

EVALUACIÓN DE ALGORITMOS DE DETECCIÓN DE COMPLEJOS QRS MEDIANTE LAS CURVAS DE FUNCIONAMIENTO ROC, DET Y EPC

RESUMEN

Se presenta una metodología para la selección de modelos utilizados en detección de eventos, empleando las curvas de funcionamiento *característica de operación del receptor (ROC - Receiver Operating Characteristic)*, *compensación del error de detección (DET - Detection Error Trade-off)* y curvas de desempeño esperado (*EPC - Expected Performance Curve*), las cuales asumen un criterio de mínimo error para evaluar modelos. Las curvas se evalúan sobre algoritmos de detección de complejos QRS en electrocardiografía utilizando la base de datos de arritmias del MIT [8]. Los resultados obtenidos muestran que la mejor curva para representar el comportamiento de los métodos de detección es la curva EPC debido a que utiliza pruebas sobre conjuntos de entrenamiento y validación. Igualmente se obtiene que el mejor detector de complejos QRS es el basado en la amplitud y la primera derivada AF3.

PALABRAS CLAVES: Selección de Modelos, curva ROC, curva DET, EPC, complejo QRS.

ABSTRACT

A methodology to select models used in detection is shown; it uses the performance curves named ROC, DET and EPC. These curves employ a criterion to evaluate the model based in obtaining a minimum error. Curves are applied over QRS complex detection algorithms using MIT Arrhythmia Database. Results show that the best curve for representing the behavior of the detection algorithms is the EPC curve, due to it uses training and test set. Equally, we obtained that the best QRS complex detector is AF3.

KEYWORDS: Model selection, ROC, DET, EPC.

1. INTRODUCCIÓN

En la tarea de detección de eventos, se aprecian dos tipos comunes de error, el detector puede dar tanto una falsa alarma como una detección fallida. La forma adecuada para presentar el funcionamiento del detector es generando una curva, para ver el comportamiento en los diferentes puntos de operación. El punto de operación de un modelo es un parámetro que puede ser ajustado dentro de este para llevar a cabo la toma de decisiones.

La curva ROC (Receiver Operating Characteristic) ha sido usada de forma tradicional para ajustar el punto de operación de un detector o para decidir cuál es el mejor en la toma de decisiones. Generalmente se grafica la tasa de falsas alarmas en el eje horizontal, mientras la detección correcta se grafica en el eje vertical. La curva DET (Detection Error Tradeoff) ha sido desarrollada para apreciar el funcionamiento del detector de forma más sencilla debido a que se grafica la desviación normal en ambos ejes, dando un tratamiento uniforme para ambos tipos de error y usa una escala logarítmica para ambos ejes; lo anterior mejora la forma de observar el

SANTIAGO SÁNCHEZ

Ingeniero Electricista
Universidad Tecnológica de Pereira
ssa@ohm.utp.edu.co

CLAUDIA PATRICIA HENAO

Ingeniera Electricista
Universidad Tecnológica de Pereira
claudiah@ohm.utp.edu.co

MAURICIO ALVAREZ

Ingeniero Electrónico, M.Sc
Profesor Auxiliar
Universidad Tecnológica de Pereira
malvarez@ohm.utp.edu.co

funcionamiento de los sistemas y produce gráficas que son cercanas a líneas rectas. Las curvas EPC (Curvas de desempeño esperado) se han introducido con el fin de buscar el modelo que presente menor costo de error y mostrar que las curvas ROC o DET deben utilizarse con conjuntos de prueba que representen las diferentes circunstancias a las cuales es expuesto el modelo; ya que estas mienten en ocasiones al comparar modelos con pruebas de desarrollo y luego con pruebas de validación.

Este documento provee en la sección dos una base de los métodos más usados para evaluar o analizar modelos y un algoritmo para promediar las curvas. La sección tres presenta una aplicación de estos modelos en detectores de complejos QRS de señales electrocardiográficas usando la base de datos de arritmias del MIT [8].

2. CURVAS DE FUNCIONAMIENTO

En la tarea de clasificación y detección de eventos, los modelos utilizados presentan en sus salidas dos tipos de error, estos pueden dar tanto falsas alarmas o falsos positivos, como falsos rechazos de un objetivo.

Generalmente los resultados del modelo se evalúan con la sensibilidad, precisión, fracción de falsas alarmas y verdaderos positivos o falsos rechazos. Donde la sensibilidad se considera como una medida de cuántos objetivos se detectan de todas las muestras analizadas por el modelo. Y la precisión es una medida de cuántos objetivos detectados o clasificados son verdaderos.

Mediante las curvas de funcionamiento se puede establecer si un modelo es conservativo o liberal. El conservativo es aquel que tiene mayor probabilidad a dar una verdadera aceptación en vez de una falsa aceptación y el liberal toma el caso contrario; además se puede identificar en qué rangos de un parámetro ajustable se presentan mayor número de falsos objetivos [1,2,3,4,5].

Las medidas de desempeño de los modelos se obtienen mediante la cuantificación de los siguientes resultados:

- VP: verdaderos positivos, son los objetivos correctamente detectados.
- FN: falsos negativos, son los objetivos que no fueron detectados.
- FP: falsos positivos, son los objetivos que el modelo marcó como objetivos detectados.
- VN: verdaderos negativos, son las objetivos que consideró no eran el objetivo.

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

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

$$FVP = \frac{VP}{VP + FN} \quad (3)$$

$$FFP = \frac{FP}{FP + VN} \quad (4)$$

Para la construcción de las curvas de funcionamiento basta con representar una pareja de medidas.

Un umbral o parámetro (γ) del modelo puede ser sintonizado de manera que decremente el número de FP y FN, lo que hace que FFP y FVP tiendan a cero y uno respectivamente. En la literatura general se escoge una única medida, la cual combina estos valores. Las medidas comunes presentes en la literatura para evaluar el desempeño son [1,5]:

- La medida F, es la media armónica de la precisión y la sensibilidad:

$$F = \frac{2 \cdot \text{Precisión} \cdot \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \quad (5)$$

- El punto de quiebre:

$$BEP = \frac{\text{Precisión}(\gamma^*) + \text{sensibilidad}(\gamma^*)}{2} \quad (6)$$

Tal que:

$$\gamma^* = \arg \min_{\gamma} |\text{Precisión}(\gamma) - \text{Sensibilidad}(\gamma)| \quad (7)$$

En lugar de precisión y sensibilidad se utiliza FVP y FFP.

El punto de quiebre es el punto donde la curva ROC intercepta la línea $y=x$ denominado la tasa de igual error (EER).

- Área bajo la curva ROC o EPC.

2.1. Curva Característica de Operación del Receptor (ROC)

Una gráfica ROC es una técnica para visualizar, organizar y seleccionar clasificadores basada en su propio funcionamiento. Este tipo de gráfica es de dos dimensiones en la cual la fracción de VP se dispone en el eje Y y la fracción de FP en el eje X. Una curva ROC presenta una compensación entre los beneficios (VP) y los costos (FP) de un modelo [1], en la figura 1 se observa una curva ROC típica.

Cuando se le aplica un conjunto de prueba a un modelo, este produce un conjunto de salidas, las cuales corresponden a un punto en la curva ROC. Así cuando cambiamos un parámetro del modelo, como por ejemplo un umbral, se puede generar la curva de funcionamiento.

El modelo óptimo es aquel cuya curva ROC supera la curva ROC de otro modelo [1,2]. Los puntos de interés en el espacio son: el punto (0,0) donde se considera el umbral o el parámetro variable del modelo es muy grande o infinito, el punto (1,1) donde la tasa de VP y la tasa de FP es muy alta y se presenta el peor funcionamiento.

Una medida de comparación de modelos a partir de la curva ROC es el área bajo la curva (AUC); donde el mejor modelo es aquel que presente una medida AUC mayor que otro. Para asumir una respuesta se debe observar el comportamiento del modelo y a criterio del diseñador tomar una decisión [1].

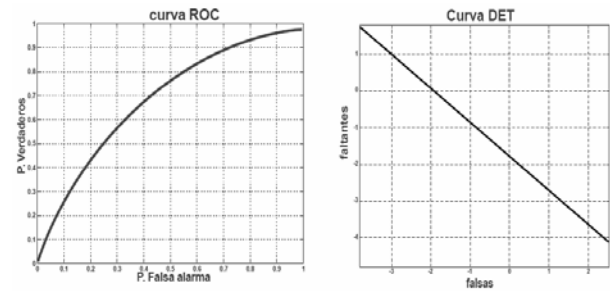


Figura 1. Curvas ROC y DET típicas.

2.2. Curva Compensación del Error de Detección (DET)

Durante la evaluación del sistema se debe determinar si la muestra desconocida es uno de los objetivos conocidos o no.

La salida del sistema es una probabilidad de que la muestra es un caso del objetivo o la información deseada. La probabilidad es usada para generar la curva de funcionamiento mostrando el rango de operación posible. Se supone que la distribución de probabilidad para objetivos y no objetivos son ambas normalmente distribuidas con sus medias μ_0 y μ_1 respectivamente y de igual varianza. Esto se puede apreciar en la figura 2.

La selección de un punto de operación x se muestra por una línea, los dos tipos de error antes mencionados son representados por áreas de las regiones para $w1$ y $w2$, falsas alarmas y faltantes respectivamente [3].

Para la curva DET en lugar de graficar las probabilidades de fallidos (FN) contra la de falsa alarma (FP), se grafica la desviación estándar que corresponde a las probabilidades. Un efecto de linealidad es resultado de la normalidad asumida de la densidad de probabilidad y una pendiente unitaria es consecuencia de las varianzas iguales en la distribución.

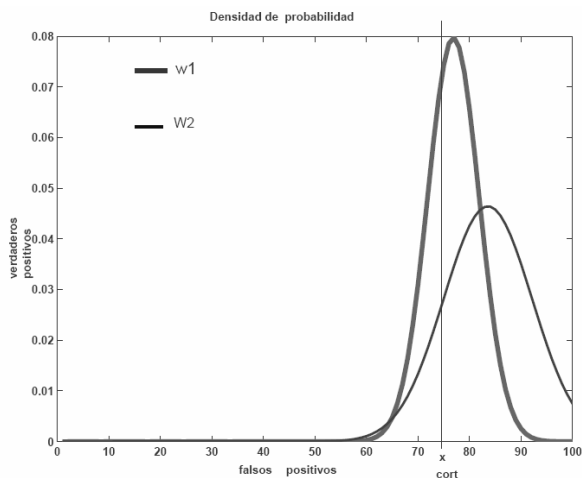


Figura 2. Distribuciones normales.

Como criterio de selección del mejor modelo se busca aquel que posea una curva DET donde se tenga menores tasas de error o se busca el modelo con el punto que minimiza la llamada función costo de detección (DCF), la ecuación de la DCF se presenta como sigue [3,5]:

$$DCF = \text{Costo}(FN)P(\text{objetivo})FFN + \text{Costo}(FP)P(\text{no objetivos})FFP \quad (8)$$

Un caso particular es conocido como la tasa media de error total con sus siglas en inglés (HTER), donde las probabilidades son iguales a 0.5 y los costos a 1.

$$HTER = \frac{FFN + FFP}{2} \quad (9)$$

2.2.1 Calculo de la Curva DET Compuesta

La curva compuesta se utiliza para promediar o combinar curvas DET o ROC [1,6].

El método transforma cada curva desde el espacio (fracción de falsos negativos, fracción de falsos positivos) en adelante (FFN, FFP) a coordenadas polares.

Cada tasa de error es función de un parámetro o umbral del modelo, estas tasas son normalizadas vía coordenadas polares y quedan con rango entre [0-1], cada curva en coordenadas polares es pesada por su número de muestras. Así como resultado la curva compuesta se ve como un promedio directo de las curvas [6].

2.3 Curvas de Desempeño Esperado (EPC)

Esta clase de curva se desarrolla con base en la optimización de un costo relativo de cada tipo de error (FPs y FNs) o una ganancia relativa de cada tipo de decisión (VPs y VNs). Tres series de curvas hacen parte de las EPC, que buscan seleccionar o implementar un modelo de acuerdo a algún criterio [5].

Como marco general las curvas EPC presentan el comportamiento de un modelo obtenido en un conjunto de datos de validación con respecto al ajuste de algún parámetro variable realizado en un conjunto ideal de pruebas. Este parámetro puede ser ajustado de varias maneras.

En una primera instancia el ajuste o la selección de un modelo se basa en el objetivo de minimizar o maximizar un criterio general tal como la función DCF o la medida F respectivamente. Otro criterio de selección es cuando se esté interesado en controlar una de las tasas de error que presentan los modelos.

Se selecciona el mejor modelo buscando que este presente una curva EPC inferior o superior en un amplio rango dependiendo del caso.

El algoritmo 1 describe el método para realizar la curva para la función de costo (DCF), una modificación se puede realizar para obtener la medida F y el algoritmo 2 presenta el método para generar la curva de funcionamiento esperado FFP y FFN.

Algoritmo 1. Método para obtener la curva DCF

- Sea *valid* el conjunto de validación
- Sea *prueba* el conjunto de prueba
- Suponga $FFP(\theta, prueba)$ es la FFP obtenida sobre el conjunto de prueba para el umbral θ
- **Para** valores $\alpha \in [0,1]$ **hacer**

$$\theta^* = \arg \min_{\theta} \left| \frac{\alpha \cdot FFP(\theta, prueba) + (1-\alpha) \cdot FFN(\theta, prueba)}{FFN(\theta, prueba)} \right|$$

Calcular $FFP(\theta^*, valid)$, $FFN(\theta^*, valid)$ y $HTER(\theta^*, valid)$.

Grafique $HTER(\theta^*, valid)$ con respecto a α .

Fin para

Algoritmo 2. Método para obtener la curva FFP y FFN

- Sea *valid* el conjunto de validación
- Sea *prueba* el conjunto de prueba
- Suponga $FFP(\theta, prueba)$ es la FFP obtenida sobre el conjunto de prueba para el umbral θ
- **Para** valores v de FFP esperado: entre 0.0 y 0.5 **hacer**

$$\theta^* = \arg \min_{\theta} |v - FP(\theta, prueba)|$$

Calcular $FFP(\theta^*, valid)$, $FFN(\theta^*, valid)$ y $HTER(\theta^*, valid)$.

Grafique $HTER(\theta^*, valid)$ con respecto a v .

Fin para

Para obtener $F(\theta^*, valid)$ respecto de v se busca el argumento que maximice $|v - FP(\theta, prueba)|$. En la región que se presente la curva menor para DCF o mayor para F este modelo es mejor.

Curvas de comportamiento obtenido contra esperado.

Con el objetivo de entender el comportamiento que toman los modelos, se puede comparar de cada medida de interés como fracción de falsos negativos (FFN) y fracción de falsos positivos (FFP); su valor esperado (en el conjunto de prueba) con el obtenido (conjunto de validación).

3. APLICACIÓN DE LAS CURVAS DE FUNCIONAMIENTO EN MODELOS DE DETECCIÓN

Los algoritmos de detección utilizados para el desarrollo de la prueba son los presentados en [7] llamados AF2, AF3, DF1, DF2, FS1, FS2, FD1 y FD2, los cuales detectan complejos QRS en señales electrocardiográficas, estos se basan en el ajuste de un umbral para asumir si un punto es candidato a complejo QRS o no.

Para la evaluación de los algoritmos se formaron dos conjuntos uno a priori donde se usa una señal normal tomada de la base de datos del MIT-BIH, la cual tiene una duración de 20 segundos. Además a esta señal se le adicionó cuatro clases de ruido con cuatro niveles de

relación señal a ruido (SNR). El otro conjunto de pruebas a posteriori utiliza 26 señales de la base de datos del MIT-BIH, con duración de 20 segundos.

En la figura 3 se presenta el funcionamiento con curvas ROC y DET del detector desarrollado por Ahstrom y Tompkins llamado FS2 ante interferencia causada por la línea de potencia (ac) del conjunto a priori.

Claramente se observa que el modelo tiene su mejor rendimiento cuando actúa en una condición de señal sin ruido (figura 3), y permite más falsas alarmas ante una SNR de 5dB, la curva ROC presenta este comportamiento; cuando se trabaja con umbrales bajos y alta SNR, las falsas alarmas del detector son pocas. A medida que se aumenta el umbral se tiene menor tasa de falsas alarmas y una tasa de detecciones que tiende a cero. De la curva DET es más fácil apreciar en qué momento el detector genera menores costos de detección dado a que esta toma los errores de detección en sus ejes; y como también asume las desviaciones estándar en lugar de las probabilidades, se presenta mayor claridad a la hora de ver el comportamiento del modelo.

En la figura 4 se presenta el comportamiento del detector FS2 ante las 26 señales electrocardiográficas del conjunto a posteriori, donde se muestra como el detector presenta para algunas señales características liberales.

Para tener un comportamiento general de cada uno de los detectores de complejos QRS en las pruebas a priori y posteriori; se realizó una normalización del umbral y se utilizó el algoritmo de cálculo de las curvas DET compuestas [6] descrito anteriormente. Esto se aprecia en la figura 5.

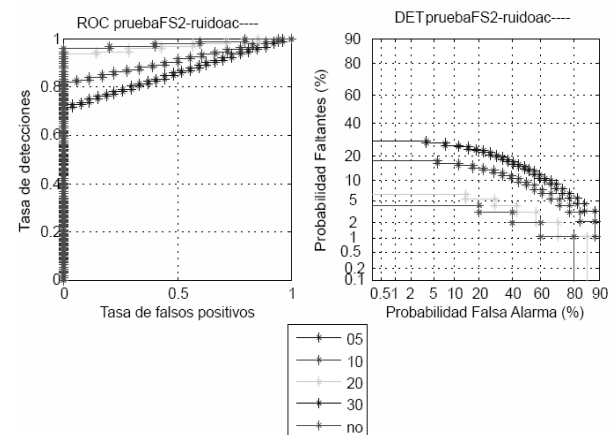


Figura 3. Curvas ROC y DET para FS2 ante ruido ac.

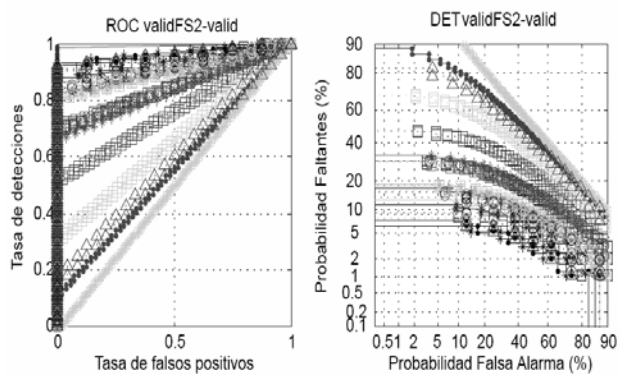


Figura 4. Curvas ROC y DET para FS2 en el conjunto de validación.

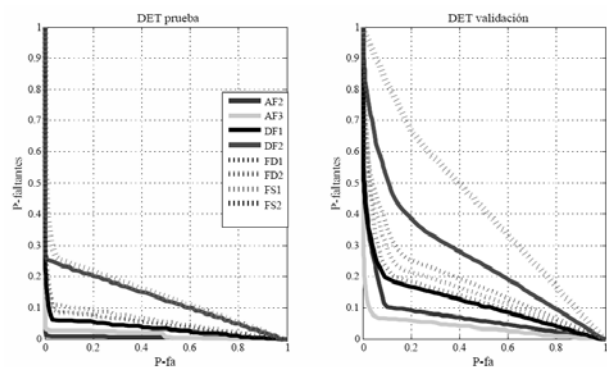


Figura 5. Curvas DET compuestas para los detectores.

En la figura 5 se aprecia que algunos detectores presentan curvas que se traslapan en varios rangos del umbral normalizado, lo cual genera conflicto a la hora de decir cuál es mejor, por esta razón se usan las curvas EPC. Las curvas EPC para todos los detectores se aprecian en las figuras 6, 7 y 8, donde se puede observar que el modelo AF3 es el que más probabilidades tiene de dar como resultado una detección correcta en vez de mentir a la hora de clasificar su salida, se observa que el comportamiento de este detector tiene menor valor de HTER en un amplio rango, menor tasa de falsas alarmas, menor tasa de falsos rechazos y en las curvas de tasa de falsas alarmas y falsos rechazos esperados contra obtenidos es el modelo que más se aproxima a la curva $y=x$; mostrando su robustez frente a diferentes clases de señales de ECG presentes en ambientes realistas.

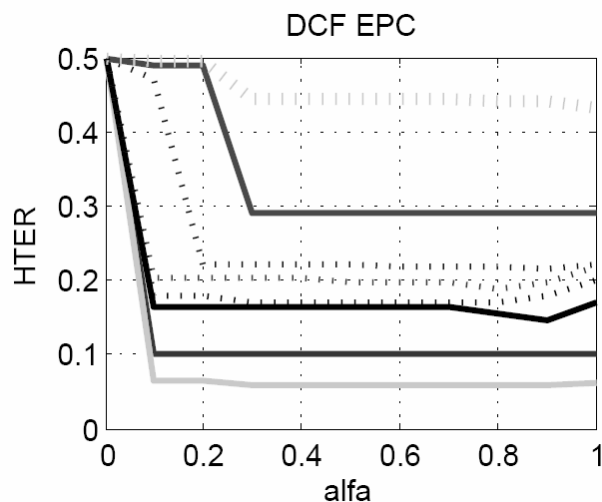


Figura 6. Curvas EPC DCF.

4. CONCLUSIONES

Mediante el uso de las curvas de funcionamiento se presenta una metodología que presta ayuda al ingeniero a la hora de decidir en cuál valor se debe ajustar el umbral de un modelo; para reducir una de las probabilidades de error de clasificación de clases o ajustar el parámetro en el punto de menor costo de error.

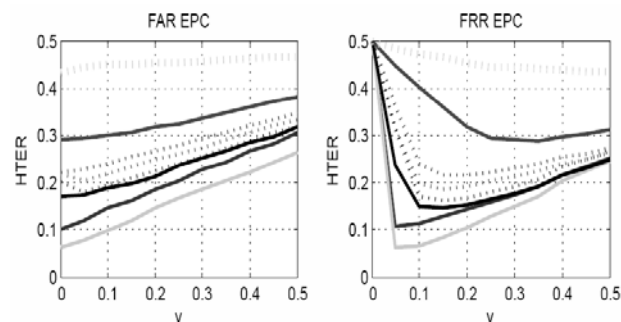


Figura 7. Curvas EPC tasa de falsas alarmas y falsos rechazos.

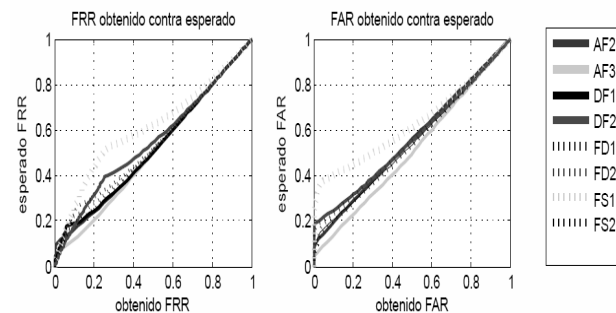


Figura 8. Curvas EPC tasas de errores esperado contra obtenido.

Resulta más sencillo sacar conclusiones y calificar un modelo mediante el uso de una curva de funcionamiento que evaluar éste en un único punto de operación.

Se presentan las ventajas de la curva DET sobre la curva ROC donde se toma la compensación de los dos errores y se hace más sencillo observar el comportamiento.

El método utilizado por Adler en [6] para promediar las curvas DET es más simple, útil y acertado que otros métodos presentados en la literatura por la manera en que mantiene la importancia de los dos ejes.

Las curvas de funcionamiento EPC presentan una ventaja notable sobre las demás curvas, dado a que tiene varias formas de selección de modelos; puesto que se puede dar mayor importancia a un tipo de error o a al costo de detección. Además se puede dar un criterio de acuerdo a una aplicación real. Y así una comparación más realista entre modelos.

5. AGRADECIMIENTOS

Este trabajo se desarrolló en el marco de los proyectos “Diseño y desarrollo de un sistema de asistencia médica para la clasificación de patologías en señales bioeléctricas”, financiado por el Centro de Investigaciones y Extensión de la Universidad Tecnológica de Pereira y el proyecto Sistema automatizado de clasificación de eventos fisiológicos a partir de patrones bioeléctricos como soporte en el tratamiento de la enfermedad de parkinson y otros desórdenes neurológicos financiado por Colciencias código 1110-14-17904.

6. BIBLIOGRAFÍA

- [1] FAWCETT, Tom. An Introduction to ROC Analysis. Pattern recognition letters. 27 pags 861-874. 2006.
- [2] A, Hero. Digital Signal Processing handbook. Chapter 13 Signal Detection and Clasification.
- [3] A, Martin. G, Doddington. T, Kamm. M, Ordowski. M, Przybocki. The DET Curve in Assessment of Detection Task Performance. National Institute of Standars and Technology. EuroSpeech 1997 Proceedings. Vol 4. 1997.
- [4] O, Lewis. HARVEY, Jr. Detection Sensitivity and Response Bias. Department of Psychology. University of Colorado. 2003.
- [5] BENGIO, Samy. MARIÉTHOZ, Johnny. The Expected Performance Curve: a New Assesment Measure for Person Authentication, IDIAP Research report, 2004.

- [6] ADLER, Andy. SCHUKERS, Michael E. Calculation of a Composite DET curve, University of Ottawa Research report, Canada.
- [7] G. M. Friesen. C, Thomas. A Comparition of the Noise Sensitivity of Nine QRS Detection Algorithms. IEEE Transactions on Biomedical Engineering. Vol 31. No 3. December 1986. pags 85-98.
- [8] PhysioNet. MIT-BIH Arrythmia Database. [en línea]. <<http://www.physionet.org/physiobank/database/mitdb/>> . [citado en 20 de marzo de 2007].