



Determinación de consonantes fricativas mediante indicadores de densidad de cruces por cero y de energía de las sílabas

Yoandy Pérez Peñate, Carlos A. Ferrer Riesgo, Yadisley Mujica Faije

RESUMEN / ABSTRACT

Se propone un algoritmo para la clasificación de sílabas en fricativas o no fricativas. En estudios anteriores se han descrito diferentes indicadores, como son el Indicador de Energía de la Sílabas (IES) y el Indicador de Sonoridad de la Sílabas (ISS) destinados a la caracterización y clasificación de consonantes. Estos indicadores separan bien las explosivas sordas de las explosivas sonoras y las consonantes nasales. Sin embargo no son capaces de resolver del todo las diferencias entre las explosivas sordas y las consonantes fricativas. En este trabajo se persigue separar estos dos tipos de consonantes. El algoritmo propuesto relaciona el Indicador de Energía de la Sílabas (IES) con la Densidad de Cruces por Cero Modificado (DCCM). Se obtienen resultados satisfactorios.

Palabras claves: Consonantes fricativas, cruces por cero, articulación.

A fricative/non-fricative consonant classification algorithm for monosyllables is proposed. Some indexes, like Syllable Energy Index and syllable Sonority Index, have been previously proposed to characterize and classify consonants. Their main drawback has been the weak separation between fricatives and voiceless plosives. This work attempts to separate fricatives and voiceless plosive consonants by means of a Zero Crossing measure. The proposal is evaluated on monosyllabic speech samples and results are satisfactory.

Key words: Fricative consonants, Zero Crossings, Articulation

Título en Inglés: Fricative consonant classification by means of zero crossing and energy measures

INTRODUCCION

Los Trastornos Motores del Lenguaje (TML) son trastornos neurológicos que afectan el correcto funcionamiento de los sistemas que participan en la producción de la voz. Una de las imprecisiones más frecuentes en presencia de estos trastornos es la sustitución de las consonantes explosivas por sus versiones fricativas o africadas. El hecho de desarrollar un sistema capaz de evaluar si estas consonantes fueron o no bien articuladas, puede contribuir al desarrollo de métodos objetivos y científicos en la evaluación y seguimiento de pacientes con problemas en el habla.

Desde mediados del siglo pasado varios estudios han investigado las características de las consonantes fricativas, tal es el caso de [1], [2]. Usando experimentos con voces simuladas, análisis de sílabas y experimentos de reconocimientos, estos estudios ofrecen datos sobre las características acústicas de las fricativas.

Estudios posteriores [3], han aumentado nuestros conocimientos sobre las fricativas. Algunos estudios más recientes [4], [5] y [6] nos permiten una mayor profundidad en el tema. Además existen varios estudios destinados a la determinación y caracterización de consonantes [7-9].

Los cruces por cero indican el número de veces que una señal continua toma el valor de cero. Para las señales discretas, un cruce por cero ocurre cuando dos muestras consecutivas difieren de signo, o bien una muestra toma el valor de cero.

Consecuentemente, las señales con mayor frecuencia presentan un mayor valor de esta característica, el ruido también genera un gran número de cruces por cero.

Partiendo de la hipótesis de que los segmentos fricativos poseen una alta Densidad de Cruces por Cero (DCC) con respecto a los no fricativos; y que cada tipo de consonante divididas en: fricativas (f, s, j), explosivas sordas (p, t, k), explosivas sonoras (b, d, g) y nasales (m, n), tienen diferencias en cuanto a su IES [9], el presente trabajo tiene como objetivo separar las sílabas en fricativas o no fricativas usando estos indicadores clasificar.

MATERIALES Y MÉTODOS

Detector de Sílabas

Para efectuar el cálculo de los indicadores se requiere ubicar las sílabas dentro de la grabación, y en particular, la transición de la consonante a la vocal. Con este propósito se emplea el detector de sílabas reportado en 10, descrito sintéticamente a continuación.

El detector se basa en la detección de máximos en la envolvente de la señal de energía de la señal, filtrada pasabajos a 50 Hz. Los máximos que no satisfacen ciertas condiciones para ser considerados como una sílaba son eliminados y el resto son considerados como centro de sílabas. Las condiciones usadas son las siguientes:

- Amplitud mayor que veinte veces el menor valor de la envolvente de energía.
 - Separación entre máximos mayor que 100 ms (cuando dos máximos están más cercanos que este valor el que posee menor amplitud es eliminado).
 - La presencia de un mínimo de menos del 75 % de la máxima amplitud entre él y los máximos posteriores y anteriores.
 - Separación entre los mencionados mínimos de más de 50 ms.

Una vez que las sílabas son detectadas el instante de liberación es determinado como el punto donde la envolvente de la energía posee su mayor valor de pendiente positiva en el segmento de tiempo entre el centro de esta sílaba y el centro de la anterior. Este método de determinación de la sílaba, tiene un 89% de detección correcta en un conjunto sílabas de pacientes disártricos [9].

Indicador de Energía de la Sílaba (IES)

El IES [9] se define a partir de la relación entre el área de la envolvente de energía $E(t)$ antes y después de la liberación de la constricción (1):

$$IES = \frac{A_b}{A_a} = \frac{\sum_{T_{on}-T_r}^{T_{on}-T_r-T_a} E(t)}{\sum_{T_{on}+T_r}^{T_{on}+T_r+T_a} E(t)} \quad (1)$$

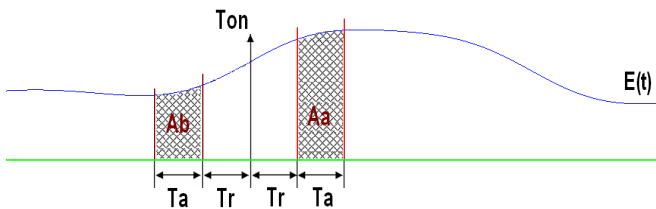


Fig. 1. Representación gráfica de las áreas y los instantes de tiempo utilizados en el cálculo de IES.

Los elementos que se tienen en cuenta en el cálculo de IES se muestran en la Fig 1. El instante de liberación de la consonante, o lo que es lo mismo el inicio de la vocal se denota como T_{on} y los valores de los intervalos de tiempo T_r (tiempo de guarda) y T_a (tiempo de análisis) son ambos de 20 milisegundos.

Densidad de Cruces por Cero (DCC)

La DCC de cada sílaba no es más que el cociente entre la cantidad de cruces por cero del segmento de 20 ms correspondiente a la consonante y la cantidad de cruces por cero del segmento de 20 ms de la vocal, representados en la Fig. 2

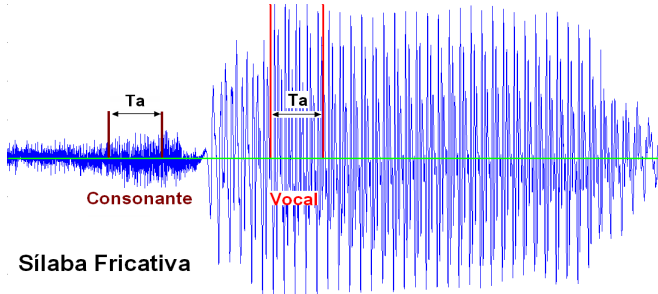


Fig. 2. Representación de una sílaba fricativa. Se observa que el segmento de la consonante existe una mayor cantidad de cruces por cero en relación al segmento representado dentro de la vocal. Por lo que es de esperar una DCC alta.

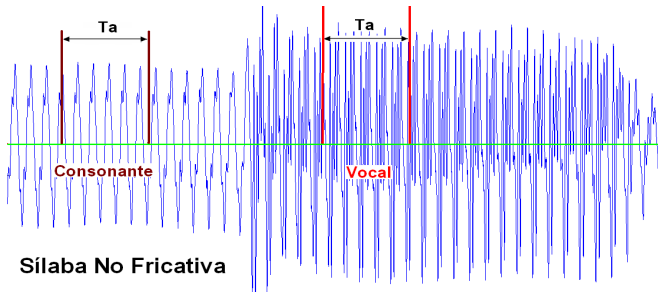


Fig. 3. Representación de una sílaba no fricativa. Se aprecia que a diferencia de las sílabas fricativas ahora el segmento de la consonante tiene una menor cantidad de cruces por cero en relación al segmento representado dentro de la vocal. Por lo que es de esperar una DCC baja.

Lo mostrado en las Fig. 2 y 3 corresponde a la sílaba 'sa' y 'ma' respectivamente, o sea una consonante fricativa y una nasal, pronunciadas por una misma persona.

Densidad de Cruces por Cero Modificado (DCCM)

Un problema semejante a la deriva de la línea base presente en la señal electrocardiográfica puede ocurrir en la señal de audio, como se ilustra en la Fig. 4. Esto atentaría en contra de una correcta clasificación de las sílabas afectadas.

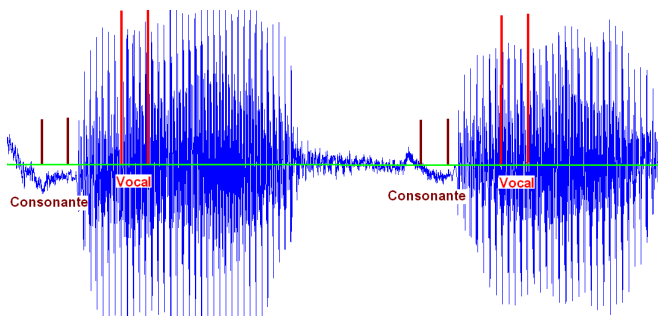


Fig. 4. En las sílabas que ocurra este problema, la determinación de los cruces por cero no aportará resultados efectivos y la clasificación de estas sílabas será incorrecta.

Con el objetivo reducir este inconveniente en el cálculo de los Cruces por Cero:

- Se divide cada segmento de consonante en cinco tramas
- En cada una se determina por separado su nivel medio
- Se calcula en cada caso los cruces por esta media

- La suma es el valor el utilizado para la implementación del indicador DCCM.

En lo demás es igual al indicador DCC. Este será el que se el utilizado en la separación del las sílabas en fricativos o no.

Señales empleadas

Se emplearon grabaciones de sujetos sanos. Se contó con grabaciones de 20 locutores, la mitad de ellos hombres y la otra mitad mujeres, todos sanos y entre 20 y 35 años de edad. Cada grabación consiste en la repetición sucesiva de una misma consonante. Se efectuaron grabaciones de 11 consonantes diferentes. (Ver Tabla 1). En cada grabación se detectaron las sílabas existentes y a cada una se le calculó el IES y la DCCM. Los resultados se fueron agrupando por tipos de consonantes para luego buscar la mejor forma de separarlas en fricativas o no.

TABLA 1. Grupos de consonantes analizados.

Grupos de Consonantes	Labial	Dental	Palatal
Explosivas Sordas	P	T	K
Explosivas Sonoras	B	D	G
Fricativas	F	S	J
Nasales	M	N	-

La adquisición se efectuó con el sistema de sonido de la PC con un micrófono Hama CS-460. La frecuencia de muestreo utilizada fue de 22050 Hz, con 16 bits de resolución.

RESULTADOS Y DISCUSIÓN

Separación mediante la DCCM

Utilizando solamente la DCCM no es posible separar a las fricativas de las demás consonantes (Fig. 5), ya que no todas las fricativas arrojaron una alta DCCM como se desearía. El mayor solapamiento con las fricativas lo tienen las explosivas sordas y en menor medida las explosivas sonoras; ya que con respecto a las nasales colocando un umbral con DCCM por ejemplo en DCCM=0.7, se lograrían separar de las fricativas con aceptable efectividad.

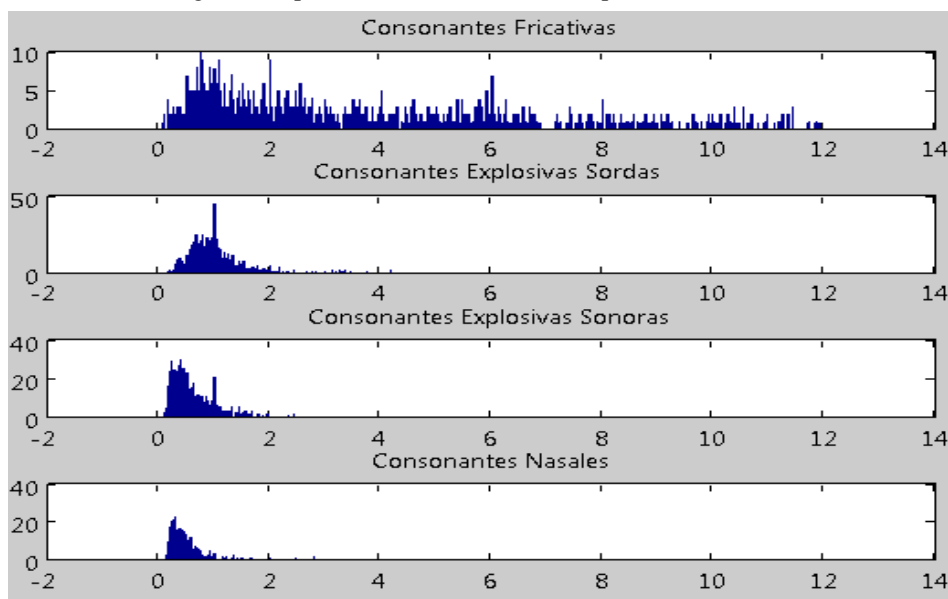


Fig. 5. Histograma de la Densidad de Cruces por Cero Modificado de cada tipo de consonante.

Separación mediante el IES

En la Fig. 6 se aprecia la diferencia existente entre cada tipo de consonante en cuanto al IES. Se corrobora que las explosivas sonoras así como las explosivas sordas tienen un IES bajo, que las fricativas un IES pudiera decirse que medio y que las nasales se pueden considerar con IES alto [9].

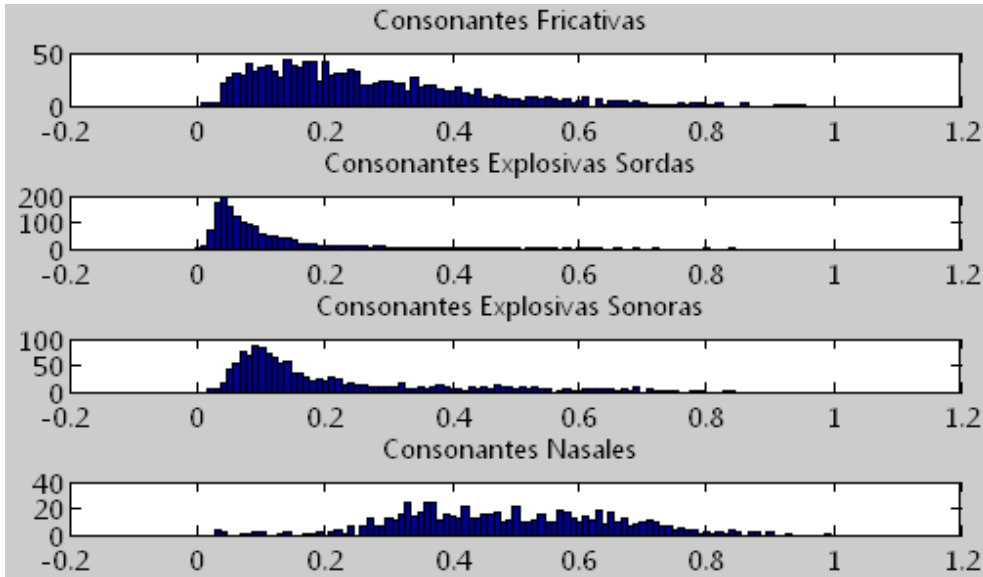


Fig. 6. Histograma IES de cada sílaba para cada tipo de consonante.

Combinación del IES y la DCCM para la separación de las consonantes fricativas

A partir de un análisis visual de la distribución de las sílabas de cada tipo de consonante en un plano IES vs DCCM se decidió crear una frontera de clasificación conformada por rectas definidas por los valores $x1$, $x2$ y $y1$. En las Figuras 7 y 8 puede entenderse la heurística de esta selección.

La función de la recta $x=x1$ es separar las fricativas de las nasales, mientras que la región limitada por las rectas $x=x2$ y $y=y1$ permiten separar las fricativas de las explosivas sordas, que quedarían siempre por debajo de estos valores.

Mediante ajustes manuales se arribó a los siguientes valores de las tres rectas de interés: $x1=0.7$, $x2=2$, y $y1=0.15$.

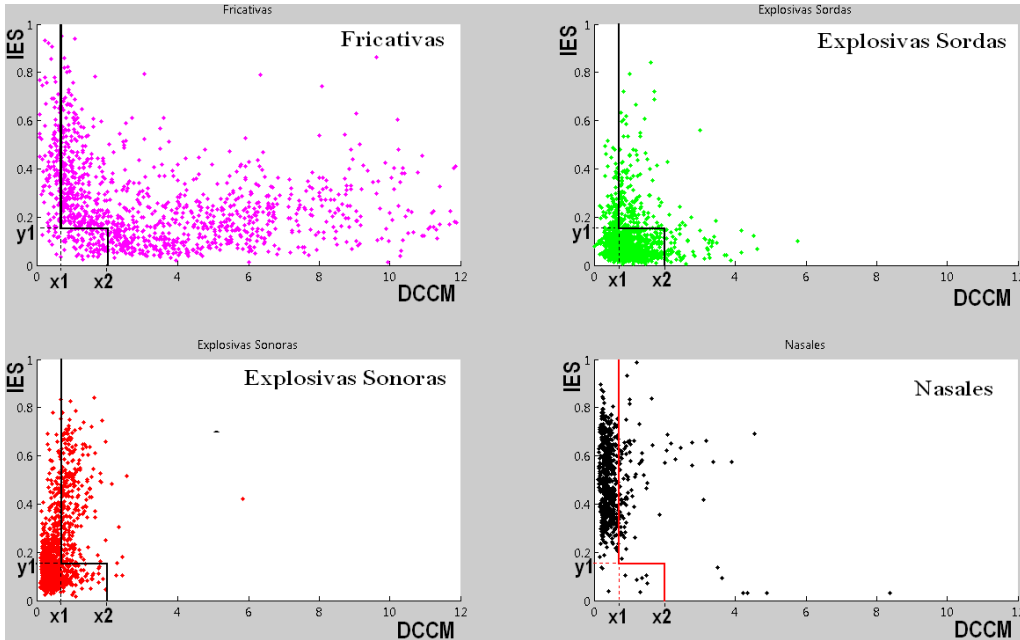


Fig. 7. IES vs DCCM para cada grupo de consonantes. Donde $x_1 = 0.7$, $x_2 = 2$, y $y_1 = 0.15$

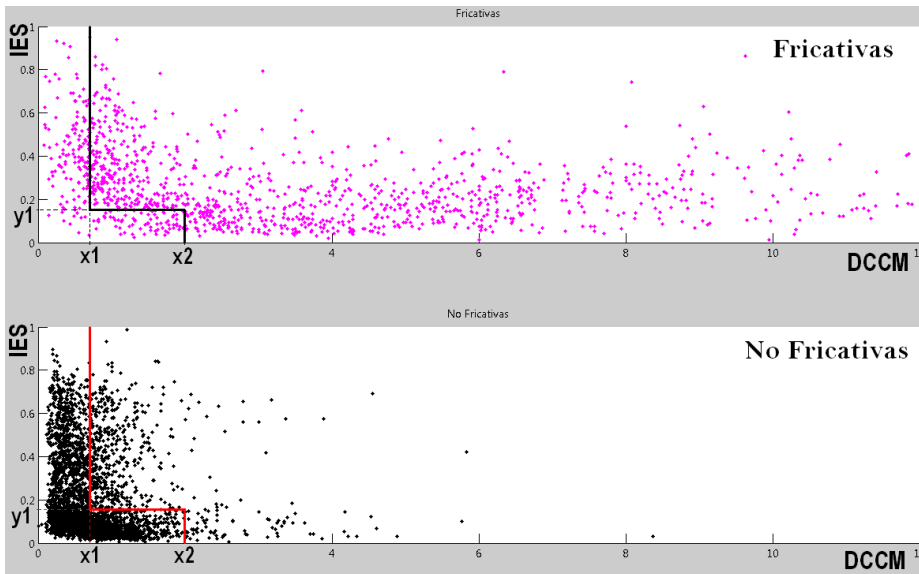


Fig. 8. IES vs DCCM para consonantes fricativas y no fricativas por separado. Donde $x_1 = 0.7$, $x_2 = 2$, y $y_1 = 0.15$

Cálculo de sensibilidad y especificidad

Los resultados obtenidos pueden ser medidos en términos de sensibilidad y especificidad, estas son definidas como la proporción de sujetos en cada subclase, correctamente clasificados por la prueba, donde:

Verdadero Positivo (VP): Sílabas fricativas correctamente clasificadas.

Falso Negativo (FN): Sílabas fricativas que por error fueron clasificadas como no fricativas.

Verdadero Negativo (VN): Sílabas no fricativas clasificadas como tales.

Falso Positivo (FP): Sílabas no fricativas clasificadas como fricativas.

La sensibilidad (*Sens*) se define como la probabilidad del algoritmo de detectar correctamente el evento positivo.

$$Sens = \frac{VP}{VP + FN}$$

La especificidad se define como la probabilidad del proceder para detectar los eventos negativos como tales.

$$Espec = \frac{VN}{VN + FP}$$

Luego de experimentos para la ubicación final de las líneas divisorias entre las consonantes fricativas y las no fricativas Fig. 7 y 8, y buscando una relación de compromiso entre la sensibilidad y la especificidad con $x1=0.7$, $x2=2$, y $y1=0.15$ se logró:

Sensibilidad Total = 83.52 %

Especificidad Total = 83.30 %

Tasa de Error = $(FP+FN)/(VP+VN+FP+FN) = 16.64$ %

Tasa de Clasif. = $(VP+VN)/(VP+VN+FP+FN) = 83.35$ %

También se determinó la especificidad en la separación como no fricativas, de los distintos tipos de consonantes por separado:

Especificidad Nasales = 86.51 %

Especificidad Explos. Sordas = 84.45 %

Especificidad Explos. Sonoras = 80.11 %

CONCLUSIONES

Se han logrado valores de sensibilidad y especificidad satisfactorios, sobre todo partiendo de que la efectividad del detector de sílabas es de un 89%. Los resultados son alentadores en cuanto a la posibilidad de separar las consonantes explosivas sordas de las fricativas, algo que no se había logrado con indicadores de energía y sonoridad [9].

Como trabajo futuro puede mencionarse el empleo de alguna técnica de optimización para la estimación de mejores valores de $x1$, $x2$ y $y1$. En este sentido ya se evalúa el empleo de algoritmos genéticos. Los resultados actuales también pueden ser mejorados mediante el empleo de combinaciones con otros índices, basados en características espectrales como relaciones de energías entre bandas de frecuencia u otros.

REFERENCIAS

1. **Hughes, G. W. and Halle, M.**, "Spectral properties of fricative Consonants: Journal of the Acoustical Society of America, 28, pp. 303-310, 1956.
2. **Manrique, A. M. B., and Massone, M. I.**, "Acoustic analysis and perception of Spanish fricative consonants", Journal of the Acoustical Society of America, 69, pp. 1145-1153, 1981.
3. **McCasland, G. P.**, "Noise intensity and spectrum cues for spoken fricatives", Journal of the Acoustical Society of America, Suppl.1, 65, S78-S79, 1979.
4. **Ali, A. M. A.**, "Acoustic features of the automatic recognition of fricatives", Technical Report, TR-CST27AUG97, Center for Sensor Technologies, University of Pennsylvania, 1997.
5. **Klatt, D.H and Klatt, L. C.**, "Analysis, synthesis and perception of voice quality variations among female and male talkers", Journal of the Acoustical Society of America, 87, pp. 820-857, 1990.
6. **Stevens, K. N., et all**, "Acoustic and perceptual characteristic of voicing in fricatives and fricative clusters", Journal of the Acoustical Society of America, 91, pp. 2979-3000, 1992.

7. **Ferrer, C., González, E.** “Determinación del grado de imprecisión de consonantes explosivas empleando niveles de energía y sonoridad”. Bioingeniería y Física Médica Cubana. Vol 5, 2004.
8. **González E., Ferrer, C. y & Hernández M. E.** “Algoritmo para la Detección de Consonantes Fricativas“. SIE´ 2005.
9. **Llanes, R., Ferrer, C., Díaz, R., Torres D.** “Caracterización de consonantes mediante indicadores de energía y sonoridad”. Reconocimiento de Patrones e Ingeniería Biomédica, 2008.
10. **Ferrer, C.; Hernández, M.E. & González, E.** Isolated Syllable Position Detector in Recordings of Patients With Motor Speech Disorders Using Speech Processing Techniques. Proceedings of the TELEC’02 International Conference, Santiago de Cuba, July. (2002).

AUTORES

Yoandy Pérez Peñate, Ing. en Telecomunicaciones y Electrónica. Centro de Ingeniería Clínica y Electromedicina. Villa Clara. (Teléfono: 271465). Especialista Superior A en Sistemas de Electromedicina. Miembro de la Sociedad Cubana de Bioingeniería, desde 2007.