes de corta duración.

## Referencias

- Bak, M. & Loeb, G. (1979). A pulsed integrator for emg analysis. *Electroencephalography and Clinical Neurophysiology*, 47(6), 738-741.
- Bonato, P., D'Alessio, T. & Naflitz, M. (1998). A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait. *IEEE Transactions on Biomedical Engineering*, 45(3), 287-299.
- Guerrero, J.A., Castillo-Galvan, M.A. & Macias-Diaz, J.E. (2018). Novel electromyography signal envelopes based on binary segmentation. *Biomedical Signal Processing and Control*, 45(1), 225-236.
- Mills, K.R. (2005). The basics of electromyography. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(2), ii32-ii35.
- Ozgülünen, K.T., Celik, U. & Kurdak, S.S. (2010). Determination of an optimal threshold value for muscle activity detection in emg analysis. *Journal of Sports Science & Medicine*, 9(4), 620-628.
- Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. New York, USA: Chapman and Hall.
- Thexton, A.J. (1996). A randomisation method for discriminating between signal and noise in recordings of rhythmic electromyographic activity. *Journal of Neuroscience Methods*, 66(2), 93-98.