

## Sistema de anotación semiautomático de señales electrocardiográficas de esfuerzo

*Juan Fajardo-Reinoso<sup>1</sup>, Darwin Astudillo-Salinas<sup>1</sup>, Lizandro Solano-Quinde<sup>1</sup>, Sara Wong<sup>1,2</sup>*

<sup>1</sup> Departamento de Ingeniería Eléctrica, Electrónica y Telecomunicaciones, Universidad de Cuenca, Av. 12 de Abril y Agustín Cueva, Cuenca, Ecuador, 010201.

<sup>2</sup> Investigador Prometeo, Senescyt, Ecuador.

Autores para correspondencia: [juan.fajardor@ucuenca.ec](mailto:juan.fajardor@ucuenca.ec), {[fabian.astudillos](mailto:fabian.astudillos@ucuenca.edu.ec), [lizandro.solano](mailto:lizandro.solano@ucuenca.edu.ec)}  
[swong@usb.ve](mailto:swong@usb.ve)

Fecha de recepción: 19 de junio del 2016 - Fecha de aceptación: 24 de julio del 2016

### ABSTRACT

Despite the availability of electrocardiographic (ECG) databases does the medical sector not yet possesses an annotated stress database. This paper introduces an annotation system using RR-time series obtained from an eight lead stress database (DICARDIA). First, the system proposes to the user a lead (reference channel) according to its statistical measures. The user then undergoes a visual inspection validating or denying the channel proposed by the system. Afterwards, the system proposes two options based on the quality of the RR-time series. If the series contains few artifacts the annotations are realized using the interval of annotations; in the case of noisy series the system allows annotations beat by beat. The preliminary results realized over 172025 beats (approximately 15% of DICARDIA database) give a sensibility and positive predictive value of 97.66% and 96.71% respectively. The system will permit the delineation of stress databases, which will be an important starting point for evaluating the performance of QRS detectors.

Keywords: ECG, RR, QRS, stress test, annotation.

### RESUMEN

A pesar de la cantidad de Bases de datos electrocardiográficas disponibles para la comunidad científica, no existe a disposición una Base de Datos de esfuerzo anotada. En este trabajo se presenta un sistema de anotación utilizando secuencias RR obtenidas de una Base de Datos de Esfuerzo de 8 derivaciones (DICARDIA). Inicialmente el sistema propone al usuario una derivación (canal de referencia) que ha sido evaluada según sus características estadísticas. El usuario realiza una inspección visual y valida o rechaza el canal de referencia. Posteriormente, el sistema propone dos opciones según la calidad de la secuencia RR, en caso de que la secuencia contenga pocos artefactos la anotación se realiza sobre ventanas de interés. En el caso de secuencias RR muy ruidosas el sistema permite la anotación latido a latido. Los resultados preliminares realizados sobre 172025 latidos (aproximadamente 15% de DICARDIA) reportan una sensibilidad y un valor predictivo positivo de 97.66% y 96.71% respectivamente. Este sistema permitirá anotar Bases de Datos de ECG de esfuerzo, lo cual será un punto de partida importante para evaluar el desempeño de los detectores de QRS.

Palabras clave: ECG, RR, QRS, prueba de esfuerzo, anotación

## 1. INTRODUCCIÓN

La prueba de esfuerzo es un estudio electrocardiográfico no invasivo comúnmente empleado para medir el efecto del ejercicio progresivo sobre el corazón, se fundamenta en que, al practicar un ejercicio físico, el organismo necesita más oxígeno y glucosa, por tanto, el corazón debe aumentar su capacidad de bombeo. Así, sometiendo el corazón a un trabajo adicional, mientras se vigila al paciente y se monitoriza

su electrocardiograma (ECG), es posible descubrir problemas cardíacos que no son evidentes en el sujeto en reposo. Esta prueba proporciona importantes datos de diagnósticos y pronósticos en una amplia variedad de pacientes, desempeña un papel fundamental, ya que es una prueba fundamental empleada en el proceso de diagnóstico del paciente con cardiopatía isquémica o enfermedad coronaria (Almeida & Brandi, 1995). Los Registros electrocardiográficos de una prueba de esfuerzo contienen una gran cantidad de ruido debido al movimiento del paciente, siendo prácticamente imposible obtener un registro libre de ruido. Adicionalmente, los latidos ectópicos y otros artefactos pueden ocurrir tanto en sujetos sanos como en sujetos con patologías cardíacas. El análisis de las secuencias *RR* latido a latido de una prueba de esfuerzo es realmente difícil de analizar. Además de los artefactos fisiológicos, se pueden presentar muchos errores en la serie-*RR* debido a los aspectos de instrumentación de registro de la señal como la mala colocación de los electrodos o relacionados con el algoritmo de detección de *QRS*.

En la actualidad se puede encontrar un gran número de algoritmos y métodos desarrollados para la detección y caracterización de las señales de electrocardiograma, desde el propuesto por Pan & Tompkins (Pan & Tompkins, 1985), pasando por los algoritmos basados en redes neuronales (Dokur, & Ólmez, 2001), en ondaletas (Martínez *et al.*, 2004), en algoritmos evolutivos (Dumont *et al.*, 2010) entre otros. Adicionalmente, se puede encontrar una gran cantidad de bases de datos con distintas anotaciones y morfologías, pero no existen validaciones ni anotaciones sobre bases de datos electrocardiográficas de esfuerzo (Goldberger *et al.*, 2000); en parte debido a la ausencia de una base de datos anotada para validar los algoritmos. Las validaciones en condiciones de esfuerzo han sido realizadas en su mayoría utilizando la base de datos de *Noise Stress Test Database* (Moody & Mark, 1990) pero a pesar de tener diversidad de ruidos no representa una fuente significativa de cambios de frecuencia cardíaca y no linealidades presentes en una prueba de esfuerzo. Recientemente se pusieron en línea tres bases de datos electrocardiográficas adquiridas durante pruebas de esfuerzo (Quintero *et al.*, 2007); sin embargo, las señales ECG no se encuentran anotadas.

En trabajos recientes se han hecho esfuerzos por realizar una anotación de las series *RR* de una prueba de esfuerzo (Armijos *et al.*, 2015; Astudillo *et al.*, 2016). La metodología propone clasificar las series *RR* en cuatro clases: referencia, alta calidad, baja calidad e inservible a partir de subdividir la serie temporal *RR* en un conjunto de ventanas para la estimación de la cantidad de artefactos en base a un valor de umbral que depende de la desviación estándar de la serie *RR*.

En este trabajo se propone un sistema de anotación de señales electrocardiográficas de esfuerzo a partir de sus secuencias *RR* validadas de forma semiautomática. A partir de las series *RR* se guía al anotador experto para validar, rechazar o editar la anotación de la onda *R*. En las siguientes secciones se describen la base de datos de *ECG* de esfuerzo, la metodología propuesta y se presentan los primeros resultados. El sistema de anotación genera un reporte basado en estándares de bases de datos electrocardiográficas de referencia. Finalmente, se presentan las conclusiones y los trabajos futuros.

## 2. MATERIALES Y MÉTODOS

### 2.1. Base de datos

La base de datos DICARDIA fue registrada en el Hospital "Universitario de Caracas", y contiene 65 datos clínicos y registros de *ECG* de 65 sujetos, 54 sujetos diabéticos y 11 sujetos sin evidencia de Diabetes o enfermedad Cardiovascular. DICARDIA es multiparamétrica contiene además del *ECG*, datos demográficos y clínicos, parámetros de laboratorio clínico y la evaluación de la prueba de esfuerzo realizada por cardiólogos expertos. Los sujetos fueron sometidos a una prueba de esfuerzo, siguiendo el protocolo de Bruce a la que se añadió una etapa de calentamiento de tres minutos. La prueba comenzó con una velocidad de 1,7 kilómetros por hora y pendiente de 0° y terminó con una velocidad de 6 mph y una pendiente de 22°. La prueba consta de siete etapas, con una duración de tres minutos cada una. Para un sujeto no entrenado la duración promedio de la prueba es de 20 minutos, esto significa que la mayoría de los sujetos finalizaron la prueba entre las etapas 3 y 5 (Quintero *et al.*, 2007). Los *ECG* fueron digitalizados a una frecuencia de muestreo de 500 Hz, 12 bits de resolución, y un rango dinámico de  $\pm 5$  mV. Estas señales se encuentran disponibles en el portal GBBAnet (Ledezma *et al.*, 2014).

2.2. Metodología

La metodología que se utilizó está organizada como se indica a continuación: (i) se describe la generación de la secuencia RR; (ii) se propone un canal de referencia (medidas estadísticas); (iii) se valida y anota el canal de referencia (por intervalos de anotación, latido a latido); (iv) se genera un reporte (archivo de anotación). La Figura 1 describe el proceso de anotación semiautomático.

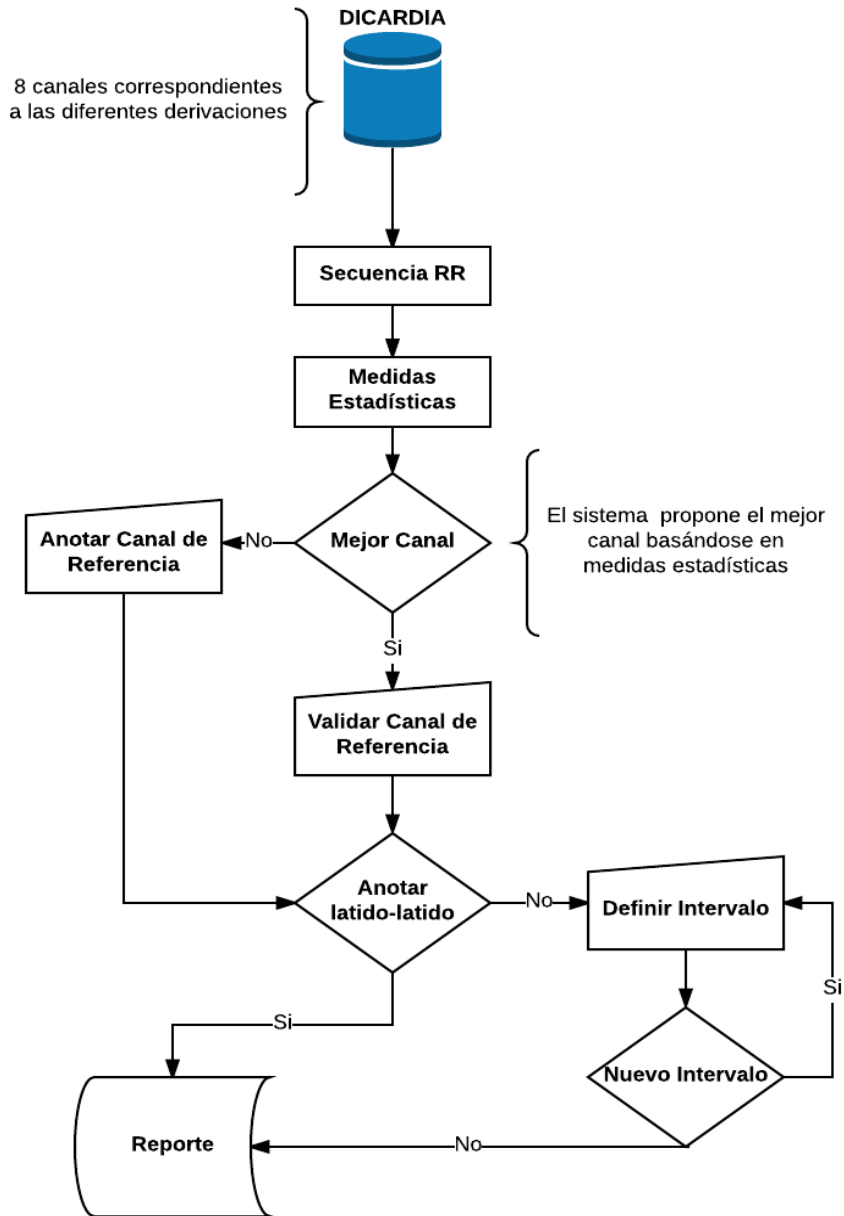


Figura 1. Diagrama del proceso de anotación.

Generación de secuencias RR

La serie RR viene dada por la diferencia entre las ondas R en el tiempo, formalmente es la oscilación del intervalo entre latidos consecutivos. La detección de los picos QRS fue realizada usando el detector de Pan & Tompkins (1985), disponible en GBBANet (Ledezma *et al.*, 2014). Se obtuvieron las 8 señales RR correspondientes a las derivaciones electrocardiográficas para los 65 registros.

### Selección de la secuencia RR de referencia

En esta etapa del proceso de anotación se selecciona una señal de referencia, para lo cual se realiza una inspección visual de las secuencias RR de cada registro. Se despliega en la pantalla las ocho series RR. El sistema propone uno o varios canales de referencia al anotador y el anotador valida las señales. En la Figura 2 se puede visualizar la interfaz, en la parte que corresponde a la anotación (parte derecha) el sistema anota el mejor canal como 1 (Astudillo *et al.*, 2016), mientras que en la parte superior izquierda se encuentran las series RR. Las series anotadas como 1 corresponden a las señales de referencia propuestas por el sistema, en la parte derecha el usuario selecciona y valida de forma manual la señal de referencia. En este punto es importante tomar en cuenta la cantidad de ruido observado y la cantidad de artefactos presentes en cada canal, la selección del canal/canales de referencia es un proceso puramente visual.

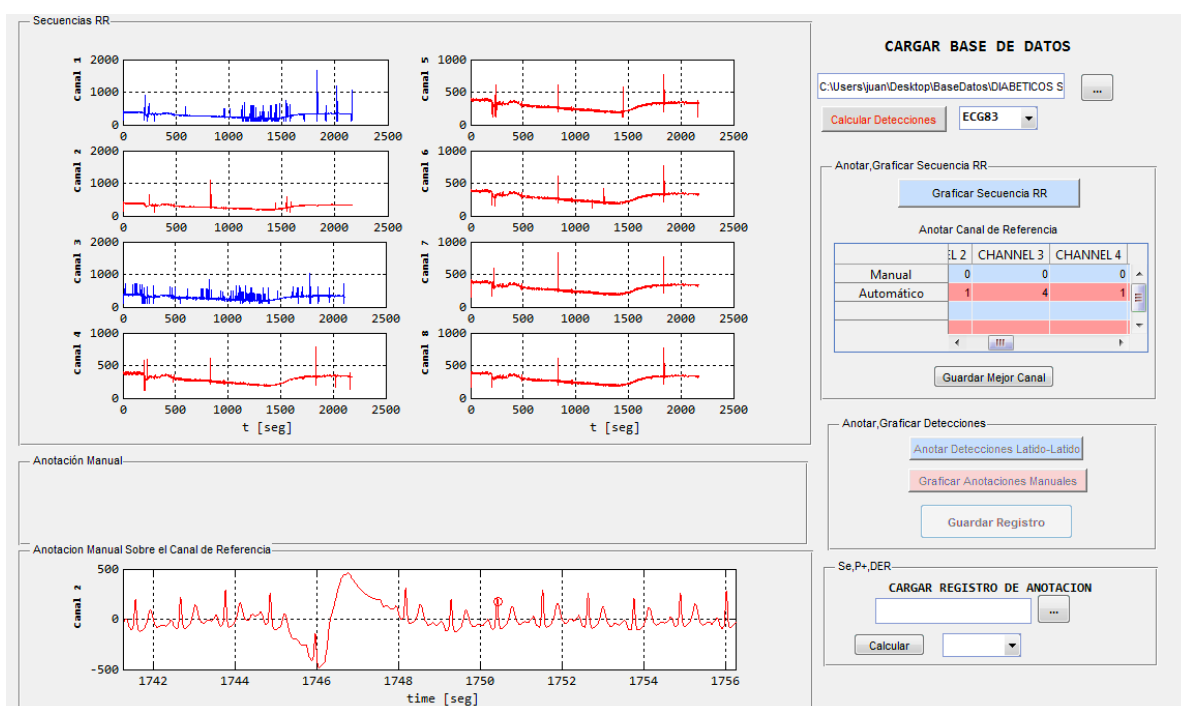


Figura 2. Secuencias RR del registro ECG83 desplegadas en la interfaz de anotación.

El método utilizado, para proponer el mejor canal se basa en la metodología propuesta por Astudillo *et al.* (2016). Consiste en dividir la secuencia RR de cada canal en ( $N = 20$ ) ventanas temporales (NW) y seleccionar el canal de referencia basándose en medidas estadísticas. Se estima la desviación estándar de cada ventana. Se aplican tres umbrales ( $T1, T2, T3$ ) basados en la máxima desviación estándar ( $SDT = 14$ ) aceptable para cada ventana y se realiza un conteo de las ventanas en las cuales no se cumple el umbral establecido. La clasificación se realiza mediante el porcentaje de ventanas (%NW) que sobrepasan el umbral establecido en cuatro clases:

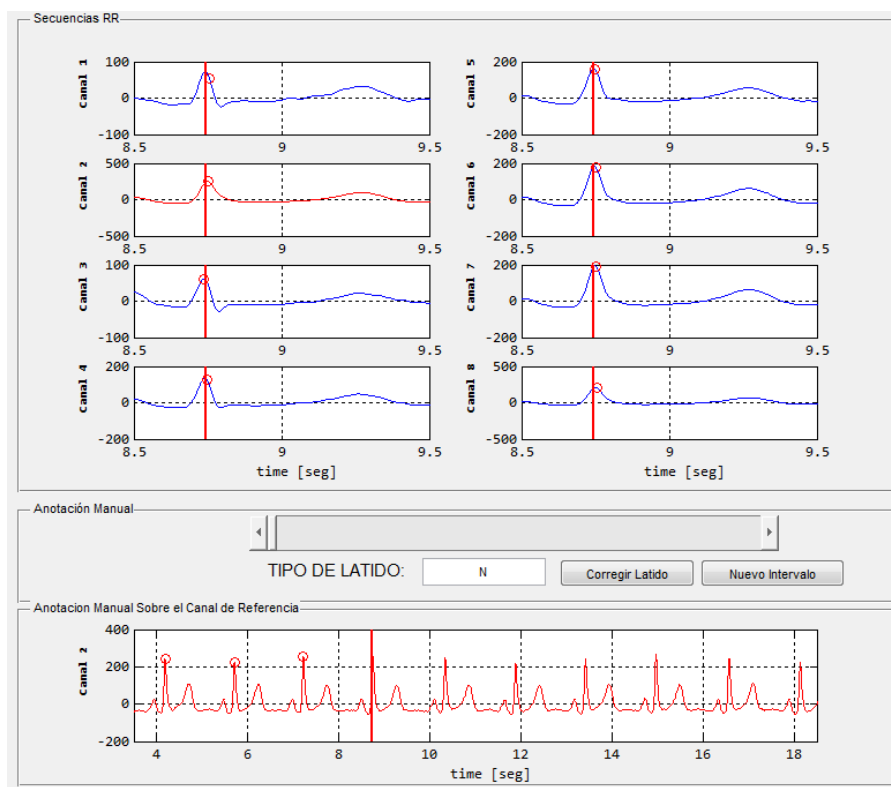
- 1) *Canal de referencia*: corresponde a una serie RR con muy pocos artefactos. Para la clasificación automática, se tiene este canal cuando  $\%NW < T1$ .
- 2) *Canal de alta calidad*: corresponde a una serie RR con un poco más de artefactos que el canal de referencia, pero que puede ser usado bajo un preprocesamiento previo. Para la clasificación automática, se tiene este canal cuando  $\%NW < T2$ .
- 3) *Canal de baja calidad*: esta señal contiene gran cantidad de artefactos y su uso puede limitarse solo a ciertos intervalos. Para la clasificación automática, se tiene este canal cuando  $\%NW < T3$ .

- 4) *Canal inservible*: representa una serie RR con gran cantidad de artefactos, de la cual es imposible extraer información útil. Para la clasificación automática, se tiene este canal cuando  $\%NW \geq T3$ .

Anotación semiautomática

Una vez seleccionado el canal de referencia, el usuario nuevamente realiza una inspección visual de la señal y decide si es necesario realizar una anotación manual latido a latido sobre toda la señal, o se define intervalos de anotación. Existen dos opciones:

- 5) *Sobre toda la secuencia RR*: Si las secuencias RR son de Baja Calidad, la anotación manual se realiza latido a latido utilizando ventanas temporales de 1seg. El anotador se desplaza con la ventana temporal sobre la serie RR obtenida de inicio a fin, analizando la naturaleza del latido y/o utilizando las detecciones QRS como referencia. El tamaño de la ventana de 1seg fue escogida de tal forma que se capture un solo latido. En la Figura 3 se muestra la interfaz para esta opción, se presentan 8 ventanas de 1seg correspondientes a los latidos de cada derivación. Además, en esta ventana también se puede visualizar las marcas de detección QRS utilizadas como referencia; en la sección inferior izquierda se presenta 15seg de la señal electrocardiográfica de referencia que se utiliza para dar un seguimiento a las anotaciones realizadas.
- 6) *Sobre los intervalos de la secuencia RR*: Al disponer de una secuencia RR de referencia con pocos artefactos (Referencia o Alta calidad) el usuario revisa de forma rápida la serie. Puede revisar algunos puntos al azar y/o ir directamente a los puntos de la serie donde se observan discontinuidades determinando si se trata de artefactos, latidos ectópicos o no detecciones. El usuario se desplaza con el cursor sobre la serie RR capturando un intervalo de interés para posteriormente realizar la anotación manual latido a latido. Las detecciones que se encuentran fuera del intervalo, son completadas utilizando las detecciones QRS automáticas definidas en el punto anterior.



**Figura 3.** Anotación latido a latido sobre el canal de referencia.

### Reporte

Finalmente, se genera un archivo de anotación bajo el formato estándar usado por el *toolkit* de *physionet* (Goldberger *et al.*, 2000). Este archivo contiene la información necesaria para evaluar la señal ECG. La estructura del archivo generado contiene dos partes: los instantes de detección de la onda *R*, y la anotación de cada latido, utilizando tres códigos de anotación de latidos: Normal (N), Ectópico (V, PPVC *Premature ventricular contraction*), Otros (O).

### 3. MEDIDA DE DESEMPEÑO

Los resultados de la detección automática fueron comparados con las anotaciones obtenidas por el sistema desarrollado en este trabajo. Entonces es posible calcular el desempeño del detector de QRS de Pan & Tompkins (Pan & Tompkins, 1985) sobre registros de ECG de esfuerzo usando la sensibilidad (*Se*) y el valor predictivo positivo (*P+*) calculados como se muestra en las ecuaciones 1 y 2 respectivamente.

$$Se = \frac{VP}{VP + FN} \quad (1)$$

$$P+ = \frac{VP}{VP + FP'} \quad (2)$$

donde:

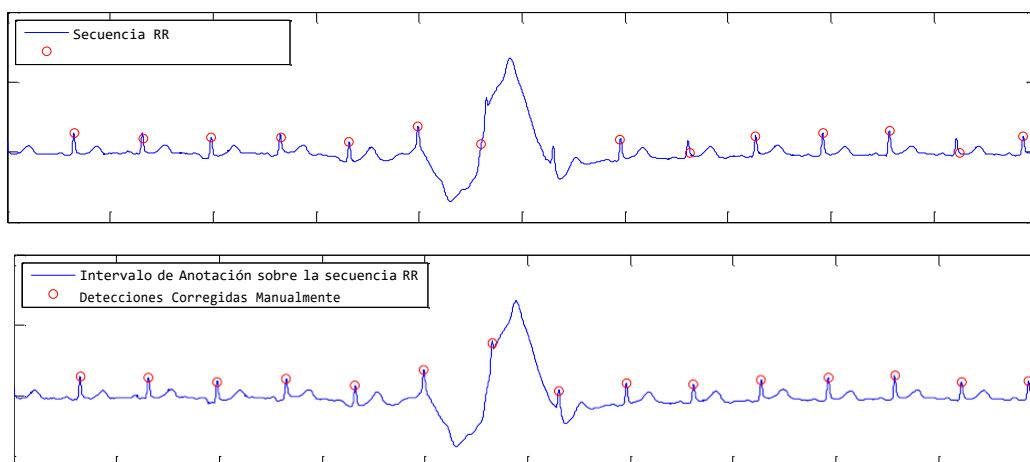
- *Verdaderos Positivos (VP)*: son los latidos o QRS detectados correctamente, se considera un QRS detectado correctamente si se detecta dentro de una ventana de 150 ms alrededor del punto fiducial.
- *Falsos Positivos (FP)*: son las detecciones señaladas donde no existe una anotación de latido. Este es el indicador de las falsas alarmas.
- *Falsos Negativos (FN)*: son los complejos QRS que no fueron detectados por el algoritmo.

### 4. RESULTADOS

Se desarrolló un sistema de anotación semiautomático tomando como parámetros lo siguiente: número de ventanas ( $NW = 20$ ), desviación estándar ( $SDT = 14$ ) y los valores de T1, T2 y T3 corresponden a 30%, 40% y 50% respectivamente. La Figura 4 muestra un ejemplo de anotación por intervalos de interés. En el panel superior se muestra un trazo del ECG donde se observó una discontinuidad sobre la serie *RR*. Al inspeccionar el *ECG* se observa que se trata de un artefacto. En el panel inferior se observa como el usuario corrige la detección sobre la serie.

La Tabla 1 muestra los resultados para el registro *ECG83*. El método automático de clasificación de Astudillo (Astudillo *et al.*, 2016), determinó las secuencias *RR* de los canales *C2*, *C4*, *C5*, *C6*, *C7* y *C8* como canales de referencia y los canales *C1* y *C3* como canales inservibles. Se validó el Canal 2 como el mejor y se anotó la señal sobre los intervalos de interés de esta secuencia *RR* (*C2*). Como se puede ver en la Tabla 1 el detector tuvo un mayor desempeño sobre el canal de referencia (*C2*) y un menor desempeño en los canales marcados como inservibles (*C1* y *C3*).

Se han anotado 10 ECG de esfuerzo de la Base de Datos DICARDIA. En todos los casos se consideraron señales de referencia o alta calidad y las anotaciones fueron realizadas sobre intervalos con artefactos de la serie *RR*. Estas anotaciones representan, considerando las ocho derivaciones, 172.025 latidos. La sensibilidad y el valor predictivo positivo sobre estos registros fue de  $Se=97.66\%$  y  $P+=96.71\%$  respectivamente.



**Figura 4.** Anotación sobre un intervalo de la secuencia RR.

El detector de QRS usado en este trabajo reporta valores de sensibilidad y valor predictivo positivo superiores al 99% para la base de datos MITDB, (Goldberger *et al.*, 2000); la cual cuenta con aproximadamente 110.000 latidos anotados. Los resultados obtenidos en este trabajo muestran un desempeño menor para el detector de Pan & Tompkins (1985), esto muestra la dificultad que representa la segmentación de un ECG de esfuerzo y la necesidad de adaptar los detectores de QRS a los ECG de esfuerzo.

**Tabla 1.** Desempeño del detector sobre el archivo de anotación ECG83

	C1	C2	C3	C4	C5	C6	C7	C8
Se	98.87	99.95	95.76	99.58	99.53	99.66	99.67	99.72
P+	96.91	99.90	98.39	99.35	99.26	99.40	98.54	99.49

## 5. CONCLUSIONES Y TRABAJOS FUTUROS

Se ha desarrollado un sistema que permite la anotación de un ECG de esfuerzo a partir de sus series *RR* calificadas como de referencia, es posible validar rápidamente registros ECG de al menos 20 min de duración, lo cual implica aproximadamente 1500 – 2500 latidos por derivación. Anotar tal cantidad de latidos requiere una gran cantidad de tiempo para el usuario. Hasta el momento se ha anotado más del 15% de la base de Datos DICARDIA, lo que constituye una cantidad de anotaciones comparables a las bases de datos electrocardiográficas de referencia (Goldberger *et al.*, 2000). La Base de Datos continúa en proceso de validación, para completar las anotaciones, pero, dado la cantidad de morfologías presentes, esta tarea requerirá de la colaboración de expertos en cardiología.

Si bien es cierto que esta metodología puede presentar algunas desventajas como la dificultad para evaluar el tiempo de retardo, presenta un valioso punto de partida para poder anotar de forma semiautomática extensas bases de datos electrocardiográficas como DICARDIA. Posteriormente estos resultados se usarán para evaluar y mejorar el desempeño de otros detectores de *QRS*, cómo por ejemplo, el basado en ondaletas (Martínez *et al.*, 2004).

## AGRADECIMIENTOS

Este trabajo fue parcialmente financiado por el programa PROMETEO de la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación de la República del Ecuador.

**REFERENCIAS**

- Almeida, D, S. Brandi, 1995. Manual de cardiología clínica. Caracas, Venezuela: Fondo Editorial. Acta Científica.
- Armijos, J., D. García, D. Astudillo, K. Palacio-Baus, R. Medina, S. Wong, 2015. *Semiautomatic validation of RR time series in an ECG stress test database*. In: 11th International Symposium on Medical Information Processing and Analysis (SIPAIM 2015), pp. 968119-968119.
- Astudillo, D., K. Palacio-Baus, L. Solano, R. Medina, S. Wong, 2016. *Characterizing artifacts in RR stress test time series*. In: Engineering in Medicine and Biology Society (EMBC), 2016 38th Annual International Conference of the IEEE (in press).
- Dokur, Z., T. Ölmez, 2001. ECG beat classification by a novel hybrid neural network. *Computer methods and programs in biomedicine*, 66(2), 167-181.
- Dumont, J., A.I. Hernandez, G. Carrault, 2010. Improving ECG beats delineation with an evolutionary optimization process. *Biomedical Engineering, IEEE Transactions on*, 57(3), 607-615.
- Goldberger, A.L., L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, H.E. Stanley, 2000. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23), e215-e220.
- Ledezma, C.A., E. Severeyn, G. Perpinan, M. Altuve, S. Wong, 2014. *A new on-line electrocardiographic records database and computer routines for data analysis*. In: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, pp. 2738-2741.
- Martínez, J.P., R. Almeida, S. Olmos, A.P. Rocha, P. Laguna, 2004. A wavelet-based ECG delineator: evaluation on standard databases. *Biomedical Engineering, IEEE Transactions on*, 51(4), 570-581.
- Moody, G.B., R.G. Mark, 1990. *The MIT-BIH arrhythmia database on CD-ROM and software for use with it*. In: Computers in Cardiology 1990, Proceedings IEEE, pp. 185-188.
- Pan, J., W.J. Tompkins, 1985. *A real-time QRS detection algorithm*. *Biomedical Engineering, IEEE Transactions on*, (3), 230-236.
- Quintero, L., S. Wong, R. Parra, J. Cruz, N. Antepara, D. Almeida, G. Passariello, 2007. *Stress ECG and laboratory database for the assessment of diabetic cardiovascular autonomic neuropathy*. In: Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, pp. 4339-4342.
- Romero, D., N. Behard, A. Martin-Yebra, J.P. Martinez, P. Laguna, E. Pueyo, A. Hernández, 2014. *Analysis of QRS alterations during stress test recordings on patients with brugada syndrome*. In: Computing in Cardiology Conference (CinC), IEEE, pp. 137-140.