

Clasificación de Señales Cerebrales Durante la Ejecución de Actividad Motora Imaginaria.

JAIME FERNANDO DELGADO SAA,
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA,
UNIVERSIDAD DEL NORTE, COLOMBIA
jadelgado@uninorte.edu.co

RESUMEN

Este trabajo describe algoritmos para clasificación de señales EEG con aplicaciones a interfaces cerebro computadora. La extracción de características se realiza mediante estimación de la densidad espectral de potencia de la señal EEG empleando métodos paramétricos y no-paramétricos. La selección de las características más importantes de la señal se realiza por medio del índice de Fisher, que es una medida de la separabilidad entre clases dando lugar a la construcción de un vector de características óptimo que permita la clasificación correcta de la actividad cerebral realizada por el usuario. Se presentan en este trabajo tres métodos de clasificación, Máquinas de vector de soporte, Análisis discriminante lineal y Redes neuronales artificiales. Los resultados obtenidos empleando data-set de la competencia BCI 2003 son comparados basados en precisión en la clasificación y el índice de Kappa Cohen. Los resultados muestran que los métodos paramétricos proveen un desempeño mayor a métodos clásicos basados en FFT en este tipo de aplicaciones, así como la conveniencia de la técnica de análisis discriminante lineal por desempeño y simplicidad.

Palabras claves: Interfaz cerebro computadora, BCI, clasificación, estimación espectral de potencia, Cerebro, reconocimiento de patrones, modelos auto regresivos.

ABSTRACT

In this work a system for classification of brain signals is described. The features employed involve the estimation of the power spectral density PSD of the EEG signal during execution of mental activity. Parametric and non-parametric methods are presented for estimation of PSD. The feature selection is done by comparing Fisher score values in different frequency bands. Finally the classification is done by three different methods for comparison. In this work data for the BCI competition 2003 was employed the results show the superiority of the parametric methods for this application.

Keywords: BCI, LDA,SVM, Brain, Autoregressive Models, classifiers.

1. INTRODUCCIÓN

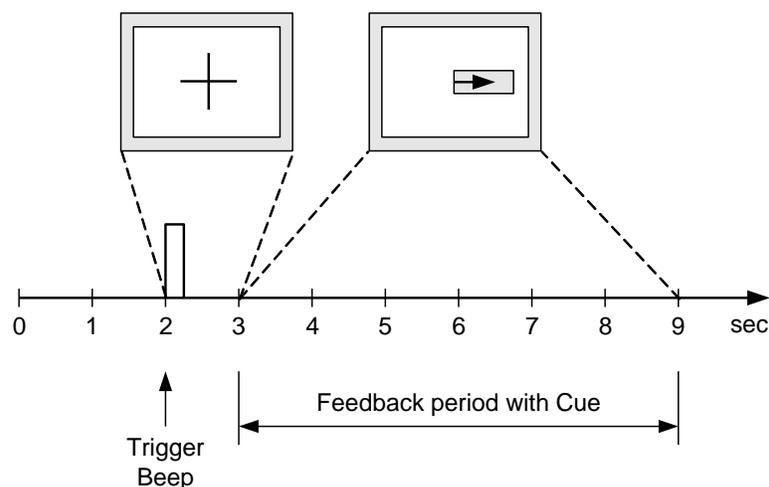
Una interfaz Cerebro-Computadora (BCI) es un sistema capaz de traducir señales electroencefalográficas (EEG) en comandos para un ordenador, haciendo posible el control de dispositivos externos. Una de las principales motivaciones para estos desarrollos es la asistencia de personas con discapacidades físicas como es el caso de pacientes con Esclerosis Lateral Amiotrófica o parálisis general, enfermedades que comprometen de manera severa el funcionamiento del sistema muscular afectando la habilidad de la persona para comunicarse a través de gestos, el habla y escritura. [2].

Una BCI incluye dentro de sus bloques generales: adquisición de las señales EEG, pre-procesamiento de las señales, extracción y selección de características, clasificación y retroalimentación. El sub-sistema de adquisición de señales debe garantizar una buena relación señal-ruido mediante la implementación de filtros y otras técnicas de procesamiento digital de señales para la eliminación de componentes no deseadas en la señal EEG.

La selección de características se realiza de manera dependiente al tipo de fenómeno fisiológico que se desea emplear en el funcionamiento de la BCI. Los fenómenos más empleados en la actualidad involucran potenciales relacionados a eventos (ERP's) como son, la actividad oscilatoria cerebral relacionada con actividad muscular o imaginación de la misma, (IM) potenciales corticales lentos (CSP) y potenciales P300 que se producen como respuesta a eventos no esperados o sorprendivos. [2], [1], [3] and [4]. En este trabajo, el fenómeno relacionado con la imaginación de actividad muscular que produce actividad oscilatoria cerebral es bandas de frecuencia específica: Alfa (8Hz-12Hz) y beta (18Hz-26Hz) es empleado. Las señales son modeladas empleando modelos auto-regresivos, obteniendo la respuesta en frecuencia de estos últimos como estimación de la densidad espectral de potencia. La clasificación de las señales se realiza por medio de 3 distintos métodos: Redes Neuronales Artificiales, Maquinas de vector de soporte y Análisis discriminante. Se presentan los resultados experimentales y se compara el desempeño obtenido con cada uno de los métodos.

2. DATA-SET EMPLEADO

En este trabajo, el data-set III de la competencia BCI III suministrado por el Departamento de Informática Médica de la Universidad de Graz fue empleado. El esquema del experimento es mostrado en la Figura 1. El data-set consiste en 140 repeticiones cada una de 9 segundos de duración. Al inicio de cada experimento una pantalla en blanco es mostrada al usuario. En $t=2s$ un tono es producido para alertar al usuario y una cruz aparece en pantalla. A partir de $t=3s$ en pantalla se muestra un flecha con dirección a la izquierda o a la derecha indicando al usuario que actividad mental ha de ser ejecutada, imaginación de movimiento de la mano derecha o de la mano izquierda. El usuario es instruido para realizar la actividad mental hasta que el experimento termina en $t=9s$. Las señales fueron adquiridas con electrodos superficiales sobre las posiciones posterior y anterior a los electrodos C3, C4 y Cz de acuerdo al estándar internacional 10/20. El sistema de adquisición muestrea las señales a una tasa de 128 muestras por segundo y un filtro con banda de paso entre 0.1Hz y 30Hz fue empleado. EL data-set provee señales para entrenamiento y señales de prueba sobre las cuales se ha de medir el desempeño del sistema.



3. EXTRACCIÓN Y SELECCIÓN DE CARACTERÍSTICAS

El fenómeno de desincronización / sincronización relacionada a eventos (ERD/ERS), presente en la ejecución o imaginación de actividad motora fue escogido como característica discriminante entre clases. Con el fin de detectar cambios en la potencia de la señal, es necesario estimar la densidad espectral de potencia de la misma. Con tal fin, métodos paramétricos y no-paramétricos fueron empleados y comparados en este trabajo.

Los métodos paramétricos involucran la determinación de modelos auto-regresivos. Dos de los métodos más empleados para generar estos modelos son el método de Burg y el método de Yuller-Walker. El método de Burg fue seleccionado debido a que, en contraste con Yulle-Walker, produce siempre un modelo estable minimizando el error en ambas direcciones.

Con el fin de obtener una representación en tiempo y frecuencia, modelos auto-regresivos de bloques de 1 segundo son calculados de acuerdo a:

$$y(n) = - \sum_{K=1}^p a(k)y(n - k) + u(n)$$

Donde $y(n)$ es el valor estimado de la señal, p corresponde al orden de modelo, $a(k)$ especifica los coeficientes del modelo y $u(n)$ es la función de ruido del sistema. La respuesta en frecuencia del modelo obtenido se emplea como estimación de la densidad espectral de potencia. Una ventana de 1 segundo es deslizada sobre la señal con pasos de una muestra proporcionando alta resolución temporal. El resultado es una representación que permite observar las componentes de la señal con mayor potencia para una componente de frecuencia específica y un instante de tiempo específico.

En [8] se describe la importancia de la selección del orden del modelo con el fin de obtener una buena estimación de la densidad espectral de potencia. El orden del modelo puede ser seleccionado basado en diversos métodos. En [8] el método Akaike's final prediction error (FPE) fue empleado, mostrando que modelos con orden tan bajo como cinco, proveen buenos resultados para segmentos cortos de señales EEG, sin embargo un orden de diez es sugerido debido a que se obtienen muy buenos resultados cuando el espectro obtenido es comparado con clásicos métodos que hacen uso de la Transformada de Fourier rápida (FFT). Para el caso del método no-paramétrico la transformada de Fourier de tiempo corto (STFT) fue seleccionada empleando también una ventana deslizante con longitud igual a 1 segundo.

La Figura 2 muestra el promedio de la potencia para cada clase (promedio sobre 70 experimentos en cada caso) en el rango de frecuencias de 7Hz a30Hz durante el desarrollo de cada experimento, usando el método de Burg y STFT, lo que permite evidenciar los cambios en la potencia en bandas de frecuencia específica dependiendo de la actividad realizada (Clase 1: imaginación de movimiento de la mano izquierda, clase 2: imaginación de movimiento de la mano derecha)

Con el fin de seleccionar las bandas de frecuencia que proveen mayor información sobre la actividad realizada, se analizó la separabilidad de los valores de potencia para cada una de las clases. Con ese propósito el índice de Fisher, una medida de la separabilidad entre dos clases fue empleada [9]:

$$Fisher_i = \frac{(m_{1i} - m_{2i})^2}{\sigma_{1i}^2 + \sigma_{2i}^2}$$

Donde m_{1i} y m_{2i} corresponden a la media de la característica i (componente de frecuencia) para la clase 1 y la clase 2 respectivamente. Las variables σ_{1i}^2 and σ_{2i}^2 corresponden al las varianzas de las característica i para la clase 1 y la clase 2 respectivamente. Esto produce una matriz, en la cual el valor del índice de Fisher puede ser

seleccionado para un valor específico de tiempo y frecuencia. Las figura 3 y 4 muestran valores del índice de Fisher para la señal EEG medida sobre los electrodos C3, C4 y Cz usando el método de Burg, y STFT.

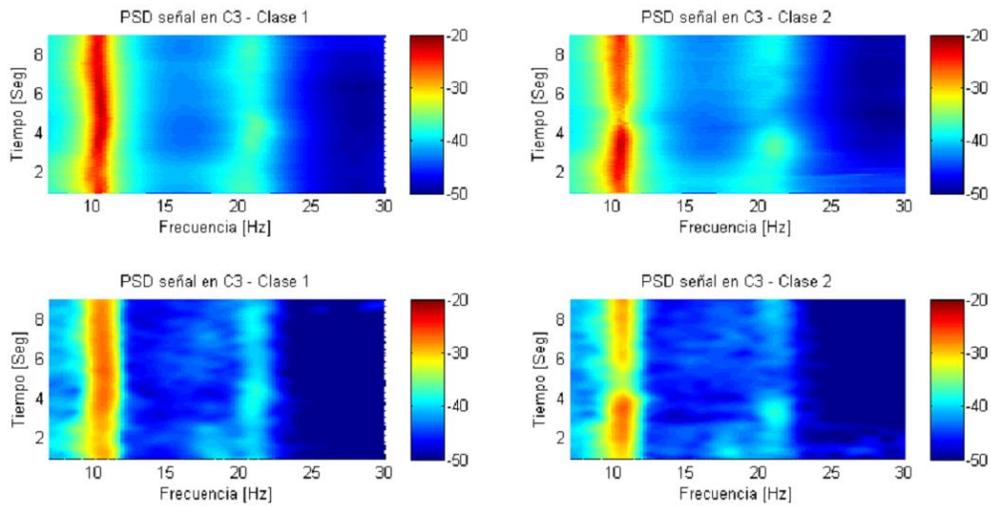


Figure 2. Potencia promedio para la clase 1 y la clase 2 empleando Burg y STFT.

Las Figuras 3 y 4 muestran que componentes al rededor de 10Hz (μ) y 20Hz (β) proveen mayor separabilidad indicado por los altos valores del índice de Fisher. Basado en esta información es posible construir vectores característica cuyos componentes son la potencia en las componentes de frecuencia con mayor índice de Fisher para los electrodos C3 y C4. Lo que da como resultado un vector de características $v \in R^5$, definido $v = \{\alpha_{powC3}, \alpha_{powC4}, \alpha_{powCz}, \beta_{powC3}, \beta_{powC4}\}$, (La banda β en el electrodo Cz no provee información significativa). Ese vector de características v es empleado como entrada a la etapa de clasificación.

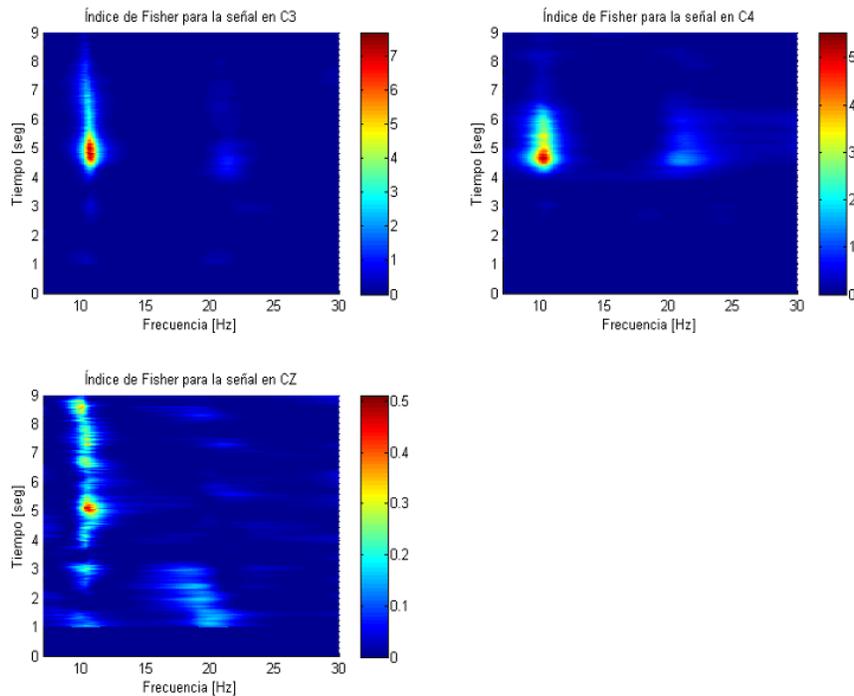


Figure 3. Índice de Fisher usando el método de BURG

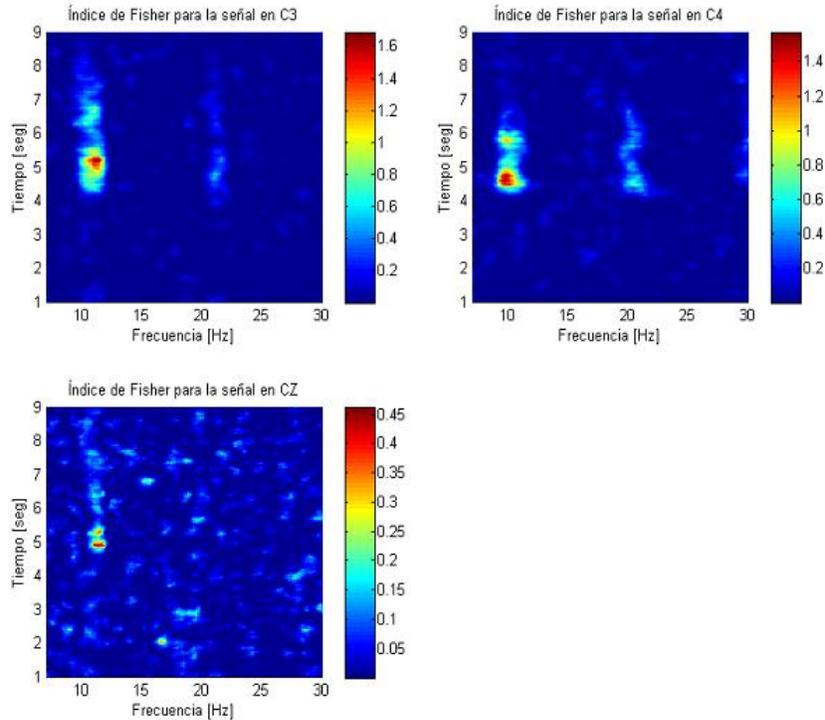


Figure 4. Índice de Fisher usando STFT

4. CLASSIFICATION

Esta sección provee información básica sobre los métodos de clasificación empleados, una descripción detallada de los mismos, está fuera del alcance de este trabajo por lo cual el lector es remitido a los fuentes empleando las referencias citadas.

4.1 MÁQUINA DE VECTOR DE SOPORTE (SVM)

Distintos métodos de clasificación han sido empleados en BCI sin embargo dichos métodos no son eficientes cuando la distribución de los datos no es conocida a priori. Las maquinas de vector de soporte funcionan bien bajo estas condiciones [10]. En el caso que se describe en este trabajo hay dos clases, representando actividad motora imaginaria de la mano derecha o de la mano izquierda. Esto significa que el espacio de muestras es de la forma $s = \{(x_1, y_1) \dots (x_m, y_m)\}$ donde \mathbf{x} corresponde al vector multidimensional v y $\mathbf{y} \in \{+1, -1\}$ corresponde a las etiquetas para cada una de las clases. Si se asume que existe separación lineal entre las clases, la función discriminante es dada por:

$$\mathbf{y} = g(\mathbf{x}) = \text{sgn}(\langle \mathbf{w}, \mathbf{x} \rangle + b)$$

Donde **sgn** hace referencia a la función “signo”, el operador $\langle \cdot, \cdot \rangle$ implica producto interno y \mathbf{w} determina la dirección del plano que separa las dos clases [10]. Los datos son separados de manera optima si no hay errores en la clasificación y el plano que separa las clases esta posicionado a la máxima distancia posible de los puntos más cercanos, conocidos como vectores de soporte.

En este trabajo el software SVM^{light} [11] y la interfaz con Matlab (R) a SVM^{light} [12] fueron empleados, haciendo uso de un kernel lineal.

4.2 ANÁLISIS DISCRIMINANTE LINEAL (LDA).

Dado un conjunto de datos dentro del cual se presentan subgrupos, LDA busca una transformación de los datos de modo en el nuevo espacio de la transformación maximiza la función:

$$j(w) = \frac{w^T S_B w}{w^T S_w w}$$

Lo que implica que maximiza la tasa entre la matriz de dispersión $S_B = \sum_c N_c (\mu_c - \bar{X})(\mu_c - \bar{X})^T$ entre clases y minimiza la matriz de dispersión al interior de cada clase $S_w = \sum_c \sum_{i \in c} (X_i - \mu_c)(X_i - \mu_c)^T$. Donde, c representa cada clase, N_c en numero muestras en la clases c , μ_c la media de las muestras x_i pertenecientes a la clase c and \bar{X} es la media de la muestra para todos los datos. Una vez que la muestra ha sido proyectada al nuevo espacio $j(w)$ es maximizada y es posible realizar la clasificación.

4.3 REDES NEURONALES ARTIFICIALES.

Una red neuronal artificial (ANN) es un sistema (algoritmo o hardware) con la capacidad de procesar datos [13]. La estructura de una ANN es desarrollada de manera similar a una red neuronal biológica [14] (BNN's) de modo que en las ANN el proceso de aprendizaje es basado en ejemplo. La estructura más empleada para ANN (feed-forward ANN) hace uso de una capa de entrada, una capa oculta y una capa de salida [13]. Los datos de entrada son enviados a la capa oculta donde son ponderados a la entrada de cada neurona artificial y una función de activación (usualmente una función sigmoide para aplicaciones de clasificación) es aplicada a la suma de las entradas ponderadas. Finalmente el resultado es enviado a la capa de salida. Los pesos de cada neurona, empleados para ponderar los datos de entrada son calculados de modo que el error cuadrático medio entre la entrada y la salida es minimizado. Este trabajo hace uso de una ANN (Feed-forward) para clasificación.

5. RESULTADOS

5.1 CLASIFICACIÓN USANDO SVM

La Figura 5 muestra los valores temporales para la precisión en la clasificación, la información mutua y la relación señal a ruido de acuerdo al criterio de evaluación empleado en la competencia BCI III [15], [16].

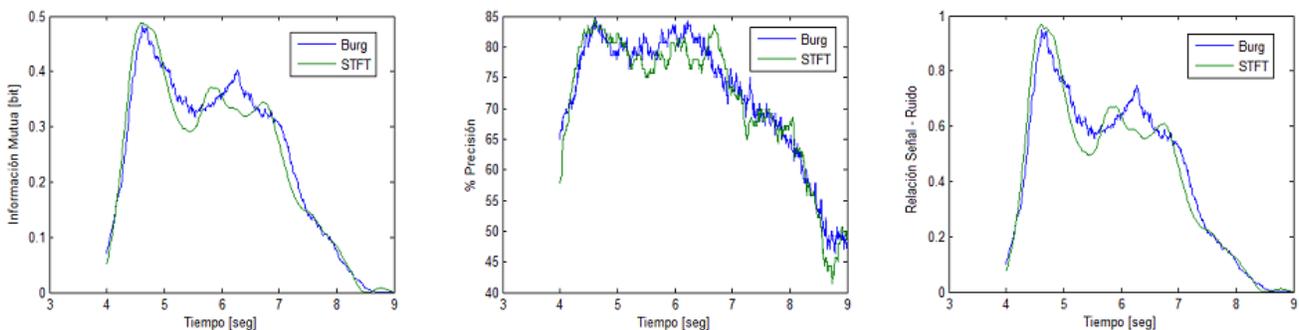


Figure 5. Resultados de la Clasificación empleando SVM.

La Tabla III muestra los valores máximos para la información mutua (IM), precisión ACC y la relación señal a ruido SNR encontrados empleando SVM para los dos métodos usados para extraer las características, STFT and BURG.

TABLE III. RESULTADOS DE LA CLASIFICACIÓN EMPLEANDO SVM

	Max MI	Max ACC	Max SNR
BURG	0.489	85,00%	0.97
STFT	0.481	85,00%	0.95

5.2 CLASIFICACIÓN USANDO LDA

La Figura 7 muestra resultados de clasificación empleando LDA, un resumen de los datos es mostrado en la tabla IV.

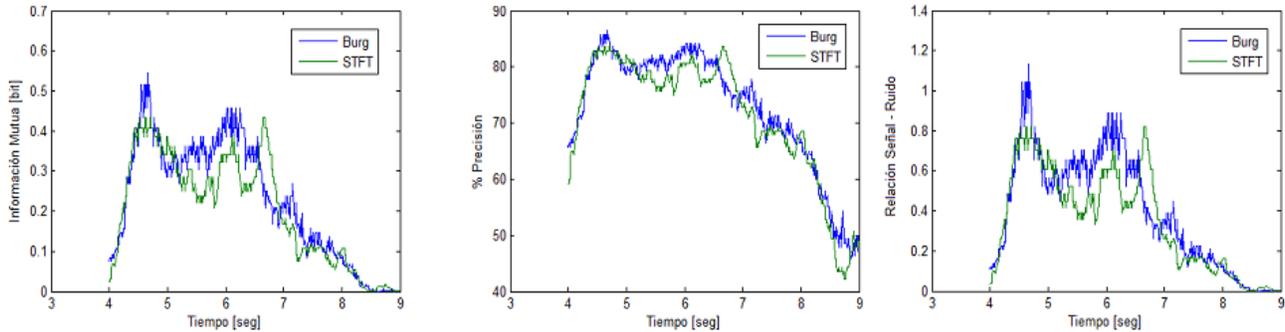


Figure 7. Resultados de la evaluación empleando LDA.

TABLE IV. RESULTADOS DE LA CLASIFICACIÓN EMPLEANDO LDA

	Max MI	Max ACC	Max SNR
BURG	0.546	86.43%	1.13
STFT	0.432	83.57%	0.82

5.3 CLASIFICACIÓN USANDO ANN

La estructura usada en la implementación de la ANN fue feed-forward. La selección del número de neuronas en la capa oculta fue realizada de manera empírica variando el número de neuronas de 1 a 25. Las tablas V and VI muestran los máximos valores encontrados para la MI, ACC and SNR para ANN's con diferentes números de neuronas en la capa oculta.

TABLE V. RESULTADOS DE CLASIFICACIÓN EMPLEANDO ANN'S CON DIFERENTES NÚMEROS DE NEURONAS EN LA CAPA OCULTA – MÉTODO BURG

	Número de Neuronas					
	1	5	10	15	20	25
Max MI	0.546	0.545	0.511	0.550	0.569	0.511
Max ACC	86.4%	85,7%	85,7%	85,0%	84,3%	85,0%
Max SNR	1.13	1.13	1.03	1.15	1.20	1.03

TABLE VI. RESULTADOS DE CLASIFICACIÓN EMPLEANDO ANN'S CON DIFERENTES NÚMEROS DE NEURONAS EN LA CAPA OCULTA – MÉTODO STFT

	Número de Neuronas					
	1	5	10	15	20	25
Max MI	0.514	0.506	0.516	0.524	0.521	0.546
Max ACC	84.3%	85.7%	86.4%	86.4%	82.9%	86,4%
Max SNR	1.04	1.02	1.04	1.07	1.06	1.13

Como se observa, el caso de la STFT un número mayor de neuronas en la capa oculta es necesario para obtener resultados iguales a los obtenidos con el método de Burg. En la figura 8 el cambio de los parámetros MI, ACC y SNR con el tiempo, para el caso de ANN's es mostrado

Una comparación de los tres métodos es mostrada en la figura 9. El valor pico más alto es obtenido mediante ANN's, sin embargo las diferencias observadas en el desempeño no son apreciables. Con el fin de poder realizar comparación entre los métodos empleados el coeficiente kappa [15] es introducido. El coeficiente kappa es un indicador del grado de acuerdo de dos observadores sobre un evento. Este indicador resulta más robusto que el medición de la precisión (ACC) debido a que tiene en cuenta la probabilidad de acierto debido al azar. La tabla VII muestra los valores de kappa y el error obtenido para con cada uno de los métodos.

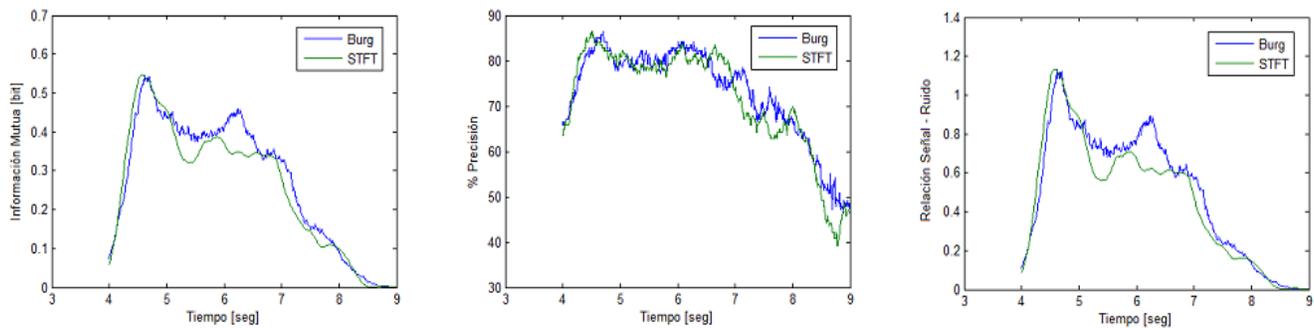


Figure 8. Resultados de la evaluación empleando ANN.

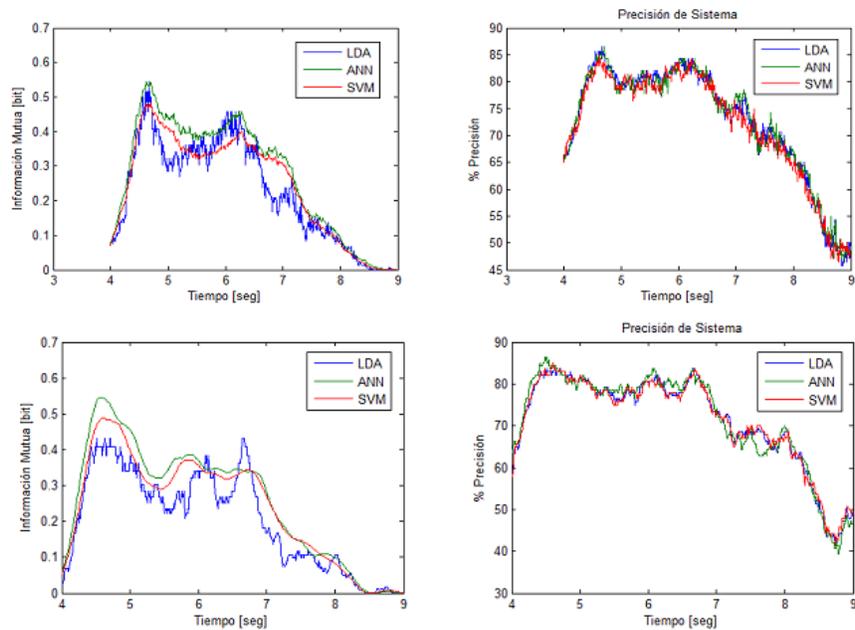


Figure 9. Comparación de los resultados obtenidos con los tres métodos

TABLE IX
RESULTADOS EMPLEANDO KAPPA CON CRITERIO DE EVALUACIÓN – BURG

	Kappa	Precisión
LDA	0.7290	86.43%
ANN	0.7280	86.40%
SVM	0.7000	85.00%

TABLE X
RESULTADOS EMPLEANDO KAPPA CON CRITERIO DE EVALUACIÓN – STFT

	Kappa	Precisión
LDA	0.6714	83.57%
ANN	0.7280	86.40%
SVM	0.7000	85.00%

6. CONCLUSIÓN.

En este trabajo se describe el uso de métodos paramétricos y no paramétricos para determinación de la densidad espectral de potencia con aplicación al diseño de interfaces cerebro computadora. Las características de la señal son obtenidas analizando las bandas de frecuencia que ofrecen mayor separabilidad de acuerdo al índice de Fisher. La clasificación de la señal es realizada mediante tres distintos tipos de clasificadores SVM, LDA y ANN. Los resultados muestran que la estimación de la densidad espectral de potencia empleando métodos paramétricos provee mejores resultados. Con respecto a los clasificadores, las ANN muestran un desempeño ligeramente mejor comparable a LDA. Dado que los métodos LDA y SVM presentados en este trabajo no permiten una frontera de separación no-lineal, los resultados resultan consistentes dado que las ANN pueden proveer una superficie de decisión no lineal. Sin embargo, LDA requiere menos poder de computo, la implementación resulta mucho más sencilla y permite un mayor análisis mediante la examinación de la distribución de las clases en el nuevo espacio (obtenido debido a la transformación empleada), lo que resulta imposible en el caso de las ANN. Por tal razón a la luz de los resultados el método sugerido resulta de la combinación de métodos paramétricos para la determinación de la densidad espectral de potencia y LDA. Por otro lado modelos no lineales basados en funciones de base

radial para SVM podrían resultar en sobre entrenamiento disminuyendo el performance del sistema, dada la naturaleza de las señales empleadas.

REFERENCIAS.

- [1] G. Pfurtscheller, C. Neuper, and N. Birbaume, "Human Brain - Computer Interface," in *Motor Cortex in Voluntary Movements*, A. Riehle and E. Vaadia, Eds. CRC Press LLC, 2005, ch. 14.
- [2] U. Hoffmann, J.-M. Vesin, and T. Ebrahimi, "Recent Advances in Brain-Computer Interfaces," *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on*, p. 17, Oct. 2007.
- [3] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, "Brain-Computer Interface Research at the Wadsworth Center," *IEEE TRANSACTIONS ON REHABILITATION ENGINEERING*, vol. 8, no. 2, pp. 222-226, Jun. 2000.
- [4] A. Kübler and L.-R. Müller, "An introduction to Brain-Computer Interfacing," in *Toward Brain-Computer Interface*, G. Dornhege, et al., Eds. London, England: MIT Press, 2007, ch. 1, pp. 1-25.
- [5] IDA. (2009, Sep.) BCI Competition II. [Online]. http://www.bbc.de/competition/ii/Graz_description.doc
- [6] W. A. MacKay, "Wheels of Motion: Oscillatory Potentials in the Motor Cortex," in *Motor Cortex in voluntary movements*, A. Riehle and E. Vaadia, Eds. USA: CRC Press LLC, 2005..
- [7] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, , . Upper Saddle River, N.J.: Prentice Hall, 1997.
- [8] B. H. Jansen, J. R. Bourne, and J. W. Ward, "Autoregressive Estimation of Short Segment Spectra for Computerized EEG Analysis," *Biomedical Engineering, IEEE Transactions on*, pp. 630-638, Sep. 1981.
- [9] X. Pei and C. Zheng, "Classification of Left and Right Hand Motor Imagery Tasks Based on EEG Frequency Component Selection," *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*, pp. 1888-1891, May 2008.
- [10] S. Sanei, *EEG Signal Processing*. West Sussex, England: Wiley & Sons Ltd, 2007.
- [11] T. Joachims, "Advances in Kernel Methods - Support Vector Learning," in *Making large-Scale SVM Learning Practical*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, USA: MIT Press, 1999, ch. 11, pp. 41-55.
- [12] T. Briggs. (2009, Mar.) Tom Briggs Assistant Professor of Computer Science, Shippensburg University. [Online]. <http://webspaceship.edu/thbrig/>
- [13] L. Francis, "The Basics of Neural Networks Demystified," *Contingencies*, pp. 56-61, Nov. 2001.
- [14] P. Isasi Viñuela, *Redes Neuronales Artificiales: un enfoque práctico*. Madrid, España: Pearson Educación, 2004.
- [15] S. A. K. J, H. E, and M. G, "Evaluation criteria in BCI research," in *Toward Brain - Computer Interfacing*, G. Dornhege, et al., Eds. Massachusetts, USA: MIT Press, 2007, ch. 19, pp. 327-342.
- [16] A. Schlogl, C. Keinrath, R. Scherer, and P. Furtscheller, "Information transfer of an EEG-based brain computer interface," *Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*, pp. 641-644, Mar. 2003.
- [17] Q. Zhao, L. Zhang, and C. Andrzej, "EEG-based asynchronous BCI control of a car in 3D virtual reality environments," *Chinese Science Bulletin*, vol. 54, no. 1, pp. 78-87, Jan. 2009.
- [18] R. Leeb, R. Scherer, C. Keinrath, C. Guger, and G. Pfurtscheller, "Exploring Virtual Environments with an EEG-based BCI through Motor Imagery," *Biomedical Engineering*, vol. 50, no. 4, pp. 86-91, Apr. 2005.
- [19] C. Vidaurre, A. Schlögl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "Adaptive On-line Classification for EEG-based Brain Computer Interfaces with AAR parameters and band power estimates," *Biomedical Engineering*, vol. 50, no. 11, p. 350-354, November.
- [20] A. Adlakha, "Single trial EEG Classification," Swiss Federal Institute of Technology, 2002.
- [21] G. N. Gracia, *Direct Brain Computer Communication Through Scalp Recorded EEG Signals*. Lausanne, Switzerland, 2004.
- [22] D. J. McFarland, A. T. Lefkowitz, and J. R. Wolpaw, "technology, Design and operation of an EEG-based brain-computer interface with digital signal processing," *Behavior Research Methods, Instruments, & Computers*, vol. 3, no. 29, p. 337-345, 1997.

Autorización y Renuncia

Los autores autorizan a LACCEI para publicar el escrito en las memorias de la conferencia. LACCEI o los editores no son responsables ni por el contenido ni por las implicaciones de lo que esta expresado en el escrito.