

# Estudio Sobre la Fiabilidad de las Anotaciones en la Base de Datos QT de Physionet

F. González Molina<sup>1</sup>, R. Alcaraz Martínez<sup>2</sup>, J.J. Rieta Ibáñez<sup>1</sup>

<sup>1</sup>BioMIT.org, Dept. de Ingeniería Electrónica, Universitat Politècnica de Valencia, España, {fgmolina, jjrieta}@upv.es

<sup>2</sup>Grupo de Inv. en Electrónica, Telecom. y Bioingeniería, Univ. de Castilla-La Mancha, España, raul.alcaraz@uclm.es

## Resumen

*La base de datos QT (BDQT) es un referente en análisis electrocardiográfico gracias a dos factores: la amplia variedad de morfologías que contiene y las anotaciones manuales, que determinan los puntos fiduciales de un buen número de latidos. Por ello, es una referencia habitual en la validación de delineadores de ECG. Sin embargo, un porcentaje significativo de sus anotaciones se ha calificado como impreciso, lo cual hace que ningún algoritmo de delineación, por eficiente que sea, pueda conseguir un resultado óptimo. En este trabajo se analizan tanto las imprecisiones de anotación, como la influencia del ruido en el ECG durante la anotación. Se han estudiado las anotaciones en ondas P porque este efecto es mayor para ondas de baja amplitud. Estas anotaciones se compararon con el resultado de aplicar un delineador automático a la señal original con ruido y, después, reduciendo éste mediante un método eficiente basado en la transformada Wavelet. Los resultados mostraron que las anotaciones automáticas se encontraban más cerca de las manuales para el caso de señal ruidosa en un porcentaje significativo de los registros, evidenciando la influencia del ruido en la localización de las anotaciones manuales. Concretamente, esta mejora irreal en la delineación se obtuvo para el inicio, pico máximo y final de onda P, en el 45.83 %, 52.29 % y 56.25 % de los registros, respectivamente. Por tanto, para mejorar la fiabilidad de la validación de delineadores, se hace necesaria la revisión de las anotaciones de la BDQT o el reemplazo de ésta por otra mejor anotada.*

## 1. Introducción

Para poder ser usado en un contexto clínico, cualquier algoritmo de procesamiento de ECG debe ser validado mediante la evaluación de su rendimiento en un contexto lo más realista posible [1]. Para ello, es necesaria la disponibilidad de bases de datos lo suficientemente extensas como para que cubran la gran diversidad de patrones de ondas que los registros de ECG pueden presentar [1]. Además, para el caso específico de métodos como los delineadores de ondas, son necesarias anotaciones que definan de forma precisa los instantes de tiempo en los que ocurren estos eventos. Para poder confiar en las anotaciones, éstas deben haber sido minuciosamente realizadas de forma manual por expertos. Sin embargo, son muchos los factores que pueden influir negativamente en este proceso y, por tanto, comprometer su resultado.

Los investigadores tienen a su disposición un número considerable de bases de datos estándares de ECG, cuyas características propias están estrechamente ligadas al objetivo específico por el cual se crearon inicialmente. Dos de ellas son la base de datos de arritmias MIT-BIH [2] y la

base de datos AHA, que se desarrollaron para evaluar el desempeño de distintos detectores de arritmias. Por otro lado, la base de datos europea ST-T se creó como respuesta al creciente interés en el análisis del segmento ST-T como indicativo de isquemia miocárdica [3]. Sin embargo, ninguna de las bases de datos hasta ahora mencionadas contiene anotaciones manuales que definan los límites y el pico máximo de las ondas del ECG, algo fundamental para la validación de algoritmos automáticos de delineación. Así, respondiendo a esta necesidad, surgieron la base de datos de mediciones multiderivación CSE [4] y la base de datos QT (QTDB) [5]. Sin embargo, debido a características distintivas, como la posibilidad de acceso libre o una mayor cantidad de ondas anotadas de forma manual por expertos, esta última base de datos se ha posicionado como la referencia más usada para la validación de delineadores de onda P en los últimos años. No obstante, las anotaciones de la BDQT no son siempre precisas, lo cual puede llegar a generar desconfianza en su uso como vía de validación de algoritmos de procesamiento de ECG. Por ello, en este estudio se evaluará la calidad de las anotaciones de la BDQT, para el caso de la onda P, y se analizará el origen de los posibles errores en la localización de las mismas.

## 2. Base de datos QT

La BDQT, disponible gratuitamente en Physionet, está compuesta por un total de 105 registros de ECG de dos canales con una duración de quince minutos cada uno. Estas señales se seleccionaron de otras bases de datos existentes, específicamente, de forma que reflejasen la amplia variabilidad de morfologías que puedan presentar las ondas del ECG en el mundo real [5]. Además, esta base de datos contiene anotaciones manuales que marcan la posición de inicio, pico máximo y final de la onda P; inicio y final del complejo QRS; pico y final de la onda T y, si estuviese presente, pico y final de la onda U. Por cada registro, se proporcionan anotaciones de, al menos, 30 latidos seleccionados para representar el patrón de forma de onda predominante en la señal.

Esta base de datos consta de dos grupos de anotaciones realizadas por sendos expertos. Sin embargo, uno de los conjuntos contiene únicamente 11 registros anotados, por lo que rara vez se ha usado para validar algún método de delineación de ECG. Debido a ello, este estudio se centra en el análisis del grupo con mayor número de anotaciones.

El proceso de anotación manual de señales es una tarea lenta y compleja. Además, variables tales como la experiencia del anotador, su grado de concentración y cansancio, la fiabilidad y precisión de la herramienta de anotación usada, o la presencia de morfologías anómalas en la señal, pueden influir en el resultado final de la anotación, comprometiendo su exactitud.

Asimismo, las señales del ECG se ven afectadas en gran medida por diversas fuentes de ruido, como el producido por la actividad eléctrica de los músculos cercanos al corazón, la red eléctrica o un mal contacto del electrodo usado para la obtención de la señal [6]. El problema de estas perturbaciones es que pueden producir alteraciones de la auténtica morfología de la señal. Además, para ondas de baja amplitud, como la onda P, este efecto se acrecienta. Por ello, la onda P se ha seleccionado en el presente trabajo como objeto para la evaluación de las anotaciones manuales de la BDQT. Además, el estudio de esta onda ha despertado un creciente interés debido a la demostrada relación existente entre sus características morfológicas y ciertas afecciones cardíacas, como la recurrencia de la fibrilación auricular [7].

### 3. Análisis de la fiabilidad de la BDQT

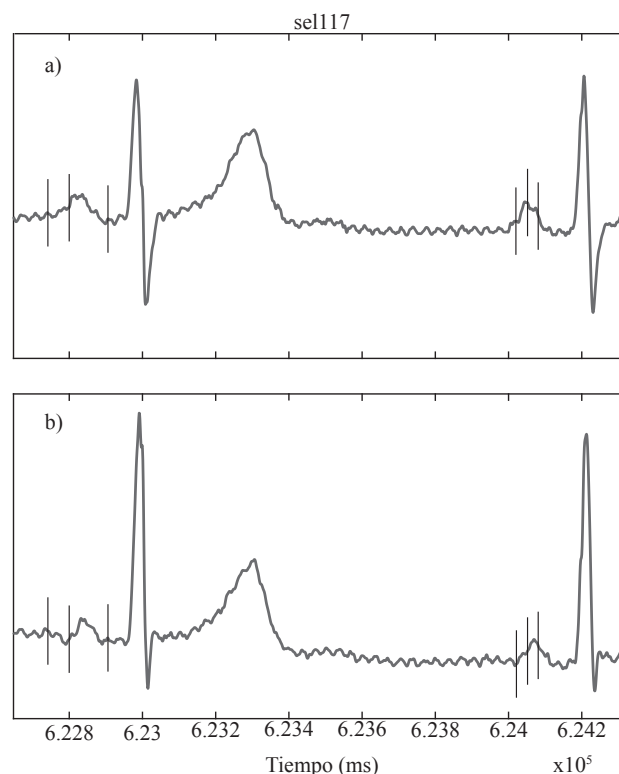
#### 3.1. Anotaciones multiderivación

En el proceso de anotación manual de los registros de la BDQT, se analizaron ambos canales simultáneamente, y la decisión sobre la posición de cada punto fiducial se tomó de forma que ésta fuese común para las dos derivaciones [5]. Sin embargo, es conocido que la disposición en el tiempo de las distintas ondas del ECG es diferente en función de la derivación considerada en cada caso, pues cada una corresponde a una proyección distinta del vector cardíaco [8]. Por tanto, este modo de proceder, unido a la falta de información acerca de cuál de los dos canales se utilizó como referencia de la anotación en cada caso, dificulta el uso de esta base de datos anotada como método de validación de algoritmos de delineación.

Para superar este obstáculo, los desarrolladores de este tipo de algoritmos se han inclinado por dos estrategias diferentes [9]. Por un lado, algunos han optado por asumir que para la anotación de cada registro se consideró únicamente una de las derivaciones, tomando así, como referencia para cada uno de ellos, el canal en el que las correspondientes al método automático se acercan más a las manuales. Por otro lado, otros desarrolladores han decidido tomar como referencia, para cada punto marcado manualmente en la base de datos, la anotación automática más cercana a él, sin importar a qué derivación corresponda. Ambas estrategias, de un modo u otro, están orientadas a obtener una ventaja injusta de la base de datos, debido a la forma en la que está anotada.

#### 3.2. Criterios de anotación variables

En la comunidad científica no existe un consenso claro acerca de la posición exacta de los límites de la onda P en el ECG. Por ello, al no conocer el criterio seguido por los

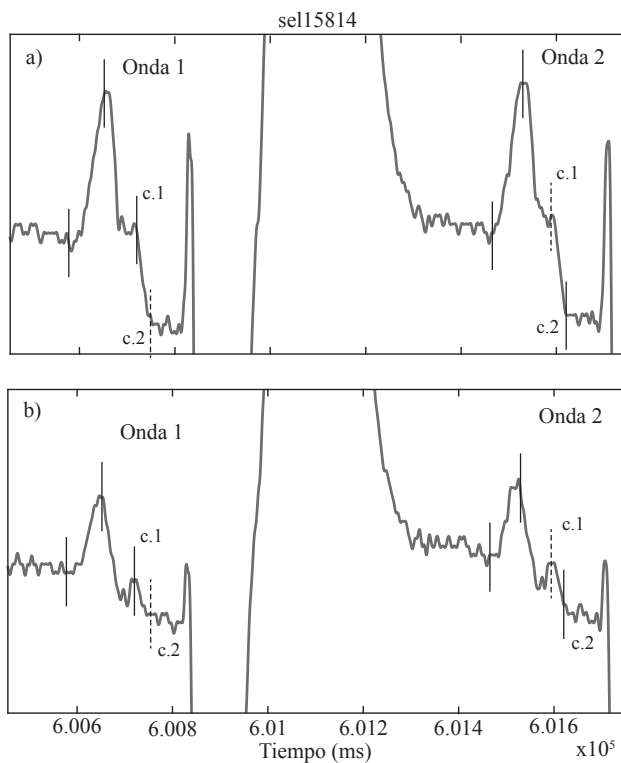


**Figura 1.** Ejemplo de los dos canales, a) y b), de un segmento de señal correspondiente al registro sel117 de la BDQT, en los que se muestran, sobre dos ondas P consecutivas, sus discutibles anotaciones manuales mediante líneas verticales. Concretamente, las ondas se corresponden a la 19ª y la 20ª onda P manualmente anotada en este registro. Obsérvense, especialmente, los finales de la segunda onda P en cada registro.

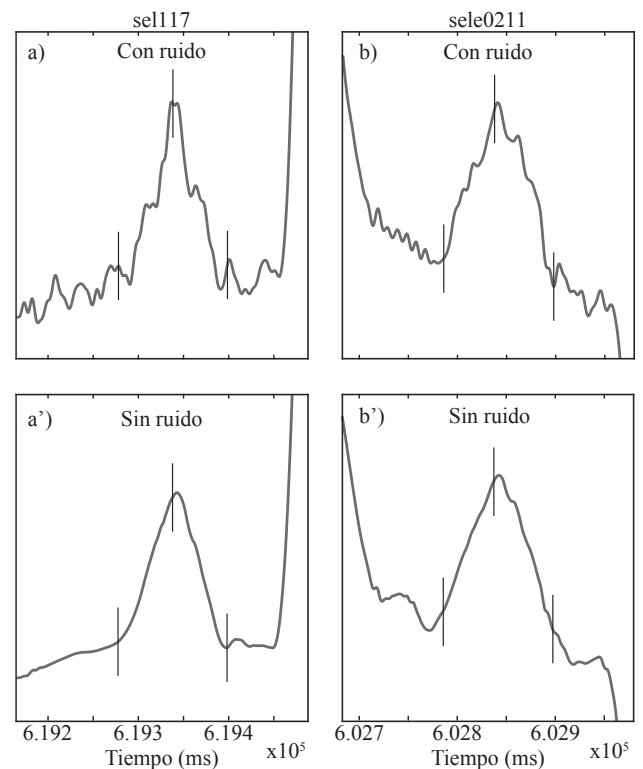
expertos para localizar los puntos fiduciales en esta onda, generalmente no es posible clasificar una anotación como errónea o imprecisa de forma categórica. Sin embargo, en algunos casos específicos, es difícilmente debatible que las anotaciones son manifiestamente defectuosas.

Uno de los casos en los que esto ocurre es cuando el punto anotado se encuentra significativamente lejos de la zona de la señal en la que una anotación puede ser considerada como correctamente localizada de forma intuitiva. En la Figura 1 se muestra un ejemplo de dos ondas P consecutivas anotadas que responden a esta situación. Se muestran ambos canales debido a la ya comentada falta de información específica acerca de cuál se tomó en cuenta durante el proceso de anotación.

Por otro lado, otra situación en la que el error en la anotación manual es incontestable, tiene lugar cuando se identifican criterios de anotación dispares en ondas cuya morfología es muy similar, tal y como se muestra en la Figura 2. En ella se puede observar cómo el final de dos ondas consecutivas se ha anotado tomando dos criterios dispares, etiquetados como c.1 y c.2. Así, para cada onda, se han marcado tanto la anotación manual original como la posición aproximada que hubiese tenido ésta si para cada caso se hubiese seguido el criterio alternativo. De esta manera, aunque ambos criterios pudieran ser aceptados como válidos, esta manifiesta discordancia evidencia un error en, al menos, una de las anotaciones.



**Figura 2.** Ejemplo de los dos canales, a) y b), de un segmento de señal correspondiente al registro sel15814 de la BDQT, en el que se muestra la disparidad de criterios seguidos para la anotación del final de las dos primeras ondas P de este registro. Con líneas continuas se marcan las anotaciones manuales y con discontinuas, la posición aproximada de la anotación si se siguiese el criterio alternativo, c.1 ó c.2., en cada caso



**Figura 3.** Ejemplo ilustrativo del efecto del ruido sobre las anotaciones manuales de la BDQT. En a) y b) se muestran, respectivamente, la 16ª onda P anotada del registro sel117 y la 5ª onda del registro sele0211. En a') y b') se presentan las ondas resultantes, junto con sus anotaciones, tras reducir la presencia de ruido en las mismas.

### 3.3. Importancia del ruido

Por último, se debe tener en cuenta que las bioseñales como el ECG, y especialmente las ondas P debido a su baja amplitud, se ven profundamente afectadas por ruido. Así, anotar señales que muestren altos niveles de ruido puede provocar desplazamientos indeseados en la posición de los puntos fiduciales marcados. Por ello, aunque los especialistas están entrenados para superar estas condiciones adversas, y muchas veces consiguen obviar la presencia de ruido; otras veces, éste hace que fallen en el proceso de anotación, pese a tomar decisiones similares. Esto puede ser fácilmente revelado por medio de un tratamiento simple de limpieza de ruido de la señal, tal y como se muestra en la Figura 3. En ella se exhiben dos ondas P anotadas, a) y b), y, bajo éstas, en a') y b'), las mismas ondas tras aplicarles un proceso de reducción de ruido basado en la transformada Wavelet [10]. Así, en a) se observan, sobre la señal ruidosa, unas anotaciones que podrían ser consideradas defectuosas, y, tras limpiar la señal, en a') comprobamos que éstas son bastante precisas. Por contra, en b) y b') se presenta el caso opuesto, en el que una onda con ruido pudiera parecer bien anotada y tras reducir la presencia de éste, estas anotaciones pasan a ser más discutibles.

## 4. Metodología y resultados

Todos los registros de la BDQT, en mayor o menor medida, están afectados por el ruido y, consecuentemente,

también sus anotaciones. Para cuantificar aproximadamente este hecho, se ha aplicado un delineador automático a las ondas P, antes y después de reducir la presencia de ruido en la señal mediante un método basado en la transformada Wavelet [10]. Después, para cada caso, se ha calculado el error absoluto medio entre las anotaciones automáticas y las manuales. Esto se ha llevado a cabo aplicando el algoritmo a los dos canales y calculando el error para cada punto fiducial como la diferencia temporal entre la anotación manual y la automática más cercana en cada caso.

Debido a su robustez bajo condiciones de ruido, para la delineación de ondas P se ha aplicado un nuevo método adaptativo basado en la generación de modelos Gaussianos de ondas P [11]. Se analizó un total de 96 registros de la BDQT, descartándose aquellos con menos de tres ondas P consecutivas anotadas.

Los resultados de este estudio mostraron que, en muchos casos, las posiciones de las anotaciones automáticas eran más próximas a las de las manuales para señales de ECG con ruido que sin él. Esta mejora irreal en el desempeño del algoritmo de delineación bajo condiciones de ruido se produjo para el inicio, pico máximo y final de las ondas P en un 45.83 %, 57.29 % y 56.25 % de los registros, respectivamente. Por lo tanto, se ha comprobado que la influencia del ruido y, quizás, otros aspectos que lleven a la incorrecta anotación de las ondas P, son altamente relevantes y deberían ser considerados seriamente.

## 5. Discusión

Como se ha constatado a lo largo de este estudio, la existencia de anotaciones imprecisas en las ondas P en los registros de la BDQT es un hecho innegable. Ya anteriormente algunos estudios han señalado esta circunstancia, asumiendo el error producido por estos defectos [12]. Asimismo, en otro estudio en el que se introdujo un nuevo delineador de onda P, se reportó una mejora significativa en los resultados obtenidos tras reanotar manualmente los registros de la BDQT [13].

Las anotaciones defectuosas no son la única fuente de error que afecta a la evaluación de los métodos automáticos de delineación. En casos donde se adoptaron criterios de anotación diferentes incluso para ondas de morfología similar, como el mostrado en la Figura 2, es imposible que un algoritmo automático consiga ser preciso siempre. Por ello, estas situaciones implicarán errores, aunque todas las anotaciones pudiesen ser justificadas como correctas.

También se han recalcado los inconvenientes que presenta una referencia anotada mediante una estrategia multiderivación, generando una mayor desconfianza en el resultado que obtenga un método comparado con esta referencia.

Por último, se ha demostrado que la presencia de ruido ha afectado considerablemente al emplazamiento de las anotaciones. Así, el resultado de la localización de los puntos fiduciales para cualquier método automático será distinto en caso de ser aplicado sobre señales con ruido que sin él. De hecho, la delineación automática bajo condiciones de ruido ha proporcionado un mejor resultado, con respecto a las anotaciones manuales de la BDQT, que la misma delineación aplicada sobre los registros limpios de ruido. Esto demuestra que, en muchos casos, un buen delineador capaz de ser completamente inmune al ruido podría verse perjudicado por ello, y otro peor, que acepte erróneamente el ruido como señal real, beneficiado. Es difícil conocer hasta qué punto estas anotaciones desplazadas han podido influir en los delineadores desarrollados en la literatura [9]. Sin embargo, lo que parece ser evidente es que este procedimiento de validación es deficiente, y que su mejora debe ser prioritaria para el desarrollo de métodos de delineación de las ondas del ECG fiables y útiles en la práctica clínica.

## 6. Conclusiones

En este estudio, se han evidenciado los defectos en las anotaciones manuales de la BDQT, generalmente aceptada para validar delineadores automáticos de ECG. Entre otros, el origen de estos errores es la falta de información sobre el canal específico anotado, la diversidad de criterios durante el proceso de anotación y/o el efecto que la presencia de ruido tuvo en el resultado final de anotación manual de los registros. Este último aspecto se ha cuantificado, demostrando que la relevancia del ruido es significativa y que puede afectar seriamente a los resultados de los diferentes delineadores desarrollados. Por todo ello, para mejorar el desarrollo de delineadores fiables y precisos, así como para poder confiar en su validación, es necesario que las anotaciones manuales de la BDQT sean revisadas y sus re-

gistros limpiados de ruido o que se adopte como referencia otra base de datos alternativa. Ésta debería contar con todas las características que se han señalado en este estudio como deseables: Una base de datos de ECGs extensa que presente una gran variedad de morfologías, con anotaciones manuales precisas sobre señales individuales en las que la influencia del ruido sea mínima.

## Agradecimientos

Trabajo financiado por los proyectos TEC2014-52250-R y DPI2017-83952-C3 MINECO/AEI/FEDER, UE.

## Referencias

- [1] Sörnmo L, Laguna P. Ch. 6 - the electrocardiogram-a brief background. In *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Academic Press. ISBN 978-0-12-437552-9, 2005; 411 – 452.
- [2] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine* May 2001;20(3):45–50. ISSN 0739-5175.
- [3] Taddei A, Distant G, Emdin M, Pisani P, Moody GB, Zeelenberg C, Marchesi C. The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *European Heart Journal* 1992;13(9):1164–1172.
- [4] Willems JL, Arnaud P, Bommel JHV, Bourdillon PJ, Degani R, Denis B, Graham I, et al. A reference data base for multilead electrocardiographic computer measurement programs. *Journal of the American College of Cardiology* 1987;10(6):1313 – 1321. ISSN 0735-1097.
- [5] Laguna P, Mark RG, Goldberg A, Moody GB. Database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In *Computers in Cardiology*. 1997; 673–676.
- [6] Bollmann A, Husser D, Mainardi L, Lombardi F, Langley P, Murray A, Rieta JJ, et al. Analysis of surface electrocardiograms in atrial fibrillation: techniques, research, and clinical applications. *Europace* Nov 2006;8:911–26.
- [7] Magnani JW, Williamson MA, Ellinor PT, Monahan KM, Benjamin EJ. P-wave indices: current status and future directions in epidemiology, clinical, and research applications. *Circ Arrh and electroph* Feb 2009;2:72–9.
- [8] Rieta JJ, Alcaraz R. The Genesis of the Electrocardiogram (ECG). In *Wiley Encyclopedia of Electrical and Electronics Engineering*. Hoboken, NJ, USA: John Wiley & Sons, Inc., February 2017; 1–15.
- [9] Beraza I, Romero I. Comparative study of algorithms for ECG segmentation. *Biomedical Signal Processing and Control* April 2017;34:166–173.
- [10] Bora PK, Sinha R, Yadav SK. Electrocardiogram signal denoising using non-local wavelet transform domain filtering. *IET Signal Processing* February 2015;9(1):88–96.
- [11] González F, Alcaraz R, Rieta JJ. Electrocardiographic P-wave delineation based on adaptive slope gaussian detection. In *Computing in Cardiology Conference (CinC)*, volume 44. IEEE, 2017; In press.
- [12] Lenis G, Pilia N, Oesterlein T, Luik A, Schmitt C, Dössel O. P-wave detection and delineation in the ECG based on the phase free stationary wavelet transform and using intracardiac atrial electrograms as reference. *Biomedizinische Technik* 2016;61(1):37–56.
- [13] Martínez A, Alcaraz R, Rieta JJ. Application of the phasor transform for automatic delineation of single-lead ECG fiducial points. *Physiol Meas* 2010;31(11):1467–1485.