



UNIVERSIDAD AUTÓNOMA METROPOLITANA - IZTAPALAPA  
DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

DETECCIÓN OPTIMIZADA DE LA RESPUESTA INFRECUENTE  
EN INTERFACES CEREBRO-COMPUTADORA

Tesis que presenta  
Alma Cecilia Lindig León  
Para obtener el grado de  
Maestra en Ciencias en Ingeniería Biomédica

Asesor: M.I. ÓSCAR YÁÑEZ SUÁREZ

Jurado Calificador:

Presidente: DR. FERNANDO ARÁMBULA COSÍO UNAM-IIMAS

Secretario: M.I. ÓSCAR YÁÑEZ SUÁREZ UAM-I

Vocal: DRA. VERÓNICA MEDINA BAÑUELOS UAM-I

México, D.F. Julio 2013

A Michi y a Pía,  
por toda la magia.

## Agradecimientos

A la Universidad Autónoma Metropolitana donde recibí, gracias a todos sus profesores, esta formación. En particular a Johny, mi maestro desde la licenciatura.

Al Consejo Nacional de Ciencia y Tecnología, por el apoyo que no sólo me brindó a mí, sino a tantos compañeros; el compromiso por tener un País cada vez mejor lo compartimos.

A mi asesor, el Profesor Óscar Yáñez. Por el entusiasmo que transmite; por la incapacidad que tiene de decir “no” a los alumnos, aunque la labor se vuelva humanamente imposible. Persona ejemplar que siempre encuentra las herramientas y soluciones para cimentar los castillos que estaban sólo en el aire.

A la Dra. Verónica Medina. Nunca tuve la suerte de cursar alguna materia que impartiera, sin embargo tuve la fortuna de que formara parte del jurado de esta tesis; lo que me permitió recibir sus valiosos y atinados consejos, muy importantes en este proceso de formación.

Al Dr. Fernando Arámbula, quien ha estado muy cerca de este proyecto y nos ha acompañado con valiosas aportaciones a lo largo de su desarrollo.

A Michi, por todo. No hay un sólo resquicio de esta vida en el que no aparezcas; siempre con grandes consejos, con el apoyo y el respaldo que han hecho todo esto posible. Sin embargo, me quedo con lo simple, con la cotidianidad de una gran infancia, con cada uno de los hot cakes dominicales y con tus hermosos jardines.

A Almita, también por todo. Por la confianza, por tu apoyo incondicional. Por todo el esfuerzo que nos ha traído hasta aquí; juntas llegamos mucho más lejos.

A Fave, jamás podré plasmar lo que significas para mí. No sólo mi gran amiga, sino un ejemplo de vida. La soledad no es tal desde que nos encontramos, así, en nuestra memoria vive.

A Miguelito, por tu gran personalidad; por cómo me haces reír. No he conocido nunca una persona con un sentido del humor como el tuyo, sólo un atisbo de tu enorme inteligencia. Te deseo todo el éxito desde lo más profundo de mi corazón.

A Mari, por tu gran carácter; por ser un ejemplo de voluntad y entrega. Por llegar a mi vida y darle fortuna.

## Resumen

El paradigma *Oddball* es uno de los más utilizados en el desarrollo de interfaces cerebro-computadora que buscan mejorar la calidad de vida de personas con altos grados de discapacidad. Esto se debe a que a diferencia de otras interfaces, éstas ofrecen a los usuarios un vasto espectro de posibilidades para interactuar con su entorno a partir de potenciales que, incluso aquellos con mayores limitaciones, son capaces de generar para conseguir controlarlas. Sin embargo, el modo tan lento con el que operan este tipo de aplicaciones representa un inconveniente para aquellos que las utilizan en su vida cotidiana; situación que es difícil de superar porque intervienen dos cuestiones intrínsecas al sistema: Por un lado, se trata de aplicaciones que requieren de una rutina que provoque los estímulos relacionados con cada una de las posibilidades que le brinda al usuario, quien debe permanecer atento y esperar, por varios segundos, a que ocurra aquél que corresponde con su elección y, por otro lado, las respuestas generadas en el electroencefalograma (EEG) presentan una razón señal a ruido muy baja, con lo que resulta prácticamente imposible detectar el potencial correspondiente. Para corregir esta situación se emplean técnicas de promediación coherente, de este modo se consigue resaltar los patrones que tienen en común las señales de EEG y, al mismo tiempo, eliminar los componentes relacionados con otro tipo de actividad o, con ciertos artefactos. Entre más respuestas se promedien, mejor es el resultado que se obtiene. Sin embargo, esto implica un fuerte incremento en el tiempo que le lleva al sistema tomar una decisión porque, para cada respuesta considerada en el promedio, es necesario presentar una rutina de estimulación. A partir de este hecho, se puede observar que existe un compromiso entre el tiempo que le lleva al sistema tomar una decisión y la precisión correspondiente. Mientras más secuencias de estimulación se emplean, mayor es el tiempo que le toma al usuario controlar al sistema; no obstante, de este modo se evita cometer errores. Y, por el contrario, entre menos secuencias se generan, menor es el tiempo necesario para controlar la interfaz, pero esto, por supuesto, en detrimento de la precisión con la que se presentan los resultados. Por esta razón, el número de secuencias de estimulación que se presentan para determinar cada uno de los elementos, es un

parámetro que debe elegirse con cuidado. En la mayoría de los casos, se opta por presentar varias secuencias y, de este modo, asegurar un buen desempeño. El inconveniente que esto conlleva es que, con frecuencia, estas secuencias son presentadas en vano, es decir, con la información registrada en el EEG desde la primera o segunda serie de estimulación, es posible hacer la inferencia correspondiente, no obstante, el sistema sigue estimulando al usuario hasta presentar todas las secuencias que se establecieron durante su configuración. Este hecho refleja lo inadecuado de considerar que, el número de secuencias de estimulación para determinar cada elemento, corresponda con un parámetro estático que destine siempre la misma cantidad de estímulos.

En este trabajo se presenta una propuesta para que dicho parámetro se establezca de manera dinámica, de modo que el sistema sea capaz de determinar automáticamente el número de secuencias de estimulación necesario para clasificar cada respuesta. Para ello, se presenta un enfoque bayesiano que permite determinar la probabilidad posterior asociada con la clasificación de cada elemento, la cual puede entenderse como la probabilidad de que la clasificación se haya realizado correctamente y, de este modo, determinar si es necesario seguir estimulando al usuario para mejorar la calidad de las señales, o asegurar con cierto grado de probabilidad, que se ha tomado ya la decisión correcta y, por lo tanto, es posible detener la estimulación.

Este nuevo esquema evita que la interfaz invierta tiempo en generar secuencias que provocan respuestas redundantes sin descuidar la precisión de su desempeño. El hecho de concebir el número de secuencias de estimulación como un parámetro dinámico, permite que el sistema se ajuste a las características de los datos y genere tantos estímulos como sean necesarios, lo que se refleja en un decremento del tiempo promedio en el control de este tipo de aplicaciones.

Vale la pena recalcar que, bajo este enfoque, el operador del sistema ya no elegirá un número determinado de secuencias de estimulación para tomar decisiones. En su lugar, debe establecer una serie de umbrales que le permiten al sistema tomar decisiones.

Los resultados obtenidos después de utilizar la metodología propuesta sobre los registros de 14 usuarios son satisfactorios, el tiempo promedio de deletreo es de  $6.1 \pm 0.63$  letras/min que, en comparación con el reportado por un sistema convencional configurado para presentar 5 secuencias de estimulación con una velocidad de deletreo constante de 3.93 letras/min, representa una mejora de más del 35%. En ambos casos, la precisión de los resultados se mantuvo por encima del 90%.

# Índice

1. Introducción .....	1
1.1 Funcionamiento de una BCI .....	2
1.2. Actividad Endógena .....	3
1.2.1. Ritmos Mu .....	3
1.2.2. Potencial Preparatorio .....	4
1.2.3. Procesos Mentales Cognitivos .....	5
1.2.4. Aprendizaje del Usuario .....	5
1.3. Actividad Exógena .....	5
1.3.1. Potenciales Visuales Provocados .....	6
1.3.2. Potencial P300 .....	6
1.4. Paradigma de Evento Raro ( <i>Oddball</i> ) .....	7
1.5. Matriz de Farwell y Donchin .....	8
1.6. Motivación .....	9
1.7. Objetivos .....	11
1.7.1. Objetivo General.....	11
1.7.2. Objetivos Específicos .....	11
2. Marco Teórico .....	12
2.1. Noción de clasificador .....	13
2.1.2. Clasificadores de Discriminante Lineal .....	13
2.1.3. Entrenamiento de los Clasificadores .....	14

2.2.	Teorema de Bayes .....	16
2.3.	Sigmoide logística .....	18
2.4.	Algoritmo de Platt .....	21
2.5.	Validación Cruzada .....	23
2.5.1.	Validación Cruzada de K Vías .....	23
2.5.2.	Validación Cruzada Aleatoria .....	24
2.5.3.	Validación Cruzada Dejando Uno Fuera .....	24
3.	Metodología .....	25
3.1	. Criterios de Inclusión y Adquisición de la Información .....	26
3.1.2.	Filtrado de la Señal de EEG .....	29
3.2.	Partición de los Datos .....	29
3.3.	Entrenamiento del Sistema .....	31
3.3.1.	Etapa 1: Construcción de Un Clasificador Lineal Sobre los Rasgos Temporales de EEG.....	33
3.3.2.	Etapa 2: Construcción de Vectores de Distancia Acumulada.....	36
3.3.3.	Etapa 3: Construcción de N Clasificadores Lineales Sobre Las Características de los Vectores de Distancia Acumulada .....	39
3.3.4.	Etapa 4: Cálculo de la n-ésima Distancia Dirigida .....	40
3.3.5.	Etapa 5: Construcción de N Modelos de Funciones Sigmoide- Logística y Determinación del Sesgo .....	42
3.3.6.	Etapa 6: Cálculo de la Probabilidad Posterior y Criterios de Decisión.....	45
4.	Resultados .....	49
4.1.	Operación del Sistema .....	50
4.2.	Validación del Sistema .....	59
4.3.	Resultados de Todos los Usuarios .....	60
4.4.	Discusión .....	61

5. Conclusiones.....	62
Referencias .....	64
Anexo 1. Resultados de cada Usuario.....	66
Anexo 2. Publicación del trabajo en la Revista Mexicana de Ingeniería Biomédica .....	88
Anexo 3. Presentación del trabajo en el Congreso Nacional de Ingeniería Biomédica. ....	106



# Capítulo 1

## 1. Introducción

Las interfaces cerebro-computadora, BCI por sus siglas en inglés (Brain-computer Interface), son sistemas capaces de extraer e interpretar la actividad eléctrica cerebral para controlar el uso de ciertas aplicaciones o diversos dispositivos.

Se han caracterizado algunos patrones de las señales de EEG, asociados con determinados procesos mentales, que pueden provocarse e identificarse fácilmente. Las BCI se basan en la detección de dichos patrones, los cuales representan comandos que el sistema interpreta para ejecutar tareas o expresar mensajes sin la intervención de los músculos ni los nervios periféricos del usuario, es decir, sin la necesidad de que mueva alguna parte del cuerpo. Por esta razón, las BCI constituyen un esquema muy importante para el desarrollo de sistemas que buscan mejorar la calidad de vida de personas que sufren alguna lesión o, alteración neuromuscular, responsable de causar altos grados de discapacidad, como la pérdida total del movimiento voluntario y, por consiguiente, de la posibilidad de interacción con su entorno [1, 2].

En la actualidad, existen diversos paradigmas sobre los que se desarrollan muchos tipos de BCI; cada uno de ellos, se ajusta a las características de las señales de EEG que, dentro del contexto de operación del sistema, se originan a causa de la voluntad que imprime el usuario al realizar cierta tarea. Los procesos mentales que producen estas características en las señales de EEG se clasifican en dos grupos: aquellos que generan actividad endógena, y aquellos que generan actividad exógena. Aun cuando las interfaces que se pueden desarrollar con cada tipo de actividad cerebral resultan de diversa índole, es posible generalizar su funcionamiento y distinguir los diferentes módulos que las componen, los cuales, salvo por el módulo de estimulación que no requieren las interfaces basadas en la detección de actividad endógena, constituyen la base fundamental de cualquier tipo de BCI [3, 4].

En la figura 1.1 se presenta un esquema del funcionamiento de una BCI a partir del cual es posible identificar los diferentes módulos que componen el sistema [5,6].



**Figura 1.1.** Esquema del funcionamiento de una interfaz cerebro-computadora.

Imagen tomada y adaptada de <http://www.softinno.co.cc/mizan/BCI.html>

- El **módulo de adquisición** consiste, en la mayoría de los casos, en un conjunto de electrodos externos que se posicionan en la cabeza del usuario de acuerdo con el sistema 10-20 para transmitir las señales de EEG a un sistema que digitaliza y amplifica la información. La señal que recibe cada uno de los electrodos se transmite a una de las entradas de un amplificador diferencial, esto significa que se requiere de un amplificador para cada par formado por el electrodo activo y el de referencia. Normalmente, la ganancia que se obtiene incrementa de 1000 a 100000 veces el voltaje entre ambos electrodos.

La cantidad y posición de los electrodos es una cuestión que debe determinarse con cuidado. Existen muchos trabajos con varias propuestas para optimizar los efectos de esta distribución, básicamente se trata de elegir las regiones más cercanas a las fuentes de las señales de interés. Existen módulos de adquisición que utilizan electrodos internos implantados directamente sobre la corteza cerebral del usuario. De este modo, además de que la calidad de las señales es mucho mejor, se evita tener que

realizar la tardada y tediosa labor de colocarlos antes de empezar cada sesión no obstante, debido a su naturaleza invasiva, resulta una práctica poco común.

- El **módulo de procesamiento** recibe las señales de salida del módulo de adquisición y, a partir de una serie de algoritmos, detecta los patrones asociados con la intención del usuario y toma las decisiones que desencadenan las acciones correspondientes. Para ello, emplea técnicas de procesamiento que acondicionan la información, posteriormente, extrae las características que le permiten realizar la clasificación de las señales, a partir de la cual determina si se generó o no, en el registro de EEG, el comando correspondiente.
- Cuando el sistema detecta el comando en las señales de EEG, desencadena la acción correspondiente a través de algún puerto de salida como, por ejemplo, una pantalla, una silla de ruedas o algún otro dispositivo, el cual consiste en el **módulo de retroalimentación** de la interfaz y es, finalmente, el que le brinda al usuario la posibilidad de interactuar con su entorno.
- Las BCI que emplean la detección de actividad exógena, además de los módulos indicados, requieren de un **módulo de estimulación**, a través del cual se consigue generar las respuestas en el EEG que le permiten al usuario controlar el sistema.

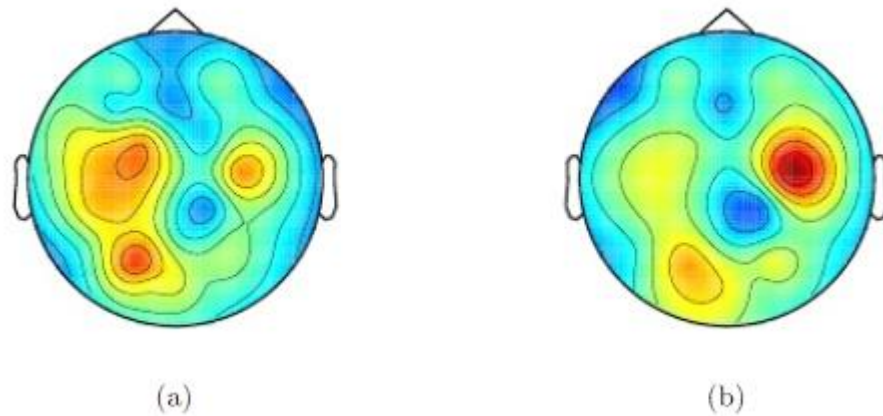
## 1.2. Actividad Endógena

La actividad endógena corresponde con respuestas asociadas con ciertos procesos mentales que el usuario puede provocar sin la necesidad de agentes externos; por esta razón, las BCI que se basan en su detección, no requieren de un módulo de estimulación. Un ejemplo de este tipo de señales corresponde con ciertas oscilaciones, denominadas ritmos, que son provocadas por la sincronía neuronal de las regiones cerebrales involucradas en procesos relacionados con el movimiento de las extremidades [7].

### 1.2.1. Ritmos Mu

En la figura 1.2 se observa un esquema topográfico que muestra la potencia de la actividad registrada en la banda alfa (8-13 Hz). Dentro de este rango, es posible analizar los ritmos mu, los cuales se originan cuando el usuario mueve o imagina el movimiento de alguna de sus extremidades. A partir del esquema, pueden observarse las diferencias en la distribución de la actividad rítmica relacionada con la extremidad que el usuario decide imaginar que mueve. Las BCI que utilizan la detección de estos ritmos suelen ser aplicaciones que le permiten al usuario tomar, de manera espontánea, una decisión

binaria que puede emplearse para llevar a cabo un movimiento continuo como, por ejemplo, subir o bajar el cursor de una pantalla y, de este modo, expresar “sí” o “no”; cerrar o abrir la pinza de un brazo mecánico; mover hacia delante o hacia atrás una silla de ruedas; aumentar o disminuir el volumen de algún reproductor, así como cambiar los canales, etc. [8, 9].



**Figura 1.2.** Distribución topográfica de potencia en la banda alfa mientras el usuario imagina el movimiento de la mano izquierda (a) y, el de la mano derecha (b). Las regiones en color rojo exhiben la presencia de fuerte actividad rítmica. Imagen tomada de [8].

### 1.2.2. Potencial Preparatorio

Otro tipo de actividad endógena asociada con el movimiento ocurre en la región cortical del hemisferio frontal, aproximadamente, 1.5 s antes de mover, o sólo imaginar que se mueve, alguna extremidad. Este proceso genera un potencial que decrece lentamente, denominado potencial *bereitschaft* o preparatorio. Al igual que los ritmos mu, estos potenciales presentan una distribución espacial característica de la extremidad utilizada en el proceso, la cual exhibe mayor amplitud en las zonas contralaterales al movimiento imaginado. El contraste que se observa cuando se utilizan las extremidades derecha o izquierda representa otro indicador viable de decisión binaria para el desarrollo de BCI [10].

### 1.2.3. Procesos Mentales Cognitivos

Además de la actividad neuronal relacionada con las tareas de movimiento, existen otros patrones asociados con procesos endógenos cognitivos, que pueden funcionar como indicadores de decisión. Algunas tareas que han demostrado generar este tipo de patrones consisten en el cálculo de multiplicaciones complejas; rotación mental de objetos geométricos y ejercicios de composición literaria entre otros. Se han desarrollado aplicaciones de BCI sensibles a este tipo de procesos, los cuales se asocian con una acción que el sistema es capaz de ejecutar cuando detecta la actividad cerebral correspondiente; de esta manera, se le ofrece al usuario la posibilidad de expresar una decisión entre muchas [11].

### 1.2.4. Aprendizaje del Usuario

Las BCI basadas en la detección de actividad endógena, demandan el entrenamiento del usuario para que sea capaz de controlar de manera satisfactoria los estados mentales a los que es sensible el sistema. Una técnica común de aprendizaje se basa en el condicionamiento operante, el cual consiste en promover el desarrollo de nuevas conductas que tienen efectos positivos. De modo que, a prueba y error, el usuario novato aprende a reconocer las acciones que le permiten controlar mejor el sistema. La duración del entrenamiento, depende tanto de su habilidad como de los algoritmos empleados para analizar la información y, en la mayoría de los casos, requiere de varias semanas [12].

## 1.3. Actividad Exógena

Las BCI que emplean la detección de actividad exógena, pueden empezar a utilizarse de manera eficaz sin la necesidad de un entrenamiento exhaustivo. Esto se debe a que dicha actividad, denominada potencial provocado, es la respuesta inherente a estímulos externos específicos y se manifiesta de manera autónoma. Para provocar este tipo de respuestas, se requiere de un módulo de estimulación que genere los estímulos asociados con los elementos de decisión a los que debe atender el usuario. Debido a que es necesario esperar a que la rutina genere los estímulos que le son relevantes, estos paradigmas presentan el inconveniente de no permitir el control espontáneo del sistema.

Usualmente, para conseguir identificar los potenciales provocados, es necesario promediar varias respuestas. Sin embargo, en comparación con la actividad endógena, resultan más fáciles de detectar y, en casi todos los casos debido a su naturaleza

inherente, les es posible a los usuarios generarlos incluso después de sufrir otras lesiones neurológicas.

Se pueden desarrollar aplicaciones basadas en la detección de potenciales provocados con un amplio repertorio de posibilidades. La forma de control que ofrecen estos paradigmas es discreta y, en la mayoría de los casos, se trata de sistemas que le permiten al usuario elegir, dentro de un conjunto de símbolos, aquél que necesita para expresar un mensaje, o para realizar una acción determinada.

### **1.3.1. Potenciales Visuales Provocados**

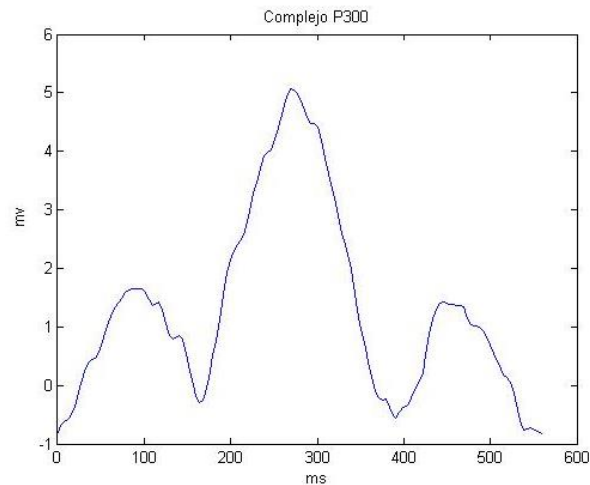
Dentro de la categoría de los potenciales visuales provocados están aquellos que se originan cuando la mirada capta un destello de luz. Las BCI basadas en su detección, utilizan esta información para determinar la dirección con la que mira el usuario, la cual corresponde con el indicador de decisión. Un ejemplo de este tipo de sistemas consiste en un teclado virtual con una rutina en la que se iluminan, una por una, todas las letras. El usuario debe mirar directamente la letra que desea seleccionar, cuando dicha letra se enciende, el potencial registrado en el EEG presenta mayor amplitud que cuando las otras lo hacen, de este modo, es posible identificar al carácter deseado [13,14].

Otro tipo de potenciales visuales provocados se origina cuando se observa un destello que se enciende y apaga a cierta frecuencia. Este fenómeno incrementa la actividad registrada en el hemisferio occipital, la cual presenta componentes con la misma frecuencia con la que se hizo la estimulación. A partir de la caracterización de la frecuencia registrada en el EEG del usuario es posible determinar el estímulo al que atiende y, de este modo, realizar la acción correspondiente. Una aplicación basada en este tipo de actividad utiliza una pantalla en la que aparecen varios botones virtuales que se encienden y apagan a distinta frecuencia. El usuario debe fijar su atención en el botón que desea seleccionar para que, a partir de los componentes de frecuencia observados, el sistema pueda inferir su elección [15].

### **1.3.2. Potencial P300**

Las BCI basadas en la detección de potenciales visuales provocados requieren de usuarios con amplio control visual que logren fijar la mirada en los estímulos relevantes y, así, consigan controlar la interfaz. Cierta tipo de potenciales provocados que han demostrado ser eficaces para el desarrollo de BCI, incluso para los usuarios más limitados, corresponden con los potenciales P300 (figura 1.3).

Esta clase de potencial consiste en un máximo que se observa en las señales de EEG registradas alrededor de las regiones parietales y se relaciona con una respuesta cognitiva que se origina, aproximadamente, 300 ms después de recibir el estímulo correspondiente, de ahí que se le conozca con esta nomenclatura [16, 17, 18].

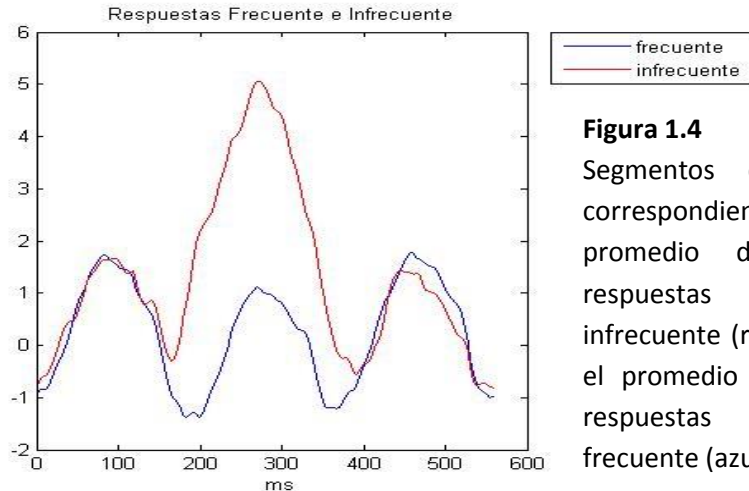


**Figura 1.3** El complejo P300 consiste en un máximo que se observa en la señal de EEG, aproximadamente, 300 ms después de haber expuesto al usuario al estímulo.

#### 1.4. Paradigma de Evento Raro (*Oddball*)

Existen muchos paradigmas con los que se pueden provocar potenciales P300, el más utilizado para el desarrollo de BCI, se conoce como paradigma de evento raro (*oddball*), el cual presenta una rutina de estimulación que genera dos tipos de respuestas, las de tipo frecuente, y las de tipo infrecuente. Como su nombre lo indica, a diferencia de las respuestas de tipo frecuente, las cuales se presentan de manera regular, las respuestas de tipo infrecuente ocurren de manera esporádica y aleatoria, y son las que presentan el complejo P300, lo que las convierte en los indicadores de decisión del usuario. Los estímulos que se emplean para generar ambos tipos de respuestas pueden ser visuales o auditivos y se encuentran relacionados con un conjunto de símbolos que permiten controlar el sistema. Para ello, debe generarse un potencial P300 en el registro de EEG del usuario cuando ocurre el estímulo relacionado con el símbolo que desea elegir. De este modo, le es posible al sistema detectar que hubo una respuesta de tipo infrecuente y por lo tanto, a partir del símbolo asociado con el estímulo correspondiente, determinar la selección del usuario [19, 20].

En la figura 1.4 se muestran dos segmentos de señal<sup>1</sup> correspondientes con el promedio de varias respuestas de tipo infrecuente, y con el promedio de varias respuestas de tipo frecuente.



**Figura 1.4**  
Segmentos de señal correspondientes con el promedio de varias respuestas de tipo infrecuente (rojo) y con el promedio de varias respuestas de tipo frecuente (azul).

## 1.5. Matriz de Farwell y Donchin

En la figura 1.5 se presenta una matriz de caracteres, desarrollada por Farwell y Donchin, correspondiente con el módulo de estimulación de un sistema, basado en el paradigma *oddball*, que permite deletrear cualquier palabra. Para ello, se generan estímulos visuales que consisten en el encendido aleatorio de los renglones y las columnas de la matriz. El usuario debe poner atención en aquél que contiene a la letra requerida para formar la palabra que desea expresar, el cual a diferencia de los demás que no contienen el carácter de interés, corresponde con un estímulo que provoca una respuesta de tipo infrecuente. Una técnica utilizada para conseguir que el usuario genere este tipo de respuesta, consiste en pedirle que cuente el número de veces que se enciende la letra que desea seleccionar. Este proceso cognitivo, además de ser fácil de realizar, ha demostrado provocar potenciales P300 en la mayoría de los sujetos [21, 22].

<sup>1</sup> En los trazos que se muestran en la figura 1.4 pueden observarse los efectos que produce la estimulación visual. Como se mencionó en la sección 1.1.4.1, el hecho de observar un destello que se enciende y apaga de manera intermitente, genera componentes en las señales de EEG que presentan la misma frecuencia con la que se hizo la estimulación. En la gráfica es posible visualizar los componentes de frecuencia correspondientes con la pausa entre estímulos (125 ms) y la duración de cada uno de ellos (62.5 ms), lo que genera una componente senoidal con un periodo, aproximadamente, de 190 ms.



SEND					
A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	_

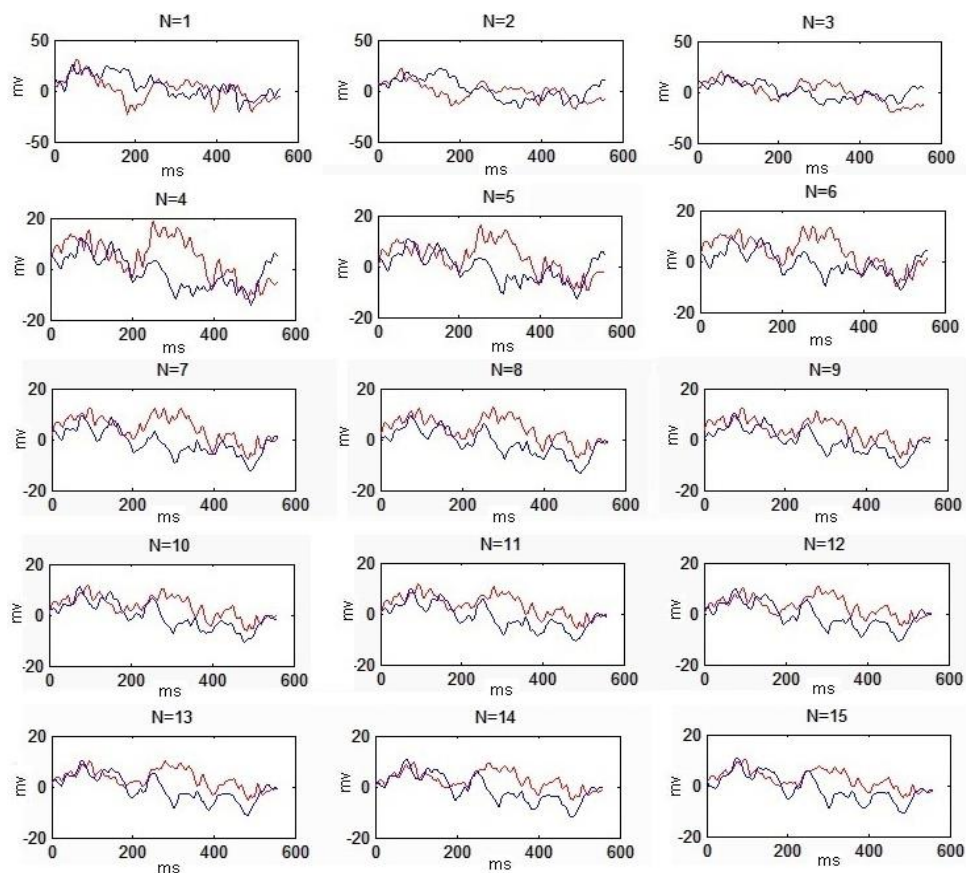
**Figura 1.5** Matriz de caracteres de Farwell y Donchin.

A partir de la figura se puede observar que una secuencia completa para conseguir elegir una letra de la matriz consta de doce estímulos, seis correspondientes con el encendido de los renglones, y otros seis correspondientes con el encendido de las columnas. Al finalizar una secuencia de estimulación, se espera contar con dos respuestas de tipo infrecuente, las cuales corresponden con el encendido del renglón y la columna que contienen al carácter de interés y cuyas posiciones determinan las coordenadas del símbolo que el sistema infiere. Las diez respuestas restantes corresponden con las de tipo frecuente y no se asocian con ninguna intención del usuario, sin embargo también son importantes puesto que lo que subyace en el módulo de procesamiento del sistema y le permite tomar decisiones, es un clasificador capaz de asignar las respuestas a una de las dos clases posibles: la infrecuente o la frecuente.

## 1.6. Motivación

En la práctica, para conseguir identificar la letra que desea seleccionar un usuario del deletreador de Farwell y Donchin, se requieren de muchas secuencias de estimulación. Esto se debe a que, en relación con la actividad de fondo que registra el EEG, la amplitud de la señal generada por el estímulo es considerablemente pequeña, es decir, la razón señal a ruido del sistema es muy baja y, difícilmente, puede detectarse el complejo P300.

Con el fin de mejorar esta relación se emplean técnicas de promediación coherente, las cuales además de resaltar las características que las señales tienen en común, eliminan los componentes relacionados con otro tipo de actividad, o con ciertos artefactos (fig. 1.6) [23].



**Figura 1.6** Secuencia de promedios de  $N=1,2,\dots,15$  respuestas de tipo infrecuente (trazo rojo), y de tipo frecuente (trazo azul). Obsérvese como, conforme aumenta el valor de  $N$ , los trazos se diferencian y exhiben la presencia o ausencia del complejo P300.

Entre más secuencias de estimulación se presenten, mejor es el resultado que se obtiene al promediar las respuestas correspondientes, lo que disminuye la tasa de error de clasificación. Sin embargo, el aumento del número de secuencias empleadas para determinar cada carácter, va en detrimento de la rapidez con la que el usuario deletrea las palabras, por esta razón, es un parámetro que debe elegirse con cuidado. En la mayoría de los casos, para asegurar una buena clasificación, se decide utilizar muchas secuencias en la rutina de estimulación. De este modo se obtienen buenos resultados no obstante, a la hora de analizar para cada secuencia la clasificación correspondiente, se observa que gran parte de las letras son clasificadas correctamente desde el tercer o segundo estímulo, y a veces, desde el primero, lo que revela la ineficacia de las secuencias subsecuentes. Sin embargo, con otras letras es diferente, y no es sino hasta después de muchos estímulos que consiguen ser clasificadas, de manera que, si se hubiera elegido presentar un menor número de secuencias, la clasificación habría sido errónea y, en la mayoría de los casos, la

palabra inferida por el sistema sería indescifrable. Este último párrafo exhibe lo inadecuado que resulta considerar el número de secuencias como un parámetro estático y, para todos los símbolos, designar la misma cantidad de estímulos. En este trabajo se presenta una propuesta para que dicho parámetro se determine de forma dinámica, de manera que el sistema sea capaz de ajustarlo automáticamente. Para ello, se calcula la probabilidad posterior asociada con la clasificación de cada estímulo. Este resultado puede entenderse como la probabilidad de que la clasificación se haya realizado correctamente y, así, mientras mayor sea este valor, mayor será la seguridad de haber clasificado bien la respuesta correspondiente. Bajo este esquema, el número de secuencias necesario para clasificar un carácter se establece cuando el valor de la probabilidad posterior correspondiente, logra sobrepasar un umbral señalado por el operador del sistema. Dicho de otro modo, el sistema seguirá estimulando al usuario mientras la probabilidad asociada con la clasificación permanezca por debajo del umbral.

## **1.7. Objetivos**

### **1.7.1. Objetivo General**

Instrumentar un esquema dinámico que determine la cantidad necesaria de estímulos para clasificar las respuestas de tipo infrecuente generadas bajo el paradigma oddball en interfaces cerebro-computadora.

### **1.7.2. Objetivos Específicos**

- Establecer un modo de agrupar la información obtenida para cada objetivo de la matriz de caracteres al término de cada secuencia de estimulación, con el fin de generar vectores de clasificación.
- Definir los criterios de decisión que permitan seleccionar los distintos símbolos de la matriz de caracteres.

## Capítulo 2

### 2. Marco Teórico

A lo largo del primer capítulo se describieron diferentes tipos de BCI que, en función de las características de la actividad cerebral que emplean para funcionar, ofrecen aplicaciones de diversa índole. Se mencionaron también los módulos que las constituyen y se dijo que el más importante corresponde con el módulo de procesamiento, el cual detecta las respuestas emitidas por el usuario y determina los comandos que le permiten controlar el sistema. Finalmente, se planteó el objetivo de este trabajo, el cual consiste en desarrollar un criterio para establecer, de manera automática, el número de estímulos necesarios para determinar cada uno de los símbolos que desea seleccionar el usuario de una interfaz basada en el paradigma *oddball*. Para ello, se define un módulo de procesamiento capaz de tomar decisiones a partir de una inferencia bayesiana que determina, en función de la probabilidad posterior asociada con la clasificación de cada respuesta, si es necesario continuar la estimulación o si, con cierto grado de certeza, se ha conseguido ya deducir la selección correcta y, por lo tanto, es posible detener el proceso.

Este capítulo presenta los conceptos necesarios para contextualizar el modelo que se propone a lo largo del tercero, y proporciona las herramientas con las que se justifican los procedimientos empleados para generar y probar dicha propuesta. Para ello, se comienza por desarrollar la noción de clasificador y se presentan las características de aquellos que utilizan una función discriminante. Posteriormente, se describe el algoritmo empleado para calcular, según este criterio, los parámetros que definen los clasificadores óptimos para cada usuario.

Los resultados generados por estos clasificadores, como se verá más adelante, corresponden con medidas de distancia que son transformadas, a partir de una función sigmoide logística, a valores de probabilidad posterior. Durante las últimas secciones de este capítulo se expone el teorema de Bayes y el concepto de probabilidad posterior, así como su relación con la sigmoide logística y el algoritmo empleado para generarla. Finalmente, se describe el método de validación cruzada, el cual es empleado con el fin de

evaluar los resultados y proporcionar una medida cuantitativa del desempeño de los clasificadores y, por consiguiente, del modelo propuesto.

## 2.1. Noción de clasificador

El componente principal del módulo de procesamiento de los sistemas basados en el paradigma *oddball* consiste en un clasificador capaz de asignar cada respuesta a una de dos clases posibles: la frecuente, o la infrecuente. Para ello se extraen las características más representativas que comparten los elementos de cada clase y se genera un nuevo conjunto de datos dentro de un espacio de dimensión  $D$ . La finalidad es conseguir que los grupos de cada clase se encuentren lo más separados posible, y a partir de las regiones correspondientes con su ubicación, definir las fronteras de decisión para asignar cualquier nuevo elemento a la clase que pertenece.

Las fronteras que delimitan las regiones correspondientes con cada clase se determinan a partir de la elección de algún modelo en particular. Por lo general para este tipo de aplicaciones se eligen modelos lineales de clasificación como, por ejemplo, aquellos que utilizan funciones discriminantes [24, 25].

Una vez elegido el modelo de clasificación los parámetros correspondientes se determinan a partir de un proceso conocido como *entrenamiento supervisado*. Durante esta etapa se utiliza un conjunto de datos etiquetados es decir, cuyas clases conocemos, con los que se busca ajustar el modelo de tal forma que, bajo algún criterio, los conjuntos de cada clase queden lo más separado posible y se consigan establecer las regiones correspondientes. Al terminar este proceso, cualquier dato nuevo no etiquetado, es clasificado en función de su posición dentro del espacio de clasificación.

### 2.1.2. Clasificadores de Discriminante Lineal

El objetivo de clasificar un conjunto de datos, consiste en asignar cada uno de ellos a una de  $K$  clases posibles, denotadas por  $C_k$ , donde  $k=1, \dots, K$ . En el mejor escenario, las clases se presentan disjuntas de modo que, cada elemento, corresponde a una única clase, la cual consiste en una de las  $K$  regiones que integran al espacio de dimensión  $D$  al que pertenecen los datos y, cuyos límites, se conocen como fronteras o superficies de decisión.

Los modelos lineales de clasificación, para delimitar las diferentes regiones, emplean funciones lineales que corresponden con hiperplanos definidos dentro del espacio de los elementos de entrada. Los conjuntos de datos con clases que pueden delimitarse de manera precisa mediante la disposición de estas funciones, se denominan linealmente separables.

El planteamiento más sencillo para clasificar un conjunto de datos consiste en proponer una función discriminante que, de manera directa, asigne cada vector  $\mathbf{x}$  (de dimensión  $D$ ) a la clase correspondiente. Para ilustrar el método, consideremos el caso de dos clases dentro de un espacio en el que se define la siguiente función lineal:

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0, \quad (\text{Ec. 2.1})$$

la cual establece la proyección del vector de entrada  $\mathbf{x}$  sobre un vector de pesos  $\mathbf{w}$  perpendicular al hiperplano que separa las clases  $y$ , cuyo sesgo está determinado por  $w_0$ . Esta función discriminante toma cualquier vector de entrada  $\mathbf{x}$  y lo asigna a una de las dos clases, para ello, evalúa la expresión y calcula el valor de  $y(\mathbf{x})$  cuyo signo, es el que determina la clase a la que pertenece el elemento correspondiente, de manera que, si  $y(\mathbf{x}) \leq 0$ , el elemento se asigna a la clase  $C_1$  y, si  $y(\mathbf{x}) > 0$ , se asigna a la clase  $C_2$ . La frontera de decisión correspondiente está definida por  $y(\mathbf{x})=0$  y consiste en un hiperplano de dimensión  $(D-1)$ .

En términos más simples, se puede visualizar la proyección de los vectores de entrada como una medida de distancia con respecto a la frontera de decisión, de manera que el signo asociado con  $y(\mathbf{x})$ , en realidad indica de qué lado del hiperplano se encuentra cada vector. Bajo este esquema, cada una de las regiones correspondientes con las clases  $C_1$  y  $C_2$ , están delimitadas por el clasificador que define dicho hiperplano. Entre más lejos se encuentre algún vector de esta frontera, se puede afirmar con mayor seguridad que pertenece a la clase asociada con el signo correspondiente y, por el contrario, los vectores que se localicen muy cerca, presentan poca evidencia para tomar una decisión. El valor escalar que se obtiene al calcular la distancia de los vectores de entrada con respecto al hiperplano, implica la reducción del espacio original a una sola dimensión. En general, las proyecciones sobre una sola dimensión provocan una considerable pérdida de información y, elementos que en su espacio original podían separarse fácilmente, desde esta perspectiva, presentan un fuerte traslape que dificulta su clasificación. Para corregir esta situación, se ajustan los componentes de  $\mathbf{w}$  y se elige aquella proyección que maximice la separación entre clases [26].

### 2.1.3. Entrenamiento de los Clasificadores

El entrenamiento del clasificador consiste en calcular, a partir de un conjunto de datos etiquetados, los parámetros que definen el hiperplano que maximiza la separación entre clases. Una reducción conveniente para plantear el modelo, consiste en agregar al vector

de entrada un término  $x_0=1$  y definir a  $\tilde{w} = (w_0, w)$  y a  $\tilde{x} = (x_0, x)$ , de tal modo que la ecuación 2.1 queda expresada en los siguientes términos:

$$y(x) = \tilde{w}^T \tilde{x}. \quad (\text{Ec. 2.2})$$

Al utilizar esta notación, la tarea de hallar el hiperplano, se reduce a encontrar los coeficientes del vector  $\tilde{w}$  que mejor separan los elementos de cada clase. Para ello, se utiliza un algoritmo que, por sus siglas en inglés, se conoce como SWLDA (Stepwise Linear Discriminant Analysis), el cual consiste en un método sistemático de regresión que agrega o elimina términos del modelo en función de su contribución para satisfacer el criterio. La cantidad máxima de coeficientes considerados en el modelo final está limitada por un parámetro que debe definirse antes de iniciar el método.

El algoritmo comienza con el planteamiento de un modelo inicial  $\tilde{w}_0$  y, paso a paso, compara el poder explicativo con otros modelos a los que, de manera gradual, se les aumenta y disminuye el número de coeficientes. Para probar los modelos con o sin algún coeficiente potencial, se calcula el valor p de una prueba F; si dicho término no está considerado en el modelo, la hipótesis nula es que, al agregarlo, ese coeficiente tendrá un valor igual a cero. Cuando existe evidencia suficiente para rechazar la hipótesis, el término se inserta en el modelo. Por otro lado, si el término en cuestión sí se encuentra considerado en el modelo, la hipótesis nula es que el valor de ese coeficiente es igual a cero; si la evidencia para rechazarla no es suficiente, el término es eliminado.

A continuación se describen los tres pasos con los que se lleva a cabo el procedimiento descrito:

1. Ajuste del modelo inicial. Por defecto, se parte de un modelo que no especifica ningún término, de manera que los coeficientes comienzan a agregarse conforme avanza el algoritmo.
2. Si algunos de los términos no considerados en el modelo tienen un valor p menor a cierta tolerancia, es decir, si resulta improbable que al agregarlos al modelo tengan coeficientes iguales a cero, se inserta aquél cuyo valor p sea el mínimo y se repite este paso, de lo contrario, se ejecuta el siguiente.
3. Si algunos de los términos considerados en el modelo tienen un valor p mayor a cierta tolerancia, es decir, si resulta improbable que se rechace la hipótesis de que estos coeficientes tomen un valor de cero, se elimina aquél cuyo valor p sea el máximo y se regresa al paso dos, de lo contrario, termina el proceso.

En función de los términos que se consideren dentro del modelo inicial y, de acuerdo con el orden con el que se insertan y eliminan los coeficientes en los pasos subsecuentes, el método puede generar distintos resultados a partir del mismo conjunto de datos de entrenamiento. El algoritmo termina cuando se estabiliza el modelo y, con el paso de las iteraciones, dejan de observarse mejoras [27].

## 2.2. Teorema de Bayes

El teorema de Bayes es una herramienta que permite calcular la probabilidad posterior asociada con algún evento. Para ejemplificar el concepto supóngase que, en épocas invernales, la probabilidad de adquirir un resfriado es del 60%. Ahora bien, si se conoce la probabilidad de tener un dolor de cabeza provocado por un resfriado, llamémosle a este hecho  $p(C|F)$ , entonces, puede saberse la probabilidad de estar resfriado si se presenta un dolor de cabeza, la cual de acuerdo con el teorema de Bayes se expresa de la siguiente manera:

$$p(F|C) = \frac{p(C|F)p(F)}{p(C)} = \frac{p(C|F)p(F)}{p(C|F)p(F) + p(C|NF)p(NF)}. \quad (\text{Ec. 2.3})$$

Donde  $p(C|F)$  y  $p(C|NF)$  corresponden con las condicionales de clase y se leen, respectivamente, como la probabilidad de presentar un dolor de cabeza dado que se tiene un resfriado y la probabilidad de presentar un dolor de cabeza dado que no se tiene un resfriado.  $p(F)$  y  $p(NF)$  son las prevalencias de clase y consisten respectivamente en valores de 0.6 y 0.4 correspondientes con la probabilidad de tener gripa y la probabilidad de no tener gripa durante las épocas invernales; ambas como puede inferirse, son mutuamente excluyentes. De esta manera es posible calcular la probabilidad de que cualquier sujeto que presenta un dolor de cabeza pertenezca a la clase de los que tienen un resfriado, así como la probabilidad de que pertenezca a la clase de los que no lo tienen y cuya cefalea se atribuye a otra causante.

Llevado al problema de clasificar las respuestas que se obtienen al utilizar la matriz de Farwell y Donchin supóngase que, después de generar cierto estímulo, se desea conocer la probabilidad de que éste haya generado una respuesta de tipo infrecuente. Dado que una rutina de estimulación consta de doce intensificaciones, de las cuales sólo dos provocan respuestas que pertenecen a la clase de tipo infrecuente ( $C_{inf}$ ) y, las diez restantes, respuestas que pertenecen a la clase de tipo frecuente ( $C_{frec}$ ), puede deducirse que la probabilidad de que este hecho ocurra es del 16.66%, este valor corresponde con la prevalencia de la clase infrecuente  $p(C_{inf})$ , del mismo modo puede inferirse que, la prevalencia de la clase frecuente  $p(C_{frec})$  es del 83.33%. Ahora bien, supóngase además



que la señal presenta rasgos característicos de las respuestas de tipo infrecuente, la probabilidad de que ésta pertenezca a dicha clase crece considerablemente y puede calcularse, mediante el teorema de Bayes de la siguiente manera:

$$p(C_{\text{inf}}|x) = \frac{p(x|C_{\text{inf}})p(C_{\text{inf}})}{p(x)} = \frac{p(x|C_{\text{inf}})p(C_{\text{inf}})}{p(x|C_{\text{inf}})p(C_{\text{inf}}) + p(x|C_{\text{frec}})p(C_{\text{frec}})}, \quad (\text{Ec. 2.3})$$

donde  $p(x|C_{\text{inf}})$  y  $p(x|C_{\text{frec}})$  corresponden, respectivamente, con las condicionales de clase de las clases infrecuente y frecuente.

Existen fundamentalmente dos formas de resolver esta expresión y obtener un resultado que permita clasificar las respuestas. Una de ellas consiste en un enfoque que, ya sea explícita o implícitamente, modela las distribuciones involucradas. De esta manera se establecen las condicionales de clase para, posteriormente, evaluarlas en el teorema de Bayes junto con las prevalencias asociadas y calcular la probabilidad posterior a partir de la cual es posible tomar una decisión. El hecho de contar con los modelos de las densidades correspondientes, permite determinar nuevos datos sintéticos; por esta razón, son métodos que se conocen como generativos.

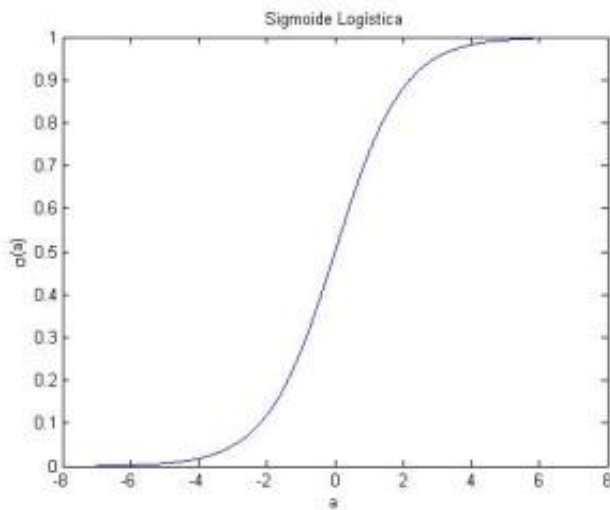
Otra manera de abordar el problema consiste en modelar directamente la probabilidad posterior a partir de un conjunto de datos de entrenamiento. Con dicho modelo es posible determinar la probabilidad posterior asociada con cualquier nuevo vector de entrada para, posteriormente, asignarlo a la clase correspondiente. Los métodos que directamente modelan la probabilidad posterior se conocen como modelos discriminativos.

Como se revisó en secciones anteriores, una alternativa que de manera directa, sin tener que calcular el teorema de Bayes, arroja la clasificación correspondiente consiste en establecer una función discriminante que, dado cualquier vector de entrada, determine la etiqueta de clase a la que éste pertenece. Bajo este esquema ninguna consideración se basa en probabilidades, y en comparación con los modelos mencionados anteriormente, representa un método mucho más sencillo de clasificación. No obstante, el hecho de contar con un resultado en términos de la probabilidad posterior proporciona una herramienta con la que pueden tomarse mejores decisiones basadas en la certeza que presenta la inferencia correspondiente. De esta manera se evita clasificar elementos con poca evidencia, lo que reduce los errores de clasificación, y por consiguiente, los riesgos que conlleva tomar malas decisiones [26].

En las siguientes secciones se presenta y describe la instrumentación de una técnica que permite, a partir de la distancia calculada por el discriminante de Fisher o por algún otro criterio, generar un valor en términos de la probabilidad posterior.

### 2.3. Sigmoide logística

Una sigmoide logística es una función que, entre sus propiedades más importantes, tiene la característica de mapear todo el eje real a un intervalo finito contenido en el rango (0,1).



**Figura 2.1.** Gráfica de una sigmoide logística. El dominio de la función comprende todo el eje real, a la mitad de este intervalo ocurre un punto de inflexión que determina la forma de “S” que presenta la gráfica, a partir de este punto, el trazo del lado izquierdo decrece de manera asintótica hacia el valor de 0, y el trazo del lado derecho crece asintóticamente hacia el valor de 1, estos extremos corresponden con el rango que toma la función. Obsérvese que el punto de inflexión ocurre a la mitad de este intervalo, y se localiza en (0,0.5).

El término *sigmoide* hace referencia a la morfología de la función, la cual se asemeja a una letra “S” (fig. 2.1) que se obtiene al evaluar la siguiente expresión:

$$\sigma(a) = \frac{1}{1 + e^{-a}}. \quad (\text{Ec. 2.4})$$

Supóngase ahora que el término **a** corresponde con el valor:

$$a = \ln \frac{p(x|C_{\text{inf}})p(C_{\text{inf}})}{p(x|C_{\text{frec}})p(C_{\text{frec}})},$$

de modo que al evaluarlo en la ecuación 2.4 se obtiene:

$$\sigma(a) = \frac{1}{1 + \exp\left(-\ln \frac{p(x|C_{\text{inf}})p(C_{\text{inf}})}{p(x|C_{\text{frec}})p(C_{\text{frec}})}\right)}$$

Cuyo resultado, al realizar la reducción correspondiente, es el siguiente:

$$\sigma(a) = \frac{p(x|C_{\text{inf}})p(C_{\text{inf}})}{p(x|C_{\text{inf}})p(C_{\text{inf}}) + p(x|C_{\text{frec}})p(C_{\text{frec}})},$$

el cual corresponde con el teorema de Bayes (Ec. 2.3) revisado en la sección anterior. Ahora bien, si se asume que las densidades de las condicionales de clase son gaussianas y que además tienen la misma matriz de covarianza, éstas pueden determinarse de la siguiente manera:

$$p(x|C_k) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu_k)^T \Sigma^{-1}(x - \mu_k)\right\}. \quad (\text{Ec. 2.5})$$

Al sustituir la expresión anterior (Ec. 2.5) en los términos de las condicionales de clase  $p(x|C_{\text{inf}})$  y  $p(x|C_{\text{frec}})$  en la ecuación 2.4 se obtiene:

$$\sigma(a) = \frac{1}{1 + e^{-(w^T x + w_0)}} = p(C_{\text{inf}}|x), \quad (\text{Ec. 2.6})$$

donde, como puede comprobar el lector, el término  $w$  corresponde con un vector cuya dimensión es la misma que la de los elementos de entrada  $x$ , y  $w_0$  representa un valor escalar, ambos quedan expresados respectivamente de la siguiente manera:

$$w = \Sigma^{-1}(\mu_{\text{inf}} - \mu_{\text{frec}}); \quad (\text{Ec. 2.7})$$

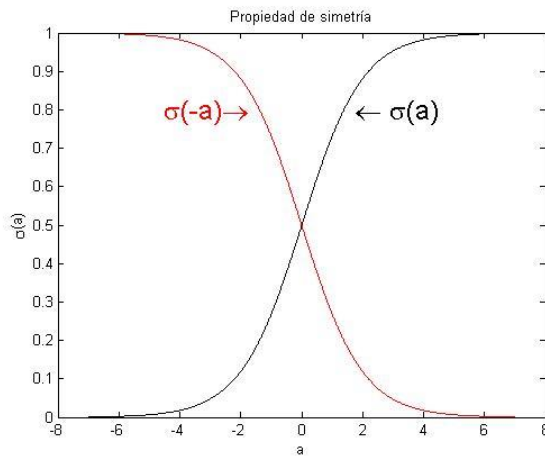
$$w_0 = -\frac{1}{2} \mu_{\text{inf}}^T \Sigma^{-1} \mu_{\text{inf}} + \frac{1}{2} \mu_{\text{frec}}^T \Sigma^{-1} \mu_{\text{frec}} + \ln \frac{p(C_{\text{inf}})}{p(C_{\text{frec}})}, \quad (\text{Ec. 2.8})$$

El resultado que se obtiene como argumento de la sigmoide logística es una función lineal, la cual como se revisó en la sección 2.2, corresponde con la ecuación de un hiperplano cuya evaluación determina la distancia que existe entre éste y cualquier otro punto. A partir de este resultado, es posible calcular la probabilidad posterior asociada con la clasificación de cada respuesta sin tener que determinar las condicionales de clase requeridas para la evaluación del teorema de Bayes puesto que, utilizar la sigmoide logística y sustituir en el argumento la distancia que existe entre cada respuesta y el hiperplano que bajo algún criterio separa las clases, representa un procedimiento equivalente.

Generar los resultados en términos de la probabilidad posterior ofrece una solución muy valiosa que permite cambiar el enfoque categórico de clasificación, en el cual se asignan las etiquetas en función del lado del hiperplano donde se ubica cada respuesta. En contraste, bajo un esquema bayesiano, se obtiene un valor entre 0 y 1 que representa la probabilidad de que determinada respuesta pertenezca a la clase  $C_k$ , de esta manera, es posible tomar decisiones de acuerdo con la certeza que presenta cada clasificación.

Otra característica importante de la sigmoide logística es la propiedad de simetría que se ilustra en la figura 2.2, la cual se expresa de la siguiente manera:

$$\sigma(-a) = 1 - \sigma(a). \quad (\text{Ec. 2.9})$$



**Figura 2.2.** Gráfica que ilustra los trazos correspondientes con  $\sigma(a)$  en negro, y con  $\sigma(-a)$  en rojo. Para el caso de dos clases, al obtener el valor correspondiente con  $p(C_1|x) = \sigma(a)$  se puede calcular de manera directa el valor de  $p(C_2|x)$  de la siguiente manera:

$$p(C_2|x) = \sigma(-a) = 1 - \sigma(a).$$

Esta propiedad cobra especial importancia para los casos que consideran solamente dos clases debido a que al determinar la probabilidad posterior asociada con una de éstas es posible obtener de manera inmediata la otra al efectuar, únicamente, la ecuación anterior. Este hecho tiene que ver con la relación mutuamente excluyente que guardan los eventos entre sí, puesto que la probabilidad de que una respuesta pertenezca

a una de las dos clases es 1, al conocer la probabilidad asociada con una de ellas, es posible determinar la otra en términos de su complemento [26].

## 2.4. Algoritmo de Platt

Para llegar al resultado de la sección anterior y conseguir expresar la probabilidad posterior en términos de la ecuación 2.6:

$$p(C_{\text{inf}}|x) = \frac{1}{1 + e^{-(w^T x + w_0)}}, \quad (\text{Ec. 2.6})$$

se hicieron algunas consideraciones. Por un lado, se asumió que las condicionales de clase  $p(x|C_{\text{inf}})$  y  $p(x|C_{\text{rec}})$  presentan una densidad de probabilidad con distribución gaussiana, y por otro lado, se hizo la suposición de que las matrices de covarianza correspondientes son iguales. Dichas consideraciones pueden generar un modelo impreciso que no se ajuste a los datos y refleje un comportamiento distinto. Para lograr corregir esta situación, John C. Platt propone una técnica que consiste en introducir dos parámetros,  $\alpha$  y  $\beta$ , en la expresión de la sigmoide logística (Ec. 2.6) de la siguiente manera [28]:

$$\sigma(y(x)) = \frac{1}{1 + e^{\alpha y(x) + \beta}}, \quad (\text{Ec. 2.7})$$

donde  $y(x)$  corresponde con el cálculo de la distancia que existe entre el hiperplano que separa las clases y cualquier dato dentro del espacio de clasificación la cual, como se revisó en secciones anteriores, se define como se muestra a continuación:

$$y(x) = w^T x + w_0. \quad (\text{Ec. 2.1})$$

Los parámetros  $\alpha$  y  $\beta$  pueden determinarse a partir de un proceso de entrenamiento con un conjunto de  $S$  datos;  $N_1$  de estos  $S$  datos etiquetados como  $y_i=1$ , y los restantes  $N_2$  como  $y_i=-1$ , una vez definido dicho conjunto, se procede a resolver el siguiente problema de verosimilitud:

$$\text{mín}_{z=(\alpha,\beta)} F(z).$$

Para un conjunto de datos definido por los vectores de características y sus respectivas etiquetas de clase, la función de verosimilitud puede escribirse en los siguientes términos:

$$p(t|w) = \prod_{i=1}^S p_i^{t_i} (1 - p_i)^{1-t_i}, \quad (\text{Ec. 2.8})$$

donde

$$p_i = \frac{1}{1 + e^{\alpha d + \beta}}, \quad y \quad t_i = \begin{cases} \frac{N_1 + 1}{N_1 + 2}, & \text{si } y_i = 1 \\ \frac{1}{N_2 + 2}, & \text{si } y_i = -1 \end{cases}, \quad i = 1, 2, \dots, S.$$

Ahora bien, si se considera el negativo del logaritmo de la expresión anterior (Ec. 2.8) es posible definir la función de costo en términos de la entropía cruzada de la siguiente manera:

$$F(z) = -\log p(t|w) = -\sum_{i=1}^S (t_i \log(p_i) + (1 - t_i) \log(1 - p_i)). \quad (\text{Ec. 2.9})$$

El algoritmo propuesto por Platt presenta dos inconvenientes importantes que pueden provocar ciertas dificultades numéricas al momento de ser ejecutado. Para evitar estos inconvenientes Hsuan-Tien Lin et al [29] presentan un algoritmo optimizado que resuelve el mismo problema de verosimilitud de una manera más eficiente. El ajuste del modelo de la sigmoide logística efectuado a lo largo de este trabajo es realizado de acuerdo con esta propuesta.

Vale la pena recalcar que, el hecho de introducir los parámetros  $\alpha$  y  $\beta$  y determinarlos a partir de un proceso de entrenamiento, permite lograr un mejor ajuste que responde al comportamiento y características de los datos, con lo que se consigue definir un modelo más adecuado que sorteja las consideraciones que fueron necesarias para determinar la probabilidad posterior en términos de la sigmoide logística.

## 2.5. Validación Cruzada

La validación cruzada es un análisis estadístico que permite evaluar el desempeño de alguna instancia generada para un modelo en particular. El objetivo de este tipo de análisis consiste en garantizar que los resultados sean independientes de la partición del conjunto de datos utilizados para generar y probar el modelo.

Para contextualizar el concepto, supóngase que se desea construir un modelo de uno o más parámetros a partir de un conjunto de datos de entrenamiento. El proceso de ajuste consiste en optimizar dichos parámetros para que el modelo se adapte a los datos lo mejor que pueda. Ahora bien, si el tamaño del conjunto de datos es pequeño o el número de parámetros del modelo es muy grande, puede ocurrir un fenómeno denominado sobreajuste, el cual se refiere al hecho de generar modelos sesgados, cuya eficacia se reduce cuando se evalúan datos nuevos.

Una manera de evitar el sobreajuste del modelo consiste en emplear técnicas de validación cruzada. Estos métodos se basan en dividir los datos en dos grupos complementarios. Uno de estos conjuntos representa los datos de entrenamiento a partir del cual se realiza el análisis, y el otro conjunto consiste en los datos de prueba con los que se lleva a cabo la validación correspondiente. De tal forma que los parámetros del modelo se ajustan únicamente en función de los datos de entrenamiento para, posteriormente, evaluar el resultado a partir de datos no vistos correspondientes con el conjunto de prueba [30].

### 2.5.1. Validación Cruzada de K Vías

Esta técnica consiste en dividir el total de datos en K subconjuntos; uno de ellos corresponde con los de prueba y, los (K-1) subconjuntos restantes, con los datos de entrenamiento (fig. 2.3).

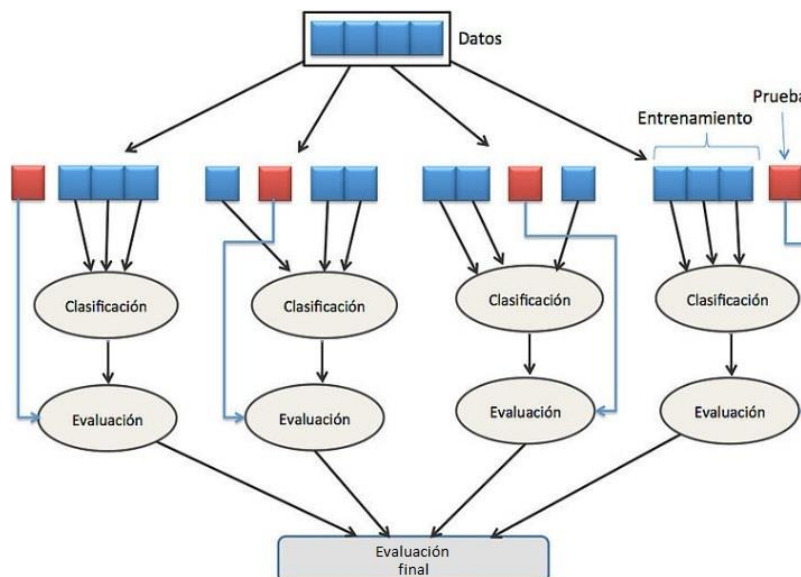


Figura 2.3. Esquema del método de validación cruzada de K vías para K=4.

El proceso de construir y evaluar el modelo se repite durante  $K$  iteraciones con cada uno de los posibles subconjuntos de datos. Finalmente, a partir de los resultados de cada iteración se calcula la media aritmética y se genera un único resultado mucho más preciso que el generado por el método de retención sin embargo, en comparación, su ejecución es considerablemente más lenta. En la práctica, el valor de  $K$  se elige de acuerdo con el tamaño del conjunto de datos; entre mayor sea éste, mayor es el número de particiones que pueden efectuarse.

### **2.5.2. Validación Cruzada Aleatoria**

Otra técnica de validación cruzada consiste en elegir los datos que integran cada uno de los subconjuntos de entrenamiento y de prueba a partir del conjunto total de datos disponibles; es decir, cada subconjunto es construido de forma independiente.

Esta técnica presenta el inconveniente de que algunas muestras no sean seleccionadas para formar parte del conjunto de prueba en ninguna de las particiones; así como puede suceder que otras muestras sean seleccionadas en más de una ocasión, de manera que los conjuntos de entrenamiento y de prueba de cada partición exhiban muchas similitudes entre sí, es decir, podría parecer que únicamente se está repitiendo el proceso de generar y evaluar el mismo modelo una y otra vez.

### **2.5.3. Validación Cruzada Dejando Uno Fuera**

Esta técnica se basa en separar los datos de tal forma que sólo uno de ellos pertenezca al conjunto de prueba, y el resto, al de entrenamiento. Bajo este esquema se generan  $N$  iteraciones, donde  $N$  corresponde con la cantidad de datos disponibles. El hecho de utilizar esta técnica brinda excelentes resultados no obstante, el gran número de iteraciones que deben realizarse, implica que se generen algoritmos lentos que demandan muchos recursos computacionales, situación que empeora conforme aumenta la complejidad del entrenamiento con el que se genera el modelo.



## Capítulo 3

### 3. Metodología

En los capítulos anteriores se sentaron las bases teóricas que justifican la metodología descrita a lo largo de las siguientes secciones; la cual en comparación con métodos convencionales, establece de forma automática el número necesario de secuencias de estimulación para identificar cada uno de los caracteres deletreados por el usuario.

Algunos métodos convencionales utilizan un clasificador lineal entrenado sobre los rasgos temporales de las señales de EEG para identificar, a partir de la distancia dirigida que presenta cada elemento con respecto al hiperplano de clasificación correspondiente, los objetivos de tipo infrecuente. Estos sistemas emiten  $N$  secuencias de estimulación para determinar cada caracter, donde  $N$  corresponde con un parámetro que establece el operador del sistema durante su configuración. Al finalizar cada secuencia de estimulación, la distancia dirigida que se obtiene a partir de cada respuesta se coloca en la  $n$ -ésima posición de un vector que contiene las  $(n-1)$  distancias calculadas en función de las respuestas sucesivas asociadas con cada uno de los doce objetivos, para  $n=1,2,\dots, N$ . Una vez concluidas las  $N$  secuencias de estimulación el sistema determina, a partir de cada uno de los doce vectores generados, la distancia acumulada y etiqueta el valor que, dentro del grupo de los seis renglones, alcanza el máximo valor absoluto asociado con el signo que representa la región que abarca la clase infrecuente; del mismo modo elige la columna correspondiente. En caso de que ninguno de los resultados presente el signo asociado con la clase infrecuente, se elige aquél cuyo valor absoluto sea el menor.

El hecho de calcular la distancia acumulada permite agrupar la información generada a partir de las  $N$  respuestas asociadas con cada objetivo y, en función de la evidencia que presentan de pertenecer a la clase infrecuente, etiquetar aquellos que muestran la mayor probabilidad. Una modificación importante que presenta la metodología propuesta consiste en generar  $\hat{N}$  clasificadores lineales para separar los datos de cada clase en función de la información generada hasta la  $n$ -ésima secuencia de estimulación, para  $n=1,2,\dots, \hat{N}$ . Donde  $\hat{N}$  representa un parámetro que debe establecer el

operador del sistema durante su configuración para indicar el número máximo de secuencias de estimulación permitidas en caso de que no sea posible, con base en la evidencia generada a partir de las respuestas de EEG del usuario, tomar antes una decisión. El énfasis señala la diferencia con el término  $N$  utilizado en los métodos convencionales con el fin de establecer el número de secuencias emitidas para seleccionar cada caracter, el cual representa un parámetro obsoleto bajo el esquema de la metodología propuesta.

Otra modificación importante consiste en el ajuste de  $\hat{N}$  modelos de funciones sigmoide-logística; los cuales permiten generar resultados en términos de la probabilidad posterior  $y$ , de esta manera, evaluar la evidencia que presenta cada objetivo de pertenecer a la clase infrecuente. Bajo este esquema el sistema es capaz de determinar, al término de cada secuencia de estimulación, si debe continuar con el proceso o, si a partir de la información generada, cuenta con la evidencia suficiente para seleccionar el renglón  $y$  la columna cuya intersección determina la posición del símbolo que desea elegir el usuario.

El desempeño del sistema se evalúa, para cada usuario, a partir de un análisis de validación cruzada sobre seis modelos generados para valores de  $\hat{N} = 2; 3; 4; 5; 6$  y  $8$ . En este rango de valores es posible observar el efecto que tiene el parámetro  $\hat{N}$  sobre los resultados que arroja el sistema. Construir un modelo para un valor de  $\hat{N} = 1$  carece de sentido porque, bajo esta condición, el esquema propuesto funciona de manera equivalente a como lo hacen los sistemas convencionales: únicamente elige el renglón  $y$  la columna correspondientes con los objetivos que presentan mayor probabilidad de ser de tipo infrecuente. Por otro lado como se verá más adelante, tanto la precisión como la tasa de deletreo, tienden a estabilizarse conforme crece el valor de  $\hat{N}$ . De manera que al elegir valores por encima de  $\hat{N} = 5$  dejan de apreciarse cambios significativos en los resultados  $y$ , debido al incremento en la demanda de datos de entrenamiento, no representa una buena decisión.

La partición de los datos para generar los conjuntos de prueba  $y$  entrenamiento, así como las condiciones de adquisición de los mismos, se describen en las dos primeras secciones de este capítulo. En la tercera  $y$  última sección se presenta, paso a paso, el proceso completo de entrenamiento del sistema. Para ello se emplea un ejemplo que ilustra, a partir de los datos generados por uno de los usuarios, el desarrollo de un modelo con un valor de  $\hat{N} = 4$ .

### 3.1. Criterios de Inclusión y Adquisición de la Información

Los datos utilizados se encuentran disponibles en una base de dominio público [31]  $y$  corresponden con una serie de archivos que contienen la información de varias sesiones

de deletreo en las que participaron diferentes usuarios. La adquisición se hizo a través de la aplicación P300\_Speller de BCI2000<sup>®</sup> [32]. El número de secuencias de estimulación emitidas para seleccionar los símbolos de la matriz de caracteres se eligió, para cada una de las distintas sesiones, dentro de un rango que varía de 1 a 15 secuencias.

Los usuarios participantes son de ambos géneros, todos con una edad que oscila entre los 21 y 25 años; se trata de estudiantes sanos de la Universidad Autónoma Metropolitana Unidad Iztapalapa a quienes se les pidió que no se medicaran ni fumaran en las horas previas al registro. Asimismo, se les solicitó que se presentaran al día siguiente de una jornada de sueño de al menos 6 horas. La adquisición de las señales de EEG se hizo de acuerdo con el sistema 10-20; se colocaron diez electrodos como se muestra en la figura 3.1 con la referencia ubicada en el lóbulo de la oreja derecha y la tierra en el mastoides del mismo lado. El amplificador utilizado corresponde con el modelo gUSBamp de 16 canales de la marca g.tec<sup>®</sup>.

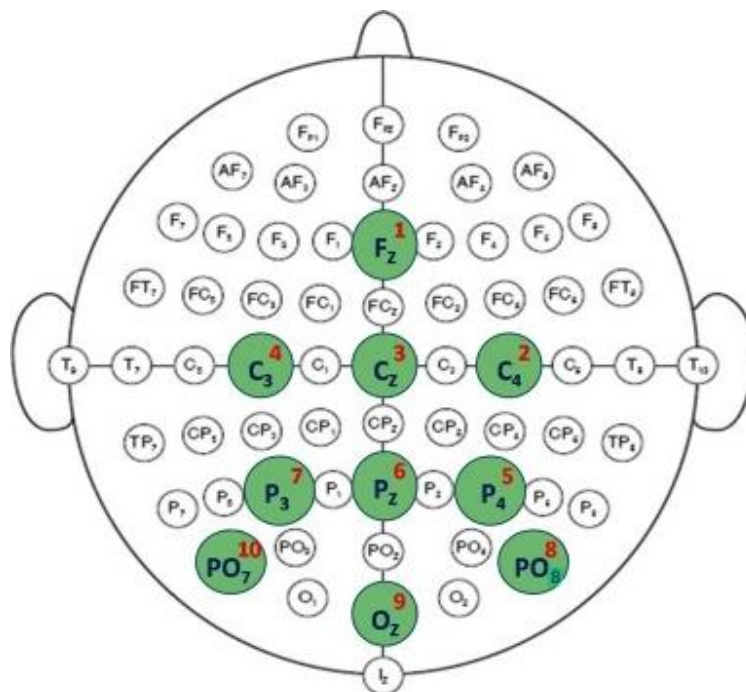


Figura 3.1. Configuración de los electrodos empleada durante las sesiones de deletreo.

En la tabla 3.1 se resumen las características del sistema con el que se llevó a cabo la adquisición, los parámetros se mantienen fijos a lo largo de todas las sesiones en las que participó cada uno de los usuarios.

Número de canales registrados	10
Frecuencia de muestreo	256 Hz
Filtro pasabanda	Chebyshev de orden 8, banda de paso [0.1-60 Hz]
Filtro Notch	Chebyshev de orden 4, banda de rechazo [58-62 Hz]
Pausa inicial	4 s
Duración de los estímulos	62.5 ms
Pausa entre estímulos	125 ms
Pausa posterior a una serie de estimulación	2 s
Pausa previa a una serie de estimulación	2 s
Tamaño de la matriz de caracteres	6 renglones x 6 columnas
Ancho y alto de la pantalla	1680 x 1050 píxeles

Tabla 3.1. Configuración del sistema de adquisición.

En la figura 3.2 se muestra el vector de estados de estimulación para una sesión correspondiente con el deletreo de una palabra formada por cinco letras y en la que, para seleccionar cada uno de los símbolos, el sistema presenta 15 secuencias de estimulación. En la gráfica se observa que el tiempo empleado para completar el proceso, corresponde con 188.75 s, el cual puede verificarse a partir de la información almacenada en la tabla 3.1.

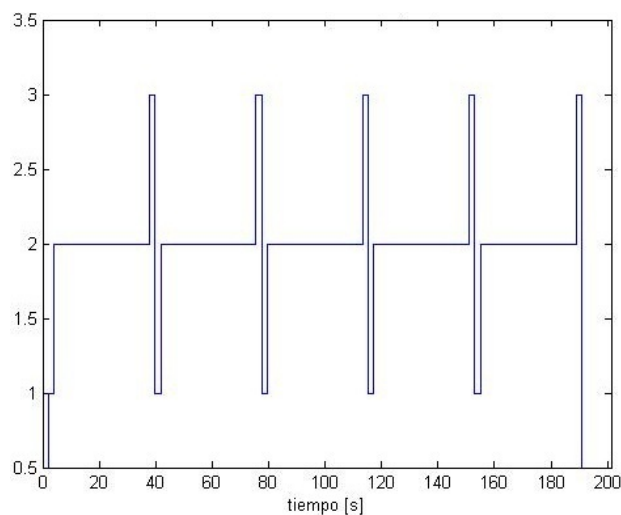


Figura 3.2. Vector de estados de estimulación. Los segmentos que toman el valor de 2 representan los tiempos en los que el sistema emite las secuencias de estimulación. Los cambios marcados con valores de 3 y 1 representan, respectivamente, la pausa posterior y previa de cada una de las secuencias de estimulación.

### 3.1.2. Filtrado de la Señal de EEG

Aun cuando el complejo P300 se encuentra predominantemente en el rango comprendido entre 0.1 y 4 Hz, también puede mostrar ciertos componentes en bandas de más alta frecuencia [33]. Por esta razón durante las pruebas realizadas a lo largo de este trabajo, se decidió abarcar un intervalo de mayor frecuencia comprendido entre 0 y 12 Hz; para delimitar este rango, se empleó un filtro pasabajos tipo Butterworth de cuarto orden.

En la gráfica del lado izquierdo de la figura 3.3 se muestran dos trazos correspondientes con el promedio de las respuestas de tipo infrecuente (trazo rojo), y con el promedio de las respuestas de tipo frecuente (trazo azul) registradas en uno de los canales de EEG con uno de los usuarios. La gráfica del lado derecho muestra las mismas señales después de aplicar el filtro mencionado y normalizarlas de tal modo que presenten un valor de media cero y un valor de varianza que oscile dentro del rango  $[-1,1]$ . Obsérvese cómo se han suavizado los trazos después de eliminar los componentes de alta frecuencia.

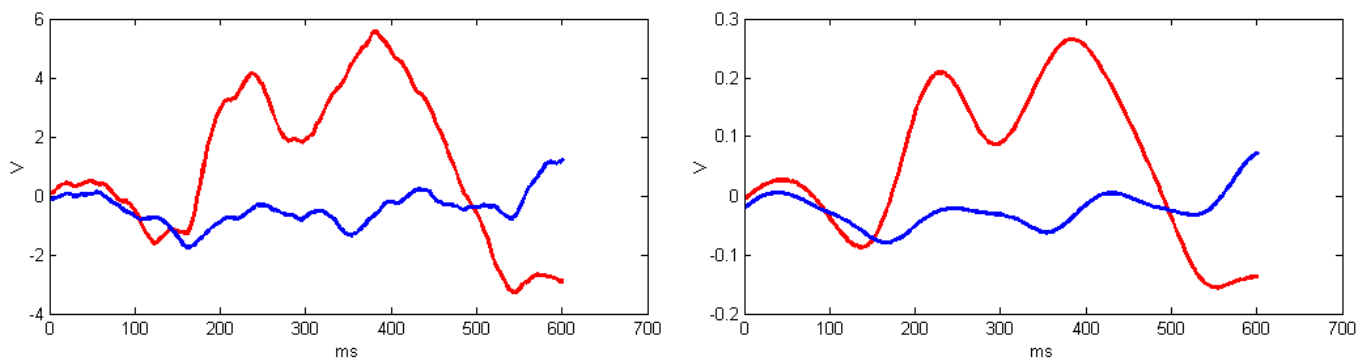


Figura 3.3. La gráfica del lado izquierdo muestra los promedios de respuestas de tipo infrecuente (trazo rojo) y los promedios de respuestas de tipo frecuente (trazo azul) registradas en uno de los electrodos que se colocaron sobre uno de los usuarios. La gráfica del lado derecho muestra las mismas señales después de aplicar un filtro pasabajos tipo Butterworth de cuarto orden con una frecuencia de corte  $f_c=12$  Hz.

### 3.2. Partición de los Datos

En la tabla 3.2 se presenta una lista de once palabras correspondientes con las once sesiones en las que participó uno de los usuarios; el número de secuencias emitidas para reconocer cada símbolo de la matriz de caracteres se indica en la columna del lado derecho.

Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
CENAR	15
COLOR	15
DULCES	15
NUBES	6
GALLETA	4
ESQUEMAS	3
ALEGRIA	3

Tabla 3.2. Lista de palabras deletreadas por uno de los usuarios.

Para ilustrar la construcción del método se describe a continuación el desarrollo de un sistema con un valor de  $\hat{N} = 4$  a partir de los datos de la tabla 3.2. La información generada durante estas sesiones se divide en dos partes: aproximadamente el 80% de los datos son utilizados en un análisis de validación cruzada de 10 vías y, el resto, en la validación de un sistema entrenado a partir de la totalidad de los datos (todas las vías).

Obsérvese que las únicas palabras cuyas letras fueron seleccionadas con menos de  $\hat{N}$  secuencias de estimulación son las últimas dos. Como se verá más adelante, los elementos empleados para la validación y prueba del sistema deben integrarse por  $\hat{N}$  respuestas sucesivas, de manera que antes de realizar la partición correspondiente, las palabras *ESQUEMAS* Y *ALEGRIA* deben ser descartadas. Las 9 palabras restantes contienen un total de 49 letras; de forma aleatoria se eligen 10 (aproximadamente el 20%) para validar el sistema y, las 39 restantes, para realizar las pruebas de validación cruzada.

Las pruebas de validación cruzada se llevan a cabo sobre 10 grupos de datos, cada uno formado por 3 letras elegidas de manera aleatoria y sin repeticiones, las cuales representan el conjunto de prueba; y por las 36 letras restantes que, en cada caso, corresponden con el conjunto de entrenamiento. En la tabla 3.3 se muestra un ejemplo de la partición de los datos, los conjuntos de prueba y entrenamiento representan una de las 10 instancias de los grupos empleados en la validación cruzada. Asimismo, la unión de ambos conjuntos, representa el grupo de entrenamiento a partir del cual se genera el sistema en cuya evaluación se emplea el conjunto de validación (segunda columna).

Palabras	Validación	Prueba	Entrenamiento
C <sub>1</sub> A <sub>2</sub> L <sub>3</sub> O <sub>4</sub> R <sub>5</sub>	C <sub>1</sub> L <sub>3</sub> C <sub>6</sub> O <sub>11</sub> I <sub>16</sub> C <sub>22</sub>	A <sub>2</sub> S <sub>19</sub> L <sub>34</sub>	O <sub>4</sub> R <sub>5</sub> A <sub>7</sub> R <sub>8</sub> I <sub>9</sub> N <sub>10</sub> S <sub>12</sub>
C <sub>6</sub> A <sub>7</sub> R <sub>8</sub> I <sub>9</sub> N <sub>10</sub> O <sub>11</sub>	O <sub>30</sub> D <sub>32</sub> C <sub>35</sub> B <sub>40</sub>		U <sub>13</sub> S <sub>14</sub> H <sub>15</sub> S <sub>17</sub> U <sub>18</sub> H <sub>20</sub>
S <sub>12</sub> U <sub>13</sub> S <sub>14</sub> H <sub>15</sub> I <sub>16</sub>			I <sub>21</sub> E <sub>23</sub> N <sub>24</sub> A <sub>25</sub> R <sub>26</sub> C <sub>27</sub>
S <sub>17</sub> U <sub>18</sub> S <sub>19</sub> H <sub>20</sub> I <sub>21</sub>			O <sub>28</sub> L <sub>29</sub> R <sub>31</sub> U <sub>33</sub> E <sub>36</sub> S <sub>37</sub>
C <sub>22</sub> E <sub>23</sub> N <sub>24</sub> A <sub>25</sub> R <sub>26</sub>			N <sub>38</sub> U <sub>39</sub> E <sub>41</sub> S <sub>42</sub> G <sub>43</sub> A <sub>44</sub>
C <sub>27</sub> O <sub>28</sub> L <sub>29</sub> O <sub>30</sub> R <sub>31</sub>			L <sub>45</sub> L <sub>46</sub> E <sub>47</sub> T <sub>48</sub> A <sub>49</sub>
D <sub>32</sub> U <sub>33</sub> L <sub>34</sub> C <sub>35</sub> E <sub>36</sub> S <sub>37</sub>			
N <sub>38</sub> U <sub>39</sub> B <sub>40</sub> E <sub>41</sub> S <sub>42</sub>			
G <sub>43</sub> A <sub>44</sub> L <sub>45</sub> L <sub>46</sub> E <sub>47</sub> T <sub>48</sub> A <sub>49</sub>			
E <sub>50</sub> S <sub>51</sub> Q <sub>52</sub> U <sub>53</sub> E <sub>54</sub> M <sub>55</sub> A <sub>56</sub> S <sub>57</sub>			
A <sub>58</sub> L <sub>59</sub> E <sub>60</sub> G <sub>61</sub> R <sub>62</sub> I <sub>63</sub> A <sub>64</sub>			

Tabla 3.3. Partición de los datos. De izquierda a derecha, la primera columna muestra los caracteres empleados para desarrollar y validar el sistema, cada uno etiquetado con un índice para rastrearlo durante la partición. La segunda columna contiene aproximadamente el 20% de los datos elegidos de manera aleatoria para utilizarlos durante la validación de un sistema desarrollado a partir del 80% restante. Las dos últimas columnas presentan una instancia de los conjuntos de entrenamiento y prueba utilizados en una de las pruebas de validación cruzada.

Las 15 letras seleccionadas con menos de  $\hat{N}$  secuencias de estimulación pertenecientes a las palabras descartadas, como se verá en la siguiente sección, son utilizadas en una de las etapas de entrenamiento del sistema, por lo que deben conservarse durante el proceso de partición.

### 3.3. Entrenamiento del Sistema

El diagrama de la figura 3.4 muestra el proceso de entrenamiento del sistema; cada etapa es representada mediante un bloque con los elementos de entrada y salida correspondientes. En las siguientes secciones se describe, paso a paso, cada una de las seis etapas que integra el proceso y se presentan los resultados generados a partir de los datos de entrenamiento de la tabla 3.3.

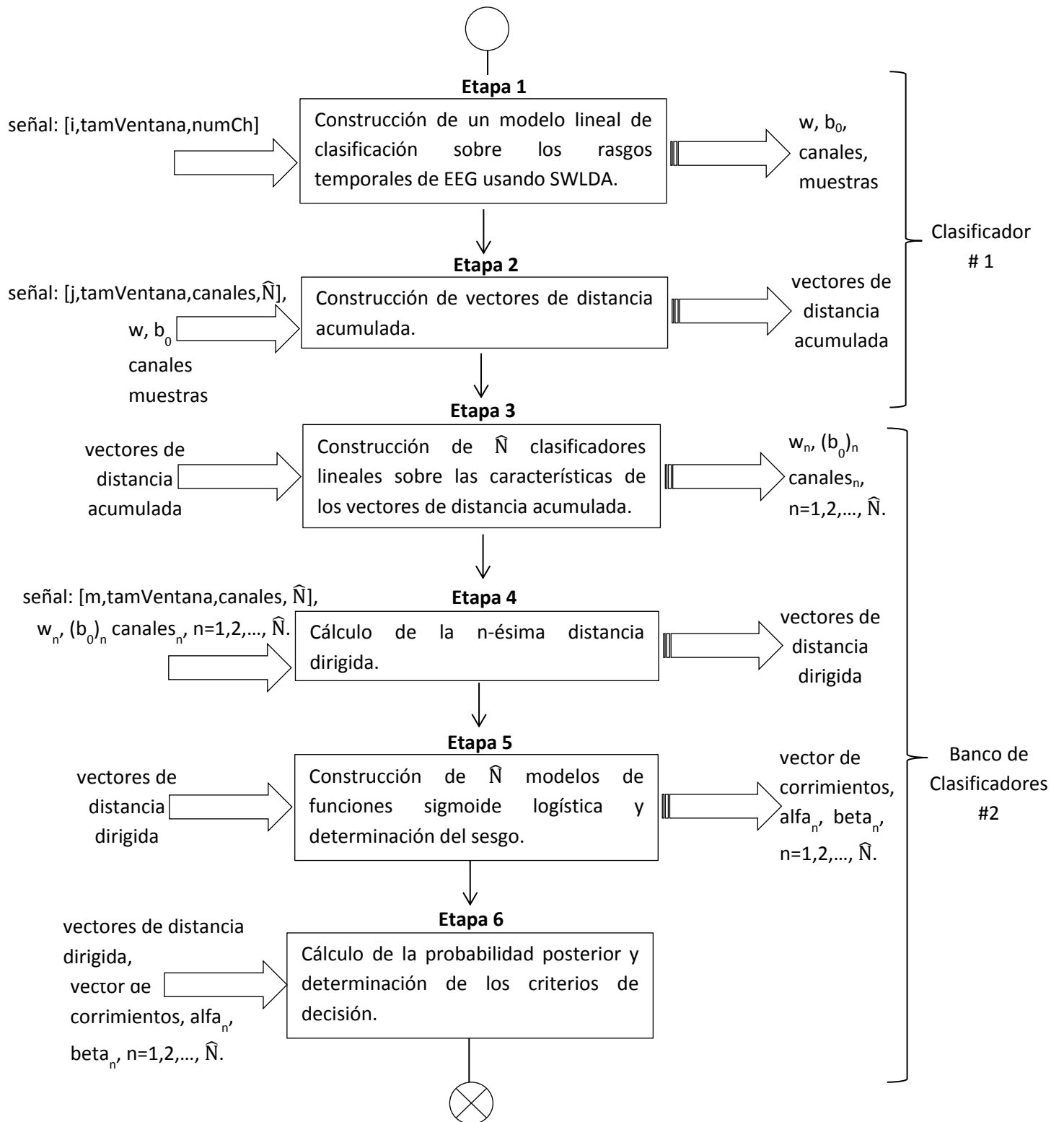


Figura 3.4. Esquema del entrenamiento del sistema.



### 3.3.1. Etapa 1: Construcción de Un Clasificador Lineal Sobre los Rasgos Temporales de EEG

Para entrenar este modelo se requieren datos correspondientes con las respuestas generadas en el EEG del usuario a partir de que se emite cualquier estímulo, sin importar la secuencia en que se presente. De manera que para generar el conjunto de entrenamiento pueden utilizarse los datos de las palabras descartadas, así como la información remanente de las letras seleccionadas con más de  $\hat{N}$  secuencias de estimulación.

La información relacionada con cada letra está contenida en un arreglo de [tamVentana,numCh,secuencias,#obj] muestras. Donde tamVentana indica el tamaño de la ventana empleada para extraer cada uno de los segmentos de señal, la cual se estableció de aproximadamente 600 ms (154 muestras) con el fin de cubrir un segmento de señal lo suficientemente grande como para asegurar que, en caso de tratarse de respuestas de tipo infrecuente, se consiga abarcar el complejo P300 en su totalidad, esto sin importar las diferencias que puedan manifestar las señales de los distintos usuarios. numCh es el número de canales utilizados durante la adquisición (10 canales); secuencias corresponde con el número de secuencias de estimulación emitidas para seleccionar cada caracter ({3; 4; 6 o 15 secuencias}) (ver tabla 3.2), y #obj (12 objetivos) representa el total de objetivos que contiene la matriz de símbolos, seis renglones y seis columnas. Dos de los 12 objetivos están etiquetados como tipo infrecuente (correspondientes con el renglón y la columna cuya intersección determina la posición del símbolo elegido por el usuario), y diez, como tipo frecuente (correspondientes con los renglones y las columnas restantes).

En el caso de las letras descartadas, las cuales fueron seleccionadas a partir de 3 secuencias de estimulación, se obtienen arreglos de [154,10,3,12] muestras. En realidad, cada arreglo contiene 3 respuestas para cada uno de los 12 objetivos, lo que equivale a un total de 36 respuestas ( $12 \cdot 3$ ). Ahora bien, puesto que se cuenta con 15 letras, se tiene un total de 540 respuestas ( $36 \cdot 15 = 540$ ); las cuales se representan, cada una, en un arreglo de [154,10] muestras. Por otro lado, dentro del conjunto de entrenamiento, se tienen 5 letras seleccionadas a partir de 6 secuencias de estimulación, esto implica que después de tomar las  $\hat{N}$  respuestas sucesivas de cada objetivo se cuenta con la información remanente de dos respuestas ( $6 - \hat{N} = 2$ ) por cada letra, lo que genera un total de 96 respuestas ( $5 \cdot 2 \cdot 12 = 120$ ), cada una, en un arreglo de [154,10] muestras. Del mismo modo, las 26 letras del conjunto de entrenamiento restantes seleccionadas a partir de 15 secuencias de estimulación presentan respuestas remanentes, cada una con la información de 11 respuestas ( $15 - \hat{N} = 11$ ) por cada objetivo, lo que genera un total de 3432 respuestas ( $26 \cdot 11 \cdot 12 = 3432$ ).

Todas estas respuestas suman un total de 4092 ( $540+120+3432=4092$ ), 3410 de tipo frecuente y 682 de tipo infrecuente, las cuales forman el conjunto de entrenamiento del primer bloque de clasificación. En la figura 3.5 puede observarse la representación de un elemento infrecuente de este conjunto; cada gráfica consiste en un segmento de señal de 154 muestras (tamVentana) registrado en cada uno de los 10 canales (numCh) a partir de que ocurre el estímulo correspondiente.

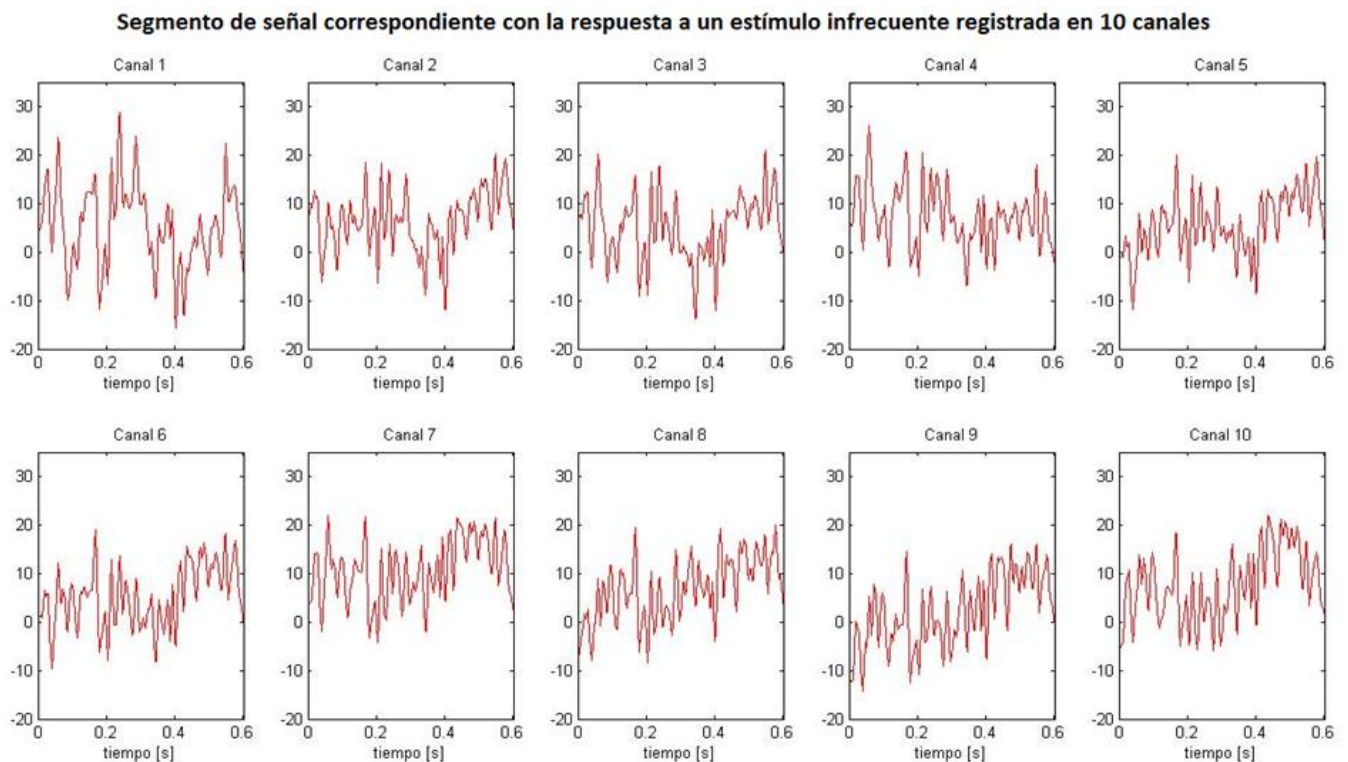


Figura 3.5. Representación de una respuesta a un estímulo de tipo infrecuente correspondiente con un arreglo de [154,10] muestras.

Una vez definido el conjunto de entrenamiento se aplica de regresión revisado en la sección 2.1.3 para determinar los parámetros que definen el hiperplano que mejor separa las clases. El resultado son cuatro elementos de salida correspondientes con los siguientes parámetros:

- **canales**, consiste en un vector cuyos elementos indican los canales a partir de los cuales, de acuerdo con el análisis de regresión (SWLDA, sección 2.1.3.), se obtiene información relevante y por lo tanto deben ser considerados en el modelo; los canales omitidos deben ser descartados<sup>2</sup>.

<sup>2</sup> Cada uno de los diez canales transmite la señal registrada por uno de los electrodos; el hecho de descartar aquellos que no aportan información relevante permitiría omitir la colocación de los electrodos correspondientes en sesiones posteriores.

- **muestras**, correspondiente con un vector de B elementos<sup>3</sup> que contiene las muestras de los canales considerados en el modelo que presentan información relevante, las muestras omitidas deben ser descartadas.
- **w**, vector de B coeficientes (perpendicular al hiperplano de clasificación).
- **w<sub>0</sub>**, valor escalar que determina el sesgo del hiperplano.

Los primeros dos parámetros permiten llevar a cabo la extracción de características sobre datos no vistos; conservar únicamente aquellos canales que demostraron detectar información relevante y tomar sólo las muestras que exhiben un comportamiento representativo de las clases frecuente e infrecuente. Al seleccionar esta información se reduce la dimensión de los datos, los cuales son mapeados a un nuevo espacio de dimensión B en el que se encuentra definido el hiperplano de clasificación por los parámetros **w** y **w<sub>0</sub>**.

Los resultados obtenidos para los parámetros **canales** y **muestras** se presentan en la tabla 3.4, obsérvese que el número de canales se mantuvo en 10, esto significa que de acuerdo con el análisis realizado sobre los datos, cada uno de los 10 canales aporta información relevante para diferenciar los elementos de las clases infrecuente y frecuente. En cada columna de la tabla se indican las muestras que contienen información útil de cada uno de los canales; el total de elementos considerados es de B=53 muestras, esto significa que después de efectuar la extracción de características definida por estos parámetros sobre un dato no visto de [tamVentana x numCh] elementos (fig. 3.5), se obtiene un vector de [1 x 53] muestras correspondiente con la consecutiva concatenación de los elementos considerados en cada canal. El valor de B determina la dimensión del espacio de clasificación donde se sitúa el hiperplano definido por los parámetros **w** y **w<sub>0</sub>**. Los B elementos obtenidos para el vector **w** se muestran en la tabla 3.5; el valor calculado para el sesgo que presenta el hiperplano es el siguiente:

$$w_0 = -0.74$$

---

<sup>3</sup> El valor de B se establece como parámetro del método de regresión, el cual determina el número máximo de coeficientes permitidos en el modelo. Durante las pruebas realizadas a lo largo de este trabajo, el número máximo de coeficientes permitidos se estableció con un valor de B=60. El hecho de que para este ejemplo, el número de coeficientes sea de B=53 se debe a que no se encontraron más elementos que mostraran características relevantes para diferenciar elementos de la clase infrecuente y frecuente, razón por la cual no hubo necesidad de truncar el modelo.

Muestras									
Canal 1	Canal 2	Canal 3	Canal 4	Canal 5	Canal 6	Canal 7	Canal 8	Canal 9	Canal 10
16	17	33	83	1	2	1	6	55	1
40	28	65		18	89	2	11	81	2
85	35	99		23	146	54	13	134	38
98	43	146		31		69	26	148	54
	96			40		96	24		69
	144			154		121	58		85
						127	76		128
							100		153
							142		
							154		

Tabla 3.4. Representación de los parámetros **canales** y **muestras** obtenidos para el primer modelo de clasificación. Los elementos indicados en cada columna corresponden con las muestras que contienen información relevante en los canales considerados por el modelo. El total de muestras,  $B=53$ , define la dimensión del espacio de clasificación.

Coeficientes									
Canal 1	Canal 2	Canal 3	Canal 4	Canal 5	Canal 6	Canal 7	Canal 8	Canal 9	Canal 10
0.1338	0.307	0.348	0.3398	-0.2869	0.3808	-2.05	0.8282	0.4873	2.2508
-0.1109	0.4706	-0.1421		-1.2489	0.6162	1.667	-2.6518	-0.2582	-2.0666
-0.0946	0.451	0.6155		0.4077	0.4283	0.8796	2.8212	-0.2331	0.2257
-0.2642	0.2224	0.2134		-2.4046		0.931	0.5407	-0.6395	-1.0757
	-0.2009			-0.6847		-0.3375	1.715		-1.0116
	-0.3483			-0.3582		-0.1892	-0.1638		-0.458
						0.4722	0.183		-0.3589
							0.2048		0.1306
							0.2876		
							0.336		

Tabla 3.5. Coeficientes del vector  $\mathbf{w}$  de  $B=60$  elementos. La posición de cada uno de ellos se indica en correspondencia con los de la tabla 3.4; de manera que los coeficientes indicados en cada canal corresponden, uno a uno, con las muestras seleccionadas.

### 3.3.2. Etapa 2: Construcción de Vectores de Distancia Acumulada

Una vez separadas las respuestas remanentes de los datos de entrenamiento de la tabla 3.3, cada elemento consiste en un arreglo de  $[\text{tamVentana}, \text{numCh}, \hat{N}, \#\text{obj}]$ , esto significa que, para cada objetivo, se tienen  $\hat{N}$  respuestas. En la figura 3.6 se muestra un elemento frecuente de este conjunto; los segmentos de señal representados en cada canal corresponden con las  $\hat{N}$  respuestas generadas en el EEG del usuario como resultado de las

$\hat{N}$  primeras intensificaciones emitidas por alguno de los objetivos de la matriz de caracteres. Obsérvese que a partir de las 36 letras que integran el conjunto de entrenamiento se obtienen 432 respuestas ( $36 \cdot 12 = 432$ ), 360 de tipo infrecuente y 72 de tipo frecuente.

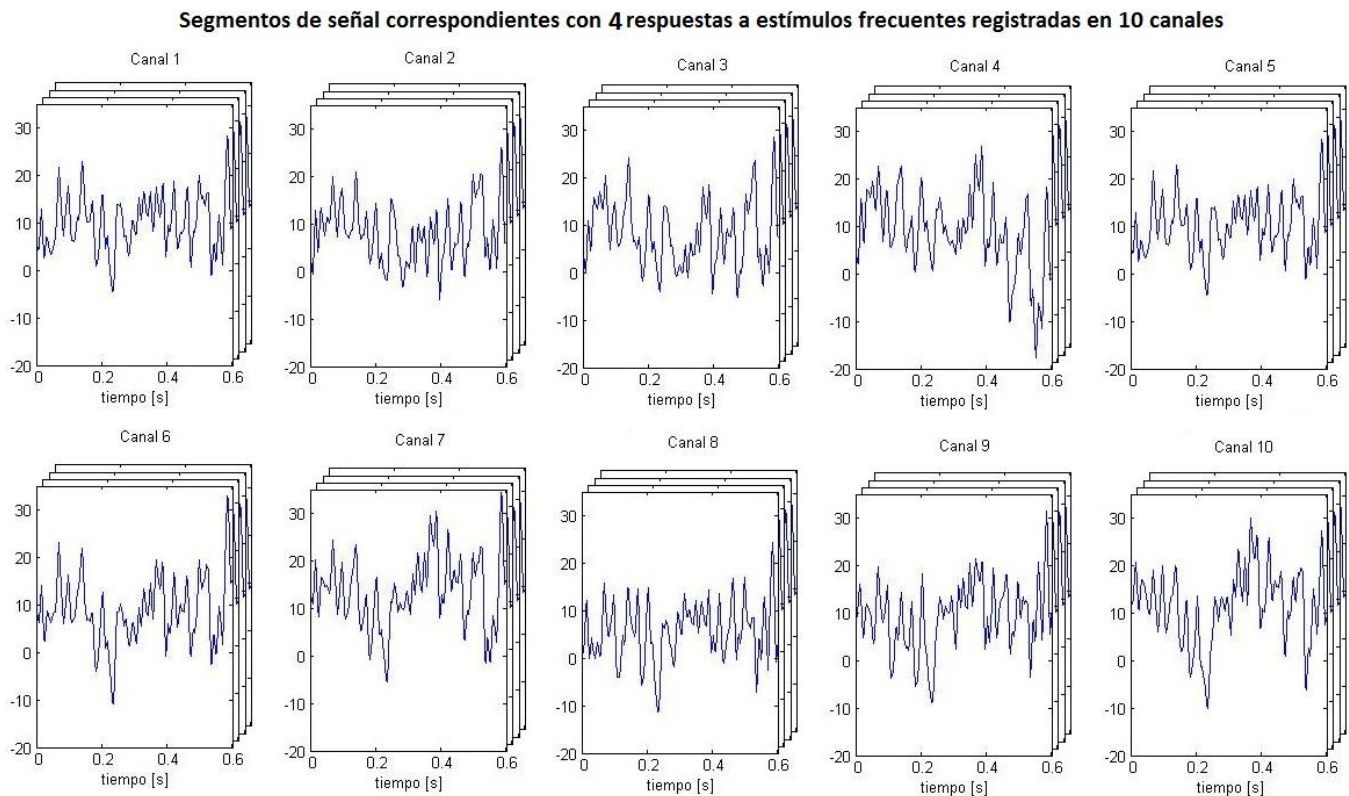


Figura 3.6. Representación de una respuesta a un estímulo de tipo frecuente correspondiente con un arreglo de  $[154, 10, \hat{N}]$  muestras.

Después de extraer las características definidas por los parámetros **canales** y **muestras**, cada una de las respuestas que integra los 432 elementos del conjunto de entrenamiento de la tabla 3.3 es mapeada al espacio de clasificación determinado en la primera etapa. Para cada respuesta situada en este espacio de clasificación, se calcula la distancia dirigida (Ec. 2.1) con respecto al hiperplano definido por  $\mathbf{w}$  y  $\mathbf{w}_0$  de la siguiente manera:

$$y(x) = \mathbf{w}^T \mathbf{x} + w_0.$$

El valor escalar resultante,  $y(x)$ , se coloca en la  $n$ -ésima posición de un vector de distancias, para  $n=1, 2, \dots, \hat{N}$ ; de manera que al finalizar el proceso se cuenta con 432 vectores de  $\hat{N}$  muestras cada uno. Finalmente, a partir de cada uno de estos vectores, se

genera otro de distancias acumuladas cuyo n-ésimo elemento se determina a partir de la siguiente expresión:

$$y_n(x) = y_{n-1}(x) + w^T x_n + w_0. \quad (\text{Ec. 3.1})$$

En la figura 3.7 se muestran cuatro proyecciones de los vectores de distancia acumulada generados a partir de los datos del conjunto de entrenamiento. Obsérvese que, conforme se incorpora la información de respuestas subsecuentes, los elementos de cada clase logran diferenciarse mejor.

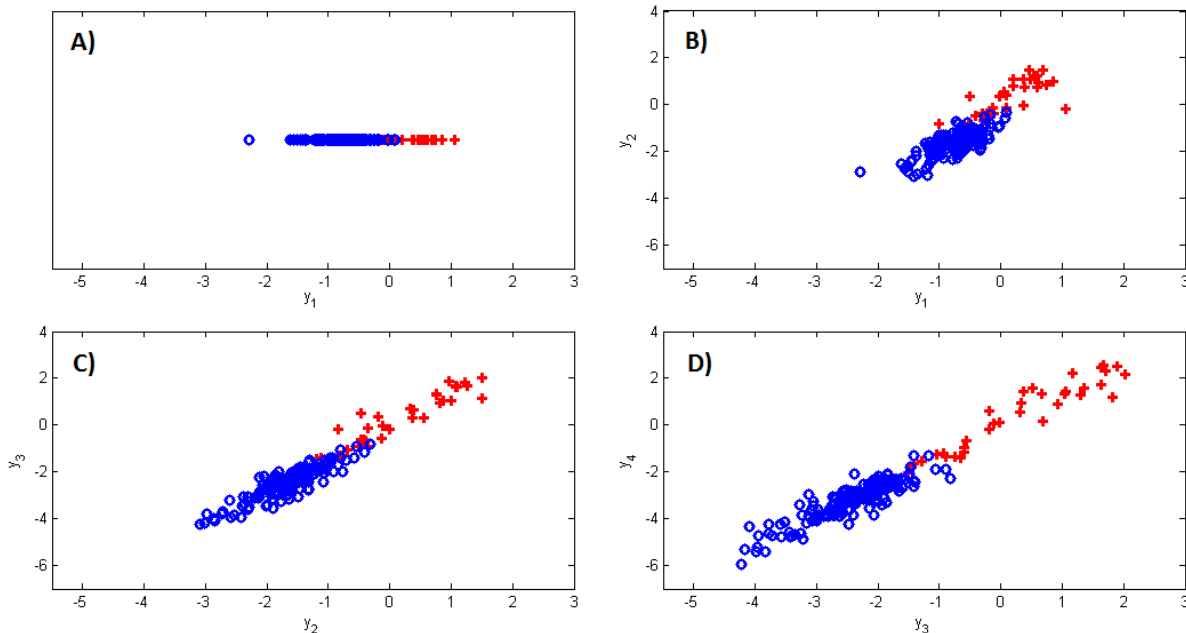


Figura 3.7. Proyecciones de los vectores de distancia acumulada,  $y_n$ , para  $n=1,2,\dots, 4$ . Los elementos infrecuentes se muestran en rojo y, los frecuentes en azul,

Los vectores de distancia acumulada consiguen agrupar la información de las respuestas actuales y las anteriores y exhiben diferencias claras entre los elementos de cada clase, los cuales presentan una separación entre sí que aumenta conforme crece el número de respuestas disponibles. Los vectores generados al finalizar cada secuencia de estimulación representan un nuevo conjunto de datos de entrenamiento a partir del cual se desarrolla un nuevo modelo de clasificación. De modo que bajo este esquema, se generan  $\hat{N}$  clasificadores para separar, en función de la información generada hasta la n-ésima secuencia de estimulación, los elementos de cada clase.

### 3.3.3. Etapa 3: Construcción de $\hat{N}$ Clasificadores Lineales Sobre Las Características de los Vectores de Distancia Acumulada

Para construir los clasificadores que integran este bloque se utiliza la mitad de los datos de entrenamiento generados en la etapa anterior (los primeros 216 vectores de distancia acumulada), esto con el fin de conservar un grupo de datos no vistos necesario para llevar a cabo las siguientes etapas. El entrenamiento de cada uno de los  $\hat{N}$  clasificadores consiste en aplicar sobre los primeros  $n$  elementos de los vectores de distancia acumulada el algoritmo de regresión descrito en la sección 2.1.3., para  $n=1,2,\dots, \hat{N}$ ; el resultado son 3 parámetros que determinan el espacio de clasificación donde se sitúa el hiperplano que mejor separa los elementos de las clases infrecuente y frecuente:

- **muestras<sub>n</sub>**, determina la extracción de características al indicar qué elementos de los vectores de distancia acumulada (para  $n=1,2,\dots, \hat{N}$ ) presentan información relevante, los elementos omitidos deben descartarse del modelo. La cantidad de elementos considerados determina la dimensión del espacio de clasificación correspondiente. Por lo general, conforme crece el valor de  $\hat{N}$ , el algoritmo de regresión (sección 2.1.3.) descarta los elementos asociados con las primeras secuencias de estimulación, y conserva aquellos que contienen más información asociados con las últimas.
- **w<sub>n</sub>**, vector de coeficientes (perpendicular al hiperplano de clasificación y con la misma dimensión que muestras<sub>n</sub>).
- **(w<sub>0</sub>)<sub>n</sub>**, valor escalar que determina el sesgo del hiperplano.

n	muestras	W <sub>n</sub>	(w <sub>0</sub> ) <sub>n</sub>
1	[1]	1.0094	-0.0201
2	[2]	0.6725	0.2233
3	[1,3]	[-0.1916,0.5508]	0.3027
4	[4]	0.3779	0.3244

Tabla 3.6. Parámetros obtenidos para cada uno de los  $\hat{N}$  modelos de clasificación.

En la tabla 3.6 se muestran los parámetros obtenidos para cada uno de los  $\hat{N}$  modelos de clasificación. Nótese que, dado que el conjunto de entrenamiento utilizado para desarrollar el primer modelo está integrado por vectores de un sólo elemento (fig. 3.7-A), el espacio de clasificación resultante está definido en una sola dimensión y, la

frontera de decisión resultante, consiste en un hiperplano de dimensión "0", es decir, con un punto. Esto significa que la única muestra correspondiente con el primer elemento de los vectores de distancia acumulada construidos a partir de los datos generados por este usuario, contiene información suficiente para diferenciar ambas clases; de otro modo, no sería posible determinar un modelo para  $n=1$ .

Por otro lado, el conjunto de entrenamiento a partir del cual se construye el modelo de clasificación para  $n=2$  consiste en datos definidos en un espacio de dos dimensiones, correspondientes con los dos primeros elementos de los vectores de distancia acumulada (fig. 3.7-B). Como se observa en la tabla 3.6., el modelo que se obtiene al efectuar el análisis de regresión, únicamente conserva el segundo elemento de los vectores. Esto indica que la información contenida en la segunda muestra es mayor que la contenida en la primera y que, en su mayoría se trata de información redundante; por lo que se puede descartar la primera muestra y obtener un modelo nuevamente definido en una sola dimensión.

En el entrenamiento del tercer modelo de clasificación se emplean datos definidos en tres dimensiones, los cuales corresponden con los tres primeros elementos de los vectores de distancia acumulada (ver proyección fig. 3.7-C). A partir de la tabla 3.6. se observa que el modelo resultante considera la primera y última muestra de estos vectores y que, la segunda muestra, es descartada. A partir de este hecho se puede inferir que la combinación del primer y tercer elementos presenta características convenientes para diferenciar los elementos de cada clase. Un rasgo característico de esta combinación, por ejemplo, corresponde con el salto que exhiben los valores de los elementos contenidos en la primera y en la tercera muestra; el cual representa el aumento en la evidencia que presentan los datos de pertenecer a alguna de las dos clases. De manera que, el modelo de clasificación resultante para  $n=3$ , está definido en un espacio de dos dimensiones y, la frontera de decisión correspondiente, consiste en un hiperplano de dimensión "1", es decir, con una recta.

Finalmente, el modelo de clasificación para  $n=4$  es entrenado a partir de los cuatro primeros elementos de los vectores de distancia acumulada (ver proyección fig. 3.7-D). Como se observa en la tabla 3.6, únicamente el último elemento de estos vectores es considerado en el modelo. Lo que implica que la información contenida en la cuarta muestra es mayor que la contenida en cualquiera de las tres primeras, o en cualquiera de sus combinaciones; razón por la cual son simplemente descartadas y, una vez más, se obtiene un modelo definido en una sola dimensión.

#### **3.3.4. Etapa 4: Cálculo de la n-ésima Distancia Dirigida**

Después de extraer las características definidas por el parámetro  $muestras_n$ , los 216 elementos que fueron apartados con el fin de conservar un conjunto de datos no vistos,



son mapeados a cada uno de los espacios de clasificación definidos por los  $\hat{N}$  modelos generados en la etapa anterior para, posteriormente, calcular la distancia dirigida con respecto al hiperplano correspondiente.

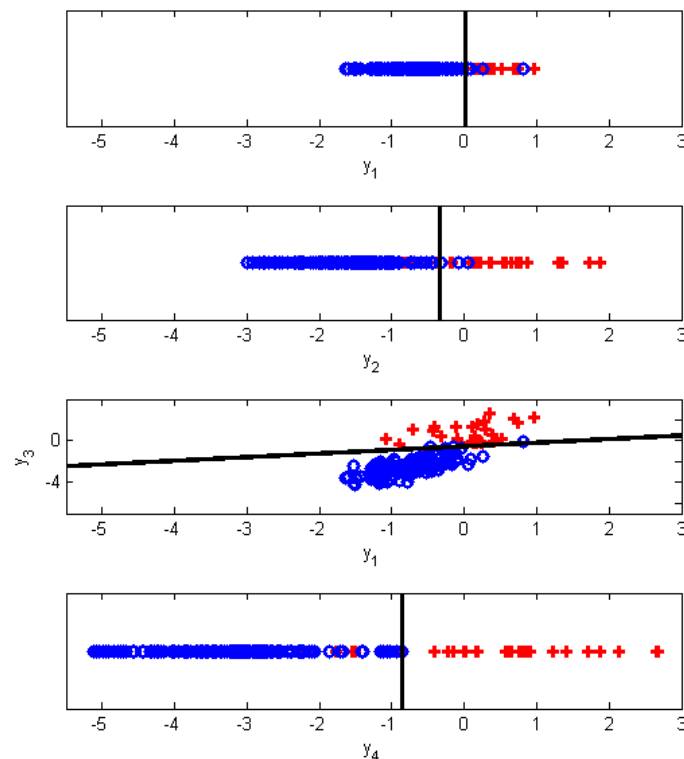


Figura 3.8. Cada gráfica muestra, de arriba hacia abajo, el mapeo del conjunto de datos no vistos a los espacios de clasificación definidos para  $n=1$ ; 2; 3 y 4. Los elementos rojos representan los datos infrecuentes y, los azules, los frecuentes.

En la figura 3.8 se observa, de arriba hacia abajo, el mapeo de los datos no vistos a los espacios de clasificación definidos para  $n=1$ ; 2; 3 y 4. Obsérvese que, conforme  $n$  aumenta, crece la separación entre los datos de cada clase y disminuye el número de elementos ubicados de manera incorrecta en algún lado de la frontera de decisión. En cada caso, para calcular la distancia dirigida con respecto al hiperplano de clasificación, se utiliza la siguiente expresión:

$$\hat{y}_n(x) = w_n^T \hat{x}_n + (w_0)_n, \quad (\text{Ec. 3.2})$$

donde el énfasis en los términos  $\hat{y}_n$  y  $\hat{x}_n$  se emplea para diferenciarlos de aquellos utilizados en la segunda etapa del entrenamiento del sistema. Nótese que en esta

expresión cada dato de entrada,  $\hat{x}_n$ , representa el mapeo de los los n primeros elementos de los vectores de distancia acumulada, para  $n=1,2,\dots, \hat{N}$ , al n-ésimo espacio de clasificación definido en la sección anterior; y que su evaluación determina la distancia dirigida,  $\hat{y}_n$ , que existe con respecto a la frontera de decisión definida por  $w_n$  y  $(w_0)_n$ ; después de calcular la distancia dirigida de los elementos con respecto al n-ésimo hiperplano de clasificación, para  $n=1,2,\dots, \hat{N}$ , se obtienen  $\hat{N}$  vectores de 216 elementos cada uno.

Como se mencionó, la etiqueta de clase correspondiente se asigna en función del signo asociado con la distancia dirigida, el cual establece de qué lado del hiperplano se encuentran los vectores de entrada. Entre más alejados de la frontera de decisión se sitúen los datos, mayor es la certeza que se tiene de clasificarlos correctamente; y por el contrario, entre más cerca se encuentren, menor es la evidencia que se tiene para tomar una decisión. De hecho un elemento situado sobre el hiperplano ( $\hat{y}_n=0$ ), presenta la misma evidencia de pertenecer a cualquiera de las dos clases. Bajo este argumento se intuye que es posible mapear cada una de las distancias dirigidas a un valor de probabilidad, de tal modo que las decisiones se tomen basadas en un nivel de evidencia y no de un modo categórico que ignore la posición específica de cada elemento. Los valores de distancia dirigida obtenidos en esta sección son utilizados a lo largo de la siguiente para ajustar  $\hat{N}$  modelos de funciones sigmoide-logística; estas funciones mapean todo el eje real a un intervalo finito comprendido entre “0” y “1” y permiten estimar la probabilidad de que algún dato no visto pertenezca a cualquiera de las dos clases.

### 3.3.5. Etapa 5: Construcción de N Modelos de Funciones Sigmoide-Logística y Determinación del Sesgo

En la primera columna de la figura 3.9 se incluye una copia de la figura anterior con el fin de analizar la correspondencia entre las distancias dirigidas que presentan los elementos con respecto a la frontera de decisión, y el modelo de la función sigmoide-logística ajustado a partir de estos resultados. Los modelos obtenidos para  $n=1$ ; 2; 3 y 4 se muestran, respectivamente, de arriba hacia abajo en la segunda columna de la figura. Para generar cada uno de estos modelos se aplicó la modificación del algoritmo de Platt (sección 2.4) sobre cada uno de los  $\hat{N}$  conjuntos de las distancias dirigidas calculadas en la sección anterior; con lo que se consigue determinar los parámetros  $\alpha_n$  y  $\beta_n$  que definen el cálculo de la probabilidad posterior en los siguientes términos (Ec. 2.7):

$$p(C_{\text{inf}}|\hat{y}_n) = \frac{1}{(1 + e^{(\alpha_n \hat{y}_n + \beta_n)})}$$

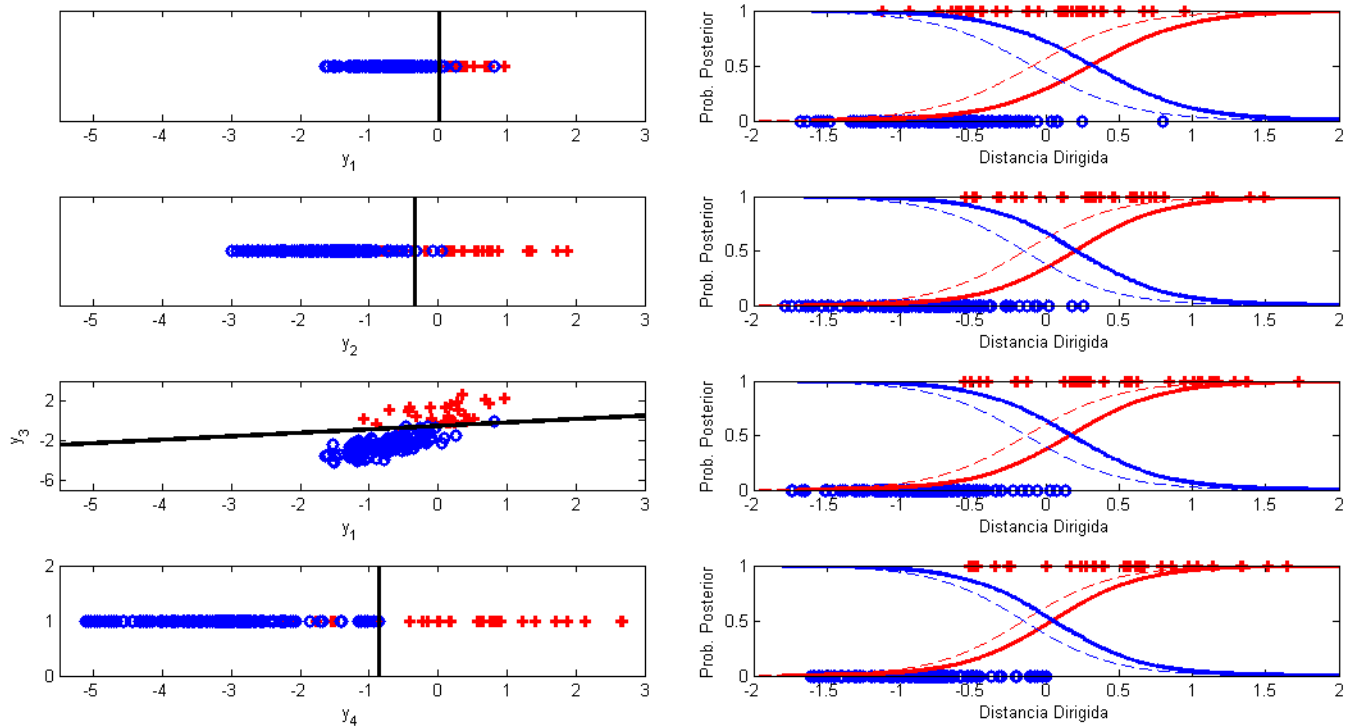


Figura 3.9. En la segunda columna de la figura se observan, de arriba hacia abajo para  $n=1;2;3$  y  $4$ , los modelos de las funciones sigmoide-logística ajustados sobre las distancias dirigidas calculadas a partir de los espacios de clasificación que se indican en la primera columna. Los datos rojos etiquetados con “1” incluidos en las gráficas de la segunda columna, representan las distancias de tipo infrecuente, y los azules etiquetados con “0” las de tipo frecuente.

Los datos etiquetados con “1” y graficados en rojo representan las distancias,  $\hat{y}_n$ , de los elementos de tipo infrecuente, y los datos azules etiquetados con “0” las distancias de tipo frecuente. Debido a la relación mutuamente excluyente que existe entre los eventos de tipo infrecuente y frecuente, es posible calcular la probabilidad de pertenecer a cualquiera de las dos clases en términos del complemento de la probabilidad posterior de la otra; como puede observarse a partir de la simetría que presentan las gráficas roja y azul correspondientes con los modelos de la sigmoide logística definidos, respectivamente, para las clases infrecuente y frecuente. El punto donde intersectan ambos trazos representa la frontera de decisión ( $\hat{y}_n=0$ ); cualquier dato que se localice sobre ésta tiene una probabilidad del 50% de pertenecer a cualquiera de las dos clases. Conforme crece la distancia que guardan los datos con respecto a este umbral, la probabilidad de pertenecer a una de las clases aumenta mientras disminuye la de la otra; lo que genera evidencia para asignar la etiqueta de clase correspondiente.

En las gráficas de las funciones sigmoide-logística se observa que existe un traslape importante entre los datos de cada clase; sobre todo en el caso de los modelos generados

a partir de valores pequeños de  $n$ , para los cuales fueron empleados datos que presentan muy poca separación entre los elementos infrecuentes y frecuentes. Es evidente que muchos de los datos cercanos a la frontera de decisión serán mal clasificados. Los errores de clasificación generan falsos positivos y/o falsos negativos; un falso positivo implica etiquetar como infrecuente una respuesta que en realidad es de tipo frecuente; lo que significa seleccionar una coordenada incorrecta de la matriz de caracteres. De modo que, si al finalizar una secuencia de estimulación se han seleccionado las coordenadas del renglón y la columna, y alguna o ambas, son incorrectas, el sistema despliega un carácter erróneo y detiene el proceso de estimulación para seguir con el siguiente símbolo. Por otro lado, un falso negativo implica etiquetar como frecuente una respuesta que en realidad es de tipo infrecuente. Esta información, aunque errónea, no conlleva una decisión tan desatinada como la que se infiere con los falsos positivos; en este caso únicamente se continúa con el proceso de estimulación hasta lograr clasificar el elemento infrecuente, con lo que se consigue cuidar la precisión de los resultados. Este último párrafo exhibe la conveniencia de sesgar el modelo de clasificación, y aun a costa de un incremento en el número de falsos negativos, evitar que surjan falsos positivos. El sesgo del modelo se lleva a cabo mediante un corrimiento del umbral de decisión; para ello se desplazan las gráficas hasta conseguir que los datos de tipo frecuente se agrupen de un sólo lado de la frontera de decisión, con lo que disminuye la probabilidad de clasificarlos de manera incorrecta. En cada una de las gráficas de la segunda columna de la figura 3.9 se presentan los modelos de las sigmoides logísticas originales mediante un trazo punteado; las gráficas con el corrimiento correspondiente, se muestran con líneas sólidas. Obsérvese que los falsos positivos se han reducido notablemente, el motivo por el cual no se propone un sesgo que los elimine por completo se debe a que, los elementos más alejados, corresponden con datos atípicos que no son representativos.

En la tabla 3.7 se muestran los parámetros obtenidos para cada uno de los  $\hat{N}$  modelos de las sigmoides logísticas junto con el corrimiento correspondiente.

$n$	$\alpha_n$	$\beta_n$	corrimiento
1	-2.957	-0.2736	-0.4
2	-3.2602	-0.4809	-0.35
3	-3.1717	-0.4208	-0.3
4	-3.2758	-0.5023	-0.2

Tabla 3.7. Parámetros asociados con cada uno de los  $\hat{N}$  modelos de las funciones sigmoide logística.

### 3.3.6. Etapa 6: Cálculo de la Probabilidad Posterior y Criterios de Decisión

Las dos primeras columnas de la figura 3.10 consisten en una copia de la figura anterior que se incluye para observar la correspondencia del mapeo de los datos  $\hat{y}_n$  a valores de probabilidad posterior para  $n=1,2,\dots,4$  que se observan en la tercera columna; los cuales son el resultado de aplicar la transformación definida por cada uno de los  $\hat{N}$  modelos de las funciones sigmoide-logística junto con el corrimiento correspondiente (segunda columna).

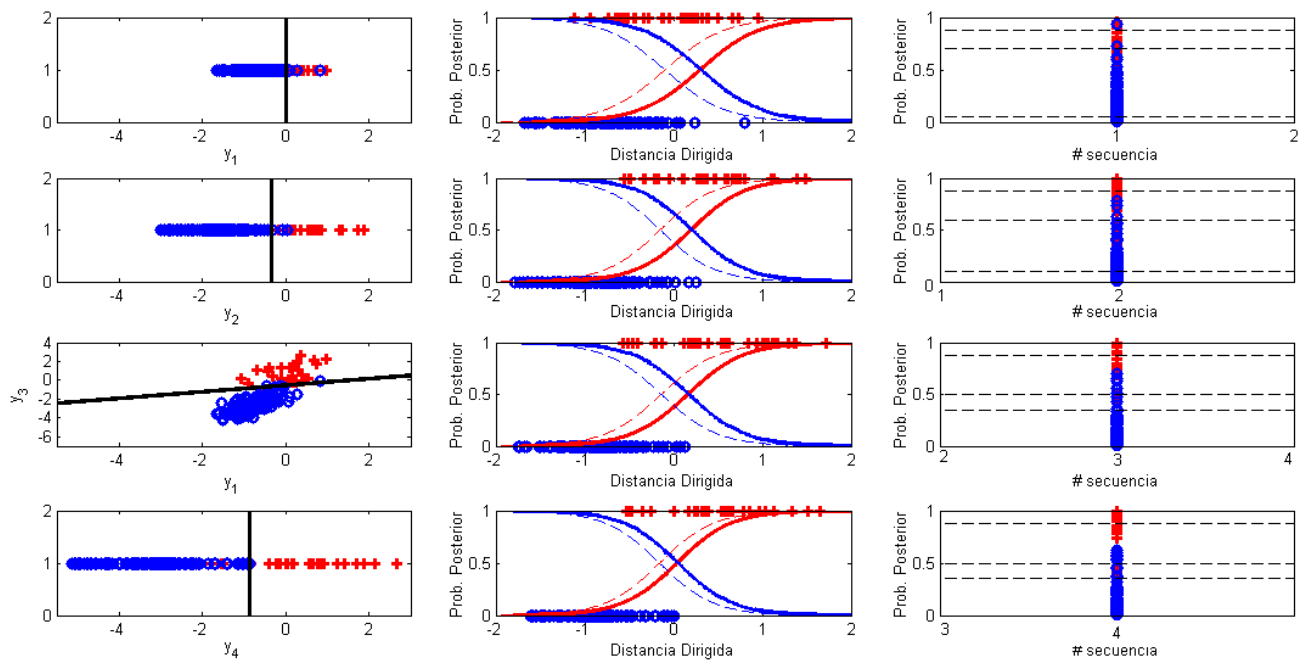


Figura 3.10. En la tercera columna se muestra, de arriba hacia abajo, el mapeo de las distancias dirigidas a valores de probabilidad posterior para  $n=1$ ; 2; 3 y 4. Estos resultados corresponden con el valor que toma la función sigmoide-logística para los correspondientes valores de distancia mostrados en la segunda columna. Las líneas ponteadas incluidas en cada gráfica de la tercera columna representan los límites de decisión a partir de los cuales es posible seleccionar el símbolo que desea elegir el usuario. Cada uno de ellos, como se verá más adelante, corresponde de mayor a menor con los límites: *maxPost*; *medPost* y *minPost*.

Nótese que conforme se incrementa el valor de  $n$ , los elementos pertenecientes a cada una de las clases se agrupan en conjuntos cada vez mejor definidos y separados entre sí. El hecho de contar con resultados en términos de la probabilidad posterior permite que el sistema tome decisiones, de manera que si algún objetivo muestra suficiente evidencia de pertenecer a la clase infrecuente consiga identificarlo y etiquetarlo; y por el contrario, si ningún objetivo presenta la certeza de ser de este tipo,

continúe con el proceso de estimulación hasta obtener información suficiente o hasta haber presentado las  $\hat{N}$  secuencias permitidas.

Como se mencionó, el símbolo elegido por el usuario se determina a partir de la intersección del renglón y la columna asociados con las respuestas de tipo infrecuente. Esto significa que de los seis renglones que integran la matriz de símbolos, se espera que sólo uno provoque una respuesta de tipo infrecuente y que, la de los cinco restantes, sea de tipo frecuente. Del mismo modo sucede con las columnas, por esta razón, la evidencia asociada con los objetivos se evalúa de forma independiente y se analizan por separado las respuestas generadas por cada grupo de renglones y columnas, lo que facilita el proceso de selección. Por supuesto, para conseguir que el sistema seleccione cualquier símbolo, es necesario reconocer tanto al renglón como la columna correspondientes; de manera que, si al finalizar una secuencia de estimulación únicamente se ha identificado uno de los objetivos, el proceso continúa hasta conseguir seleccionar ambos o, hasta haber presentado las  $\hat{N}$  secuencias permitidas.

Para evaluar la evidencia que presentan los objetivos de pertenecer a la clase infrecuente, se utiliza una serie de umbrales de decisión a partir de los cuales se establecen cuatro criterios que permiten identificar los objetivos de tipo infrecuente. Cada uno de estos umbrales se representa con una línea punteada en las gráficas de la tercera columna de la figura 3.10 y, de abajo hacia arriba corresponden, respectivamente, con el límite *maxPost*; *medPost* y *minPost*, los cuales se describen a continuación:

1. **Si existe algún valor de  $p(C_{inf}|\hat{y}_n)$  que sea mayor o igual que *maxPost*, elige la coordenada del objetivo con el valor más grande de  $p(C_{inf}|\hat{y}_n)$ .** El parámetro *maxPost* representa un valor, dentro del rango [0,1], que determina la probabilidad a partir de la cual se tiene suficiente evidencia para clasificar como infrecuente a cualquier elemento que la sobrepase. Durante las pruebas realizadas se eligió un valor *maxPost* = 0.88, esto implica que cualquier elemento con una evidencia mayor o igual al 88% de pertenecer a la clase infrecuente debe asignarse a la misma. Si no existe ningún elemento que presente tal evidencia se aplica el segundo criterio.
2. **Si existe un único valor de  $p(C_{inf}|\hat{y}_n)$  que sea mayor o igual que *medPost*, elige la coordenada del objetivo correspondiente.** El parámetro *medPost* consiste en un límite más permisivo que *maxPost* en cuanto a que, en promedio, se establece alrededor del 60%. No obstante, este criterio implica cierta rigidez puesto que sólo es posible tomar una decisión si existe un único valor de  $p(C_{inf}|\hat{y}_n)$  por encima de *medPost*. Si existen dos o más objetivos con valores de  $p(C_{inf}|\hat{y}_n)$  mayores que *medPost* se determina que no es posible tomar una decisión y se continúa con la adquisición de datos hasta conseguir diferenciar los niveles de evidencia de los

distintos elementos. El valor inicial del umbral *medPost* se establece alrededor de 0.7, y el final, sobre 0.5; esto significa que es necesaria una probabilidad del 70% de pertenecer a la clase infrecuente para conseguir tomar una decisión después de las primeras secuencias de estimulación; y una del 50% después de las últimas. Este valor puede reducirse debido a que los resultados se vuelven más confiables conforme aumenta la información de respuestas subsecuentes. En caso de que ningún valor de  $p(C_{inf}|\hat{y}_n)$  logre sobrepasar el límite *medPost* es posible aún tomar una decisión basada en el tercer criterio.

3. **Si existe sólo un valor de  $p(C_{inf}|\hat{y}_n)$  que sea mayor o igual que *minPost* elige la coordenada del objetivo correspondiente.** El tercer criterio aplica cuando ningún elemento muestra suficiente evidencia de pertenecer a la clase infrecuente. En tal caso es posible tomar una decisión por omisión, es decir, determinar cuántos elementos pertenecen a la clase frecuente para el conjunto de renglones y/o columnas correspondiente, si resultan ser cinco, entonces el único objetivo descartado puede asignarse a la clase infrecuente. De lo contrario, se continúa con la adquisición de datos hasta conseguir diferenciar los niveles de evidencia de los distintos objetivos. El parámetro *minPost* se establece en términos de  $p(C_{inf}|\hat{y}_n)$ , recordemos que (Ec. 2.9):

$$p(C_{frec}|\hat{y}_n) = 1 - (C_{inf}|\hat{y}_n).$$

De manera que un nivel de evidencia del 95% para asignar los elementos a la clase frecuente, representa un valor de *minPost*=0.05. El valor inicial del parámetro *minPost* se sitúa alrededor de 0.05 y, el final, se establece cercano a 0.35. Lo que implica que debe haber 5 objetivos con una probabilidad del 95% de pertenecer a la clase frecuente después de las primeras secuencias de estimulación para etiquetar como infrecuente al único descartado; conforme más información generada en secuencias subsecuentes está disponible esta tolerancia se reduce hasta el 65%. Si no se consigue satisfacer este criterio y el número máximo de secuencias de estimulación no ha sido excedido, el sistema continúa estimulando al usuario hasta que consiga diferenciar los resultados. Si el sistema no fue capaz de tomar una decisión después de la última secuencia de estimulación se aplica el cuarto criterio.

4. **Si ninguno de los criterios anteriores se satisface y el número de secuencias de estimulación es igual a  $\hat{N}$ , etiqueta como infrecuente aquél que presenta el máximo valor de  $p(C_{inf}|\hat{y}_n)$ .**

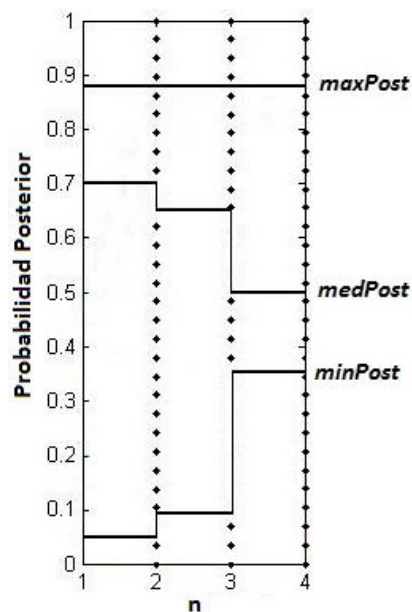
Después de aplicar los criterios de decisión sobre los datos, el sistema determina si consiguió seleccionar las coordenadas del renglón y la columna correspondientes. De ser así, selecciona el carácter indicado para desplegarlo en pantalla y comenzar con el

siguiente símbolo. Por el contrario, si sólo consiguió seleccionar las coordenadas de uno de los objetivos, o en el peor de los casos de ninguno, el sistema continúa estimulando al usuario hasta lograr tomar una decisión, o hasta haber presentado las  $\hat{N}$  secuencias de estimulación permitidas.

n	maxPost	medPost	minPost
1	0.88	0.7	0.05
2	0.88	0.6	0.1
3	0.88	0.5	0.35
4	0.88	0.5	0.35

Tabla 3.8. Umbrales de decisión en función del número de estimulación.

En la tabla 3.8 se muestran los valores de cada uno de los umbrales de decisión en relación con el tamaño de  $n$ , los cuales fueron determinados de manera heurística a partir de los resultados observados. Asimismo, se incluye en la figura 3.11 la representación gráfica de cada uno de estos umbrales.



3.11. Umbrales de decisión descritos en cada uno de los criterios.



## Capítulo 4

### 4. Resultados

A lo largo del capítulo anterior se describió el entrenamiento del sistema propuesto a partir del desarrollo de un modelo generado con un valor de  $\hat{N} = 4$ , utilizando los datos de uno de los usuarios. El proceso descrito se llevó a cabo con el conjunto de entrenamiento definido para una de las instancias de los grupos de datos empleados durante las pruebas de validación cruzada. Como se mencionó, cada una de estas instancias consta de un conjunto de entrenamiento formado por la información contenida en el registro de 36 letras y por un conjunto de prueba formado sólo por tres; asimismo, fueron apartadas 10 letras (aproximadamente el 20% de la totalidad de los datos de este usuario) con el fin de validar el sistema. Cabe recalcar que el hecho de efectuar las pruebas de validación cruzada tiene como objetivo asegurar la independencia entre el desempeño y la partición de los datos con los que se entrena y prueba el sistema. Por otro lado, una vez que se ha conseguido comprobar que existe esta independencia, es posible construir un sistema más robusto a partir de la totalidad de los datos y validarlo sobre otro conjunto de elementos no vistos.

En este capítulo se exponen, paso a paso, la operación y validación del sistema a partir de los datos contenidos en los registros de las tres letras que forman el conjunto de prueba de la misma vía de la que se obtuvo el conjunto de entrenamiento. Al finalizar, se incluyen los resultados obtenidos a partir de las nueve vías restantes; los resultados de la validación general del sistema para valores de  $\hat{N} = 2; 3; 4; 5; 6$  y  $8$  generados a partir de los registros de este usuario, así como de los trece restantes, se incluyen al final de este documento en el anexo titulado *Resultados de cada Usuario*.

Con el fin de comparar el desempeño de la metodología propuesta con el alcanzado por sistemas convencionales, se incluyen los resultados obtenidos con la aplicación P300\_Speller® de BCi2000 a partir de los mismos conjuntos de prueba. Es importante mencionar que, en el caso del entrenamiento, se utilizaron los mismos datos para desarrollar el sistema convencional que los empleados en el diseño de la primera

etapa de clasificación de la metodología propuesta, de hecho corresponden con los mismos clasificadores; de esta manera se evita inducir un sesgo al momento de evaluar y comparar ambos métodos.

Para evaluar los sistemas se utilizan tres indicadores de desempeño correspondientes con la precisión; el tiempo promedio de deletreo (letras/minuto) y el número de secuencias promedio necesario para determinar cada carácter (secuencias/letra).

#### 4.1. Operación del Sistema

En la figura 4.1 se presenta nuevamente la matriz de símbolos y un código de color; el cual emplea seis colores para diferenciar cada uno de los seis renglones y cada una de las seis columnas. Los renglones se cuentan de arriba hacia abajo y, las columnas, de izquierda a derecha.

	1	2	3	4	5	6
	C_1	C_2	C_3	C_4	C_5	C_6
R_1	A	B	C	D	E	F
R_2	G	H	I	J	K	L
R_3	M	N	O	P	Q	R
R_4	S	T	U	V	W	X
R_5	Y	Z	1	2	3	4
R_6	5	6	7	8	9	_

Figura 4.1. Matriz de caracteres y el correspondiente código de color empleado para identificar cada uno de los objetivos que integra al conjunto de los seis renglones y las seis columnas.

Los resultados que se presentan a continuación, correspondientes con los elementos generados para cada uno de los doce objetivos durante el proceso de prueba del sistema, se muestran de acuerdo con el código de color propuesto. Recordemos que, el conjunto de prueba de la tabla 3.3, está formado por tres letras que fueron apartadas durante todo el proceso de entrenamiento, de manera que corresponden con una

secuencia de datos no vistos a partir del cual es posible evaluar el desempeño del sistema. Las letras del conjunto fueron determinadas a partir de 15 secuencias de estimulación; las respuestas obtenidas para cada letra a partir de las 11 secuencias remanentes ( $15-\hat{N}=15-4=11$ ) son simplemente desechadas.

En la figura 4.2 se muestran 4 proyecciones de los vectores de distancia acumulada correspondientes con los elementos  $y_n$ , para  $n=1,2,\dots,\hat{N}$ ; calculados para cada uno de los 12 objetivos con el fin de seleccionar la primera letra del conjunto de prueba.

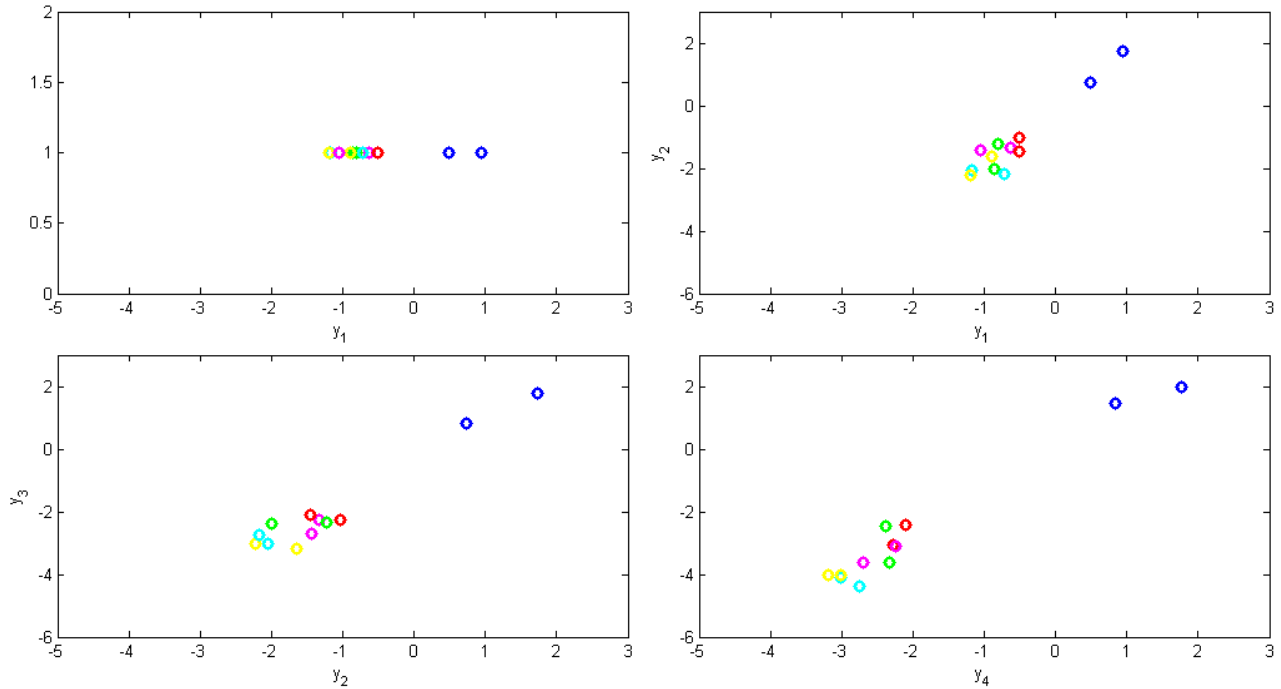


Figura 4.2. Proyecciones de los vectores de distancia acumulada generados a partir de las respuestas asociadas con la primera letra del conjunto de prueba.

Nótese cómo desde la primera secuencia de estimulación es posible observar dos elementos que se separan considerablemente del resto, lo que facilita el proceso de selección de los objetivos de tipo infrecuente. En la figura 4.3 se presenta el proceso de selección del objetivo infrecuente para el caso de los renglones; cada una de las tres gráficas representa de izquierda a derecha, para  $n=1$ , el mapeo de los vectores de distancia acumulada,  $y_n$ , al  $n$ -ésimo espacio de clasificación; la transformación a través de la evaluación del  $n$ -ésimo modelo de la sigmoide-logística de las distancias dirigidas,  $\hat{y}_n$ , a valores de probabilidad posterior,  $p(C_{inf}|\hat{y}_n)$ ; y la aplicación de los criterios de decisión sobre los resultados correspondientes. Obsérvese a partir de la primera gráfica que el elemento de color azul (primer renglón, fig. 4.1), se sitúa muy alejado del resto y en la región opuesta delimitada por la frontera de decisión. Nótese también en las dos gráficas restantes que la distancia dirigida resultante,  $\hat{y}_1$ , genera un valor mayor a  $medPost$

cuando es mapeada, a través de la sigmoide-logística correspondiente, a un valor de probabilidad posterior; mientras que la evidencia que presentan los otros cinco elementos se mantiene por debajo de este umbral, lo que satisface el segundo criterio de decisión y permite etiquetar como infrecuente al primer renglón de la matriz de símbolos.

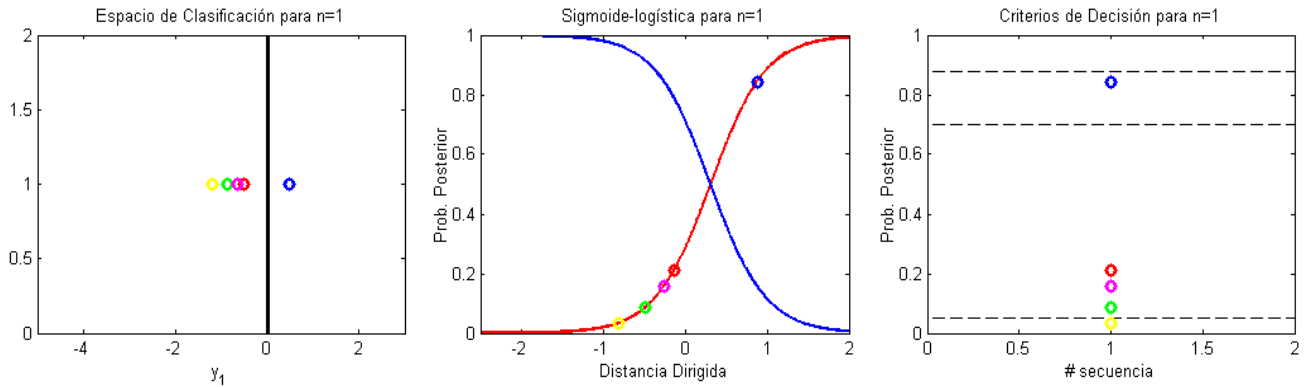


Figura 4.3. Proceso de selección para el caso de los renglones de la primera letra del conjunto de prueba. Las gráficas representan, de izquierda a derecha, el mapeo de los vectores de distancia acumulada al espacio de clasificación definido para  $n=1$ ; las distancias dirigidas con respecto al hiperplano de clasificación correspondiente junto con el modelo de la sigmoide-logística; y los resultados de transformar cada una de las distancias dirigidas, a partir de la sigmoide-logística, a valores de probabilidad posterior junto con los límites de decisión. Al finalizar la primera secuencia de estimulación el elemento de color azul, correspondiente con el primer renglón, es etiquetado con base en el segundo criterio como infrecuente.

De manera simultánea se realiza el mismo análisis sobre las respuestas asociadas con cada uno de los seis objetivos del conjunto de las columnas. Obsérvese en la figura 4.4 que el elemento azul (primera columna, fig. 4.1) se sitúa muy separado de los demás y en el lado opuesto del hiperplano de clasificación. Dado que el resultado que se obtiene al mapear la distancia dirigida correspondiente a través de la sigmoide logística definida para  $n=1$ , es mayor que  $medPost$  y, la evidencia generada por los cinco elementos restantes se mantiene por debajo de este umbral, es posible etiquetar como infrecuente al elemento azul con base en el segundo criterio. Esto significa que al finalizar la primera secuencia de estimulación el sistema consigue etiquetar la primera columna como infrecuente y debido a que, como al mismo tiempo fue identificado el primer renglón, es posible seleccionar el símbolo elegido por el usuario, el cual corresponde con la letra "A" situada en la posición (1,1) de la matriz de caracteres.

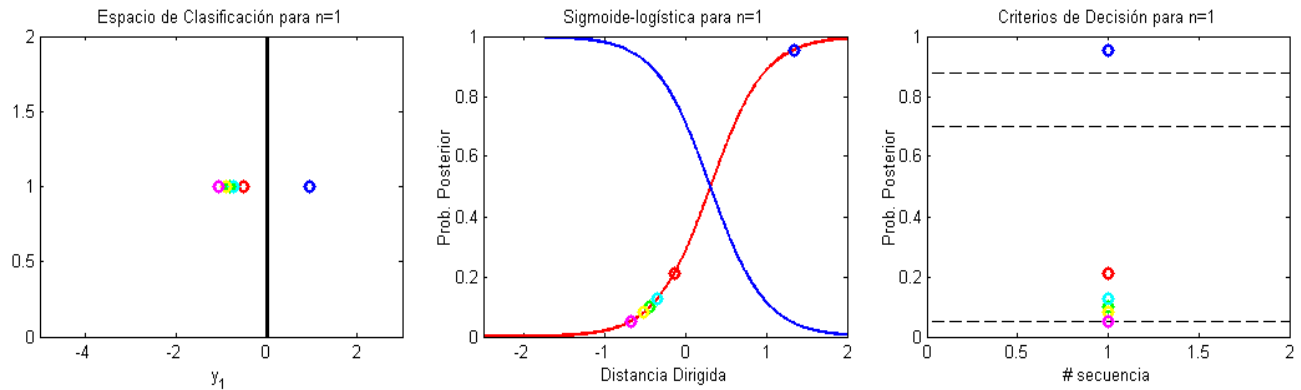


Figura 4.4. Proceso de selección para el caso de las columnas de la primera letra del conjunto de prueba. De izquierda a derecha se muestra para  $n=1$  el mapeo de los vectores de distancia acumulada al  $n$ -ésimo espacio de clasificación; las distancias dirigidas con respecto al hiperplano correspondiente junto con el  $n$ -ésimo modelo de la sigmoide-logística y el mapeo de las distancias a valores de probabilidad posterior. Al finalizar la primera serie de estimulación el elemento azul correspondiente con la primera columna es identificado como infrecuente.

A partir de la tabla 3.3 se observa que el primer caracter perteneciente al conjunto de prueba, en efecto, corresponde con la letra “A”; de manera que se ha llevado a cabo una clasificación correcta únicamente con las respuestas generadas en la primera secuencia de estimulación. Después de seleccionar la primera letra del conjunto de prueba, el sistema continúa con la segunda. Las proyecciones de los vectores de distancia acumulada,  $y_n$ , para  $n=1,2,\dots, \hat{N}$ , de cada uno de los doce objetivos se muestran en la figura 4.5.

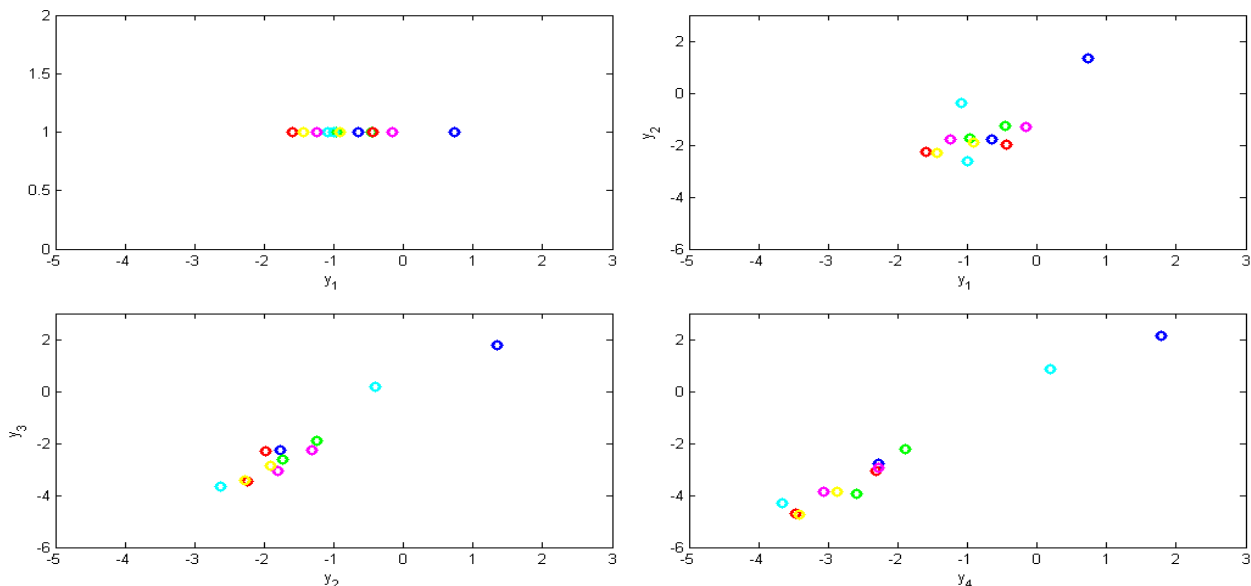


Figura 4.5. Proyecciones de los vectores de distancia acumulada generados a partir de las respuestas asociadas con la segunda letra del conjunto de prueba.

Obsérvese que al finalizar la primera secuencia de estimulación uno de los elementos se aleja considerablemente del resto, sin embargo no es posible identificar otro elemento que aparente ser de tipo infrecuente sino hasta el término de la tercera secuencia.

El proceso de selección para el caso de los renglones se ilustra en la figura 4.6. Obsérvese que al finalizar la primera secuencia de estimulación ninguno de los criterios se cumple, razón por la cual el sistema debe seguir estimulando al usuario. Después de la segunda secuencia de estimulación el elemento cian, el cual representa al cuarto renglón (fig. 4.1), consigue diferenciarse un poco del resto y, aun cuando ni siquiera se encuentra del lado opuesto del hiperplano de clasificación, es el único elemento que al mapear las distancias dirigidas,  $\hat{y}_2$ , a valores de probabilidad posterior, no presenta gran evidencia de pertenecer a la clase de tipo frecuente, por lo que, con base en el tercer criterio, puede ser etiquetado como infrecuente.

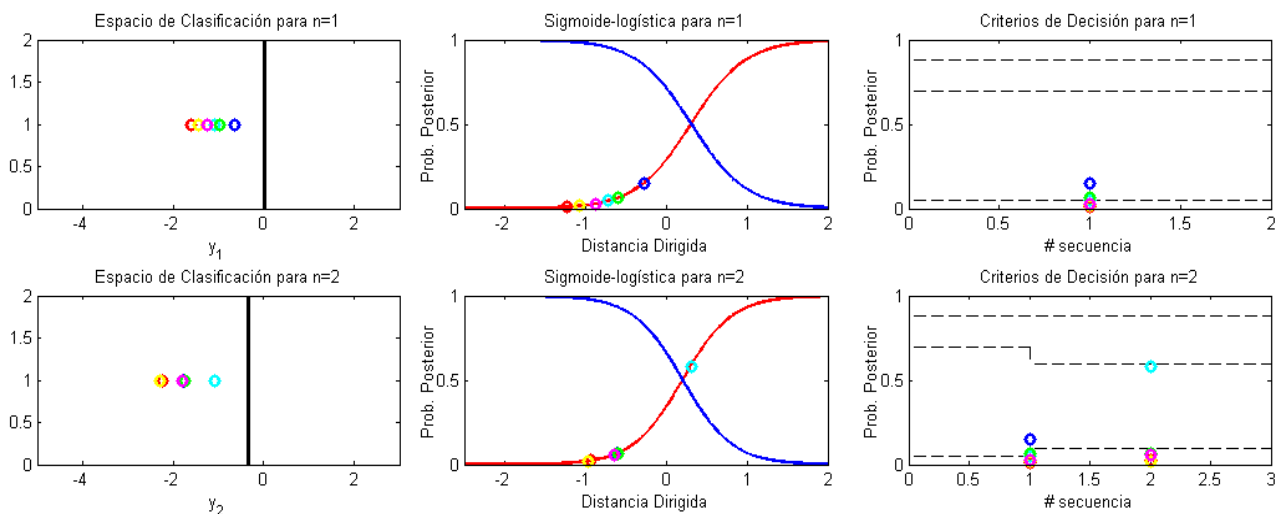


Figura 4.6. . Proceso de selección para el caso de los renglones de la segunda letra del conjunto de prueba. La primera columna presenta, de arriba hacia abajo, el mapeo de los vectores de distancia acumulada a los espacios de clasificación definidos para  $n=1$  y  $2$ . En la segunda columna se muestran las distancias dirigidas con respecto al hiperplano de clasificación correspondiente junto con el  $n$ -ésimo modelo de la sigmoide logística, para  $n=1$  y  $2$ . Los resultados de transformar cada una de las distancias dirigidas, a partir de la sigmoide logística correspondiente, a valores de probabilidad posterior se muestran en las gráficas de la tercera columna junto con los límites de decisión. Al finalizar la segunda secuencia de estimulación el elemento de color cian, correspondiente con el cuarto renglón, es etiquetado con base en el tercer criterio como infrecuente.

En la figura 4.7 se presenta el mismo proceso de selección para el caso de las columnas. Nótese que al finalizar la primera secuencia de estimulación, el elemento azul se localiza muy separado del resto y del lado opuesto del hiperplano de clasificación. Asimismo, presenta un valor de probabilidad posterior mayor que el límite *maxPost*, lo que permite asignarlo a la clase de tipo infrecuente. Sin embargo, puesto que al término de la primera secuencia de estimulación aún no ha sido posible identificar el renglón correspondiente, el sistema continúa estimulando al usuario.

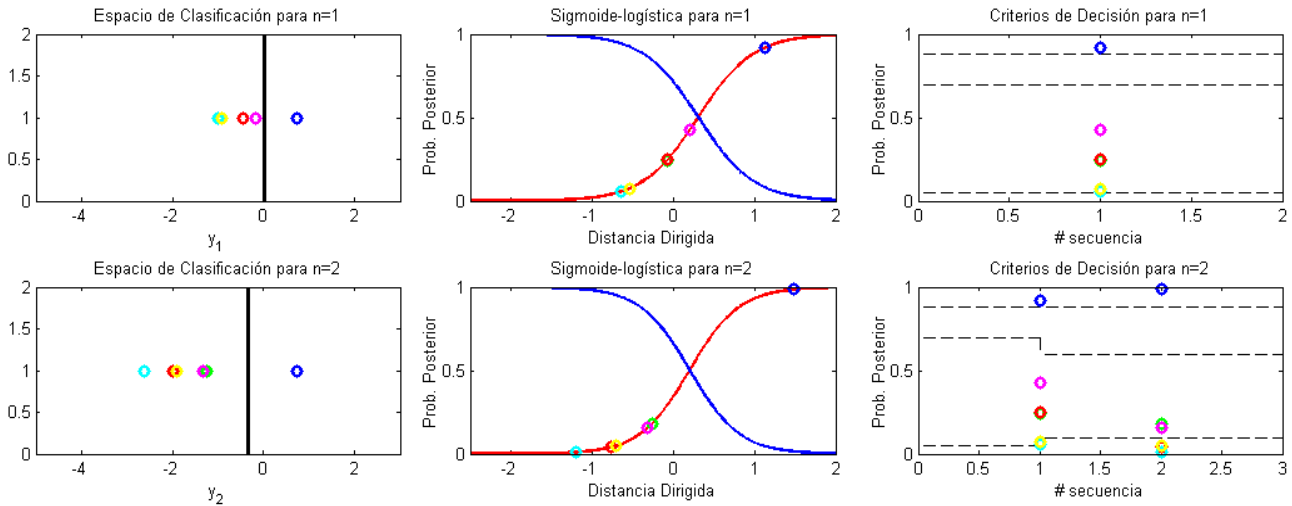


Figura 4.7. Proceso de selección para el caso de las columnas de la segunda letra del conjunto de prueba. De izquierda a derecha se muestra el mapeo de los vectores de distancia acumulada al  $n$ -ésimo espacio de clasificación; las distancias dirigidas con respecto al hiperplano correspondiente junto con el  $n$ -ésimo modelo de la sigmoide-logística y el mapeo de las distancias a valores de probabilidad posterior. Al finalizar la primera serie de estimulación el elemento azul correspondiente con la primera columna es identificado como infrecuente, no obstante, debido a que el renglón es seleccionado hasta el término de la segunda secuencia, es necesario continuar con el proceso.

Al finalizar la segunda secuencia de estimulación el elemento azul, correspondiente con la primera columna de la matriz de caracteres (fig. 4.1), presenta aún más evidencia de pertenecer a la clase infrecuente y, como simultáneamente se ha seleccionado el cuarto renglón, es posible identificar a partir de las coordenadas (4,1) el símbolo que desea elegir el usuario; el cual consiste en la letra “S” que, como puede verificarse a partir de la tabla 3.3, equivale al segundo elemento del conjunto de prueba. De manera que se ha conseguido clasificar correctamente el caracter elegido por el usuario a partir de la información generada en las dos primeras secuencias de estimulación.

Finalmente se inicia el análisis para determinar el tercer elemento del conjunto de prueba; las proyecciones de los vectores de distancia acumulada,  $\hat{y}_n$ , para  $n=1,2,\dots,N$ , correspondientes con cada uno de los doce objetivos se muestran en la figura 4.8.

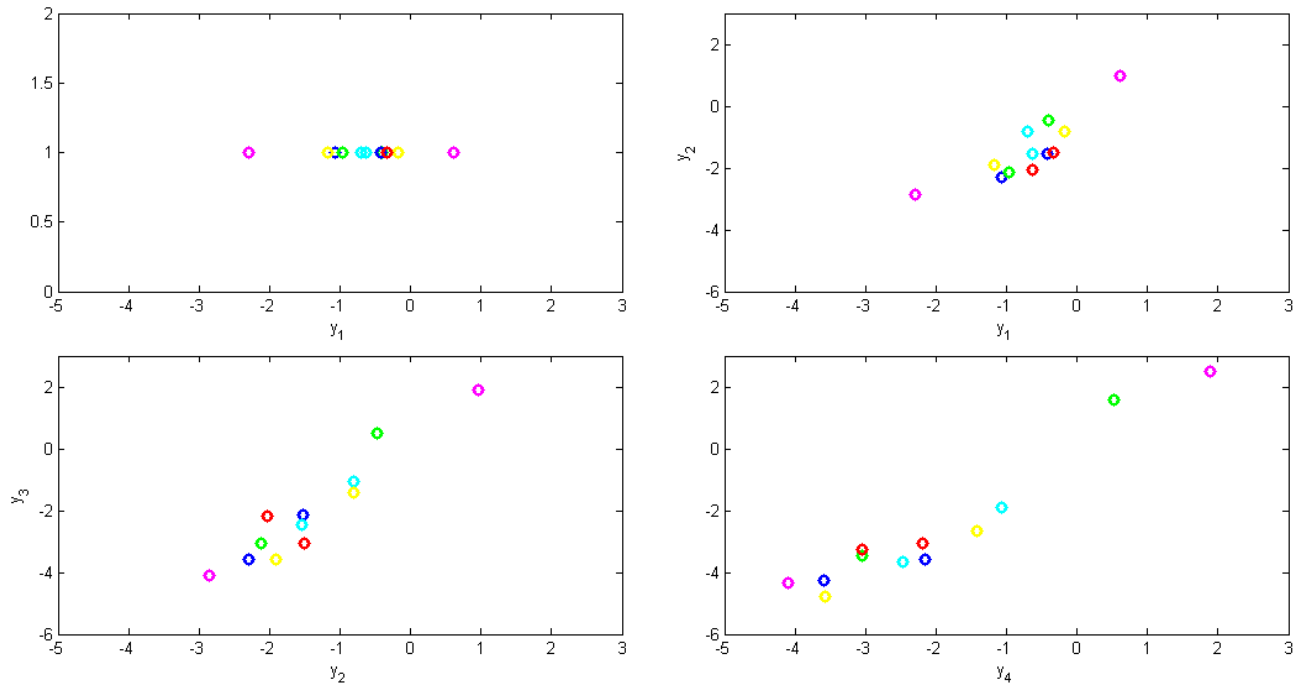


Figura 4.8. Proyecciones de los vectores de distancia acumulada generados a partir de las respuestas asociadas con la tercera letra del conjunto de prueba.

En la figura 4.9 se presenta el proceso de selección del objetivo infrecuente para el caso de los renglones; nótese a partir de la tercera columna de la figura 4.9 que al finalizar la primera secuencia de estimulación ninguno de los criterios se cumple, razón por la cual el sistema debe continuar estimulando al usuario. Al término de la segunda secuencia, los elementos consiguen diferenciarse mejor no obstante, los criterios siguen sin cumplirse y es necesario presentar otra secuencia de estimulación. Después de la tercera secuencia el elemento de color verde cumple con el primer criterio puesto que presenta un valor mayor a  $maxPost$  (0.88); esto significa que el renglón correspondiente presenta una probabilidad mayor al 88% de haber provocado respuestas de tipo infrecuente y, por lo tanto, es etiquetado como tal. A partir de la figura 4.1 se observa que el color verde representa el segundo renglón de manera que, la letra elegida por el usuario, debe encontrarse ubicada en la segunda línea de la matriz de caracteres.



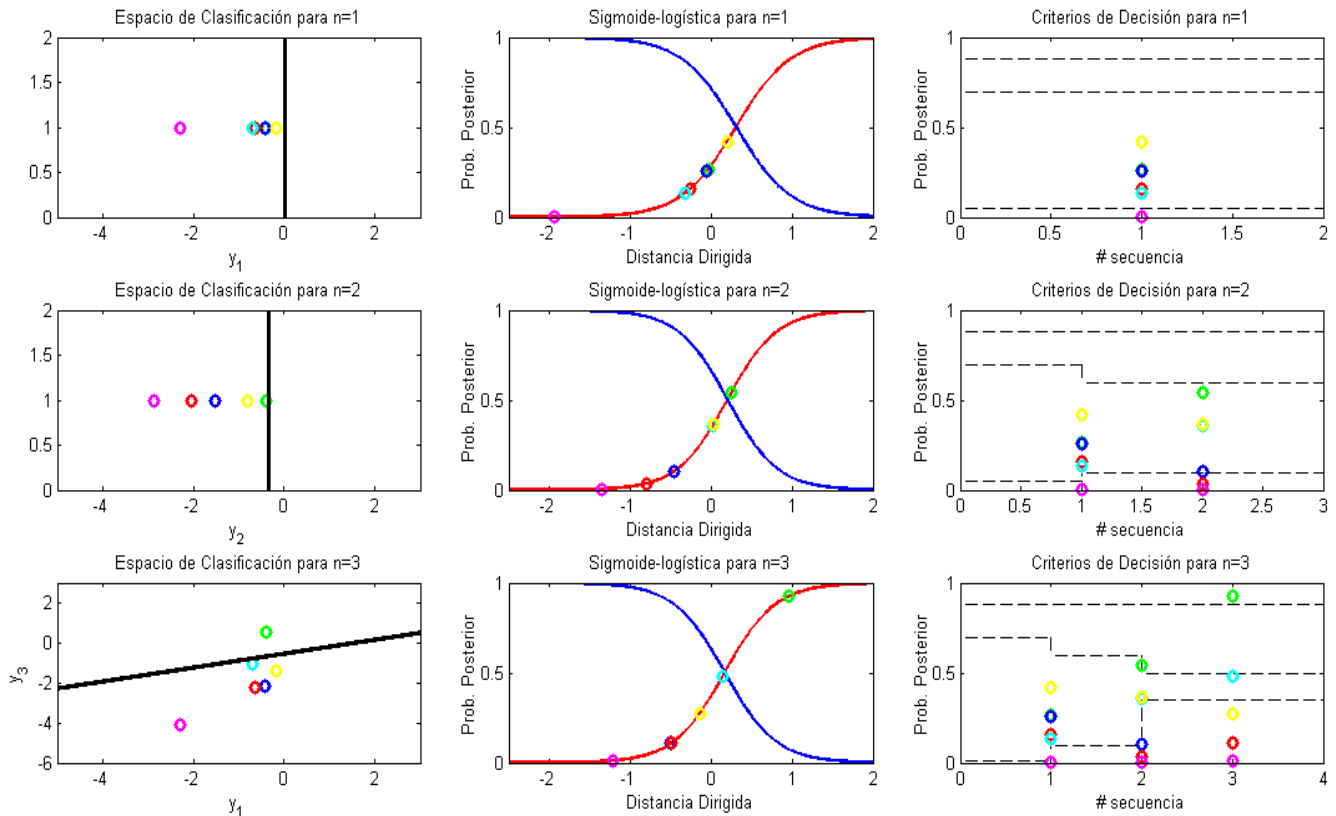


Figura 4.9. Proceso de selección para el caso de los renglones de la tercera letra del conjunto de prueba. La primera columna presenta, de arriba hacia abajo, el mapeo de los vectores de distancia acumulada a los espacios de clasificación definidos para  $n=1$ ; 2 y 3. En la segunda columna se muestran las distancias dirigidas con respecto al hiperplano de clasificación correspondiente junto con el  $n$ -ésimo modelo de la sigmoide logística, para  $n=1$ ; 2 y 3. Los resultados de transformar cada una de las distancias dirigidas, a partir de la sigmoide logística correspondiente, a valores de probabilidad posterior se muestran en las gráficas de la tercera columna junto con los límites de decisión; al finalizar la tercera serie de estímulos el elemento de color verde, correspondiente con el segundo renglón, es etiquetado con base en el primer criterio como infrecuente.

El mismo proceso de selección se muestra en la figura 4.10 para el caso de las columnas. Obsérvese a partir de la tercera columna de gráficas que desde la primera secuencia de estimulación el sistema podría etiquetar al elemento morado como infrecuente puesto que, al tener éste un valor de  $p(C_{\text{inf}}|\hat{y}_n)$  mayor que  $\text{maxPost}$  (0.88) cumple con el primer criterio. No obstante, debido a que al finalizar la primera secuencia de estimulación aún no se ha conseguido identificar el renglón correspondiente, el sistema continúa estimulando al usuario.

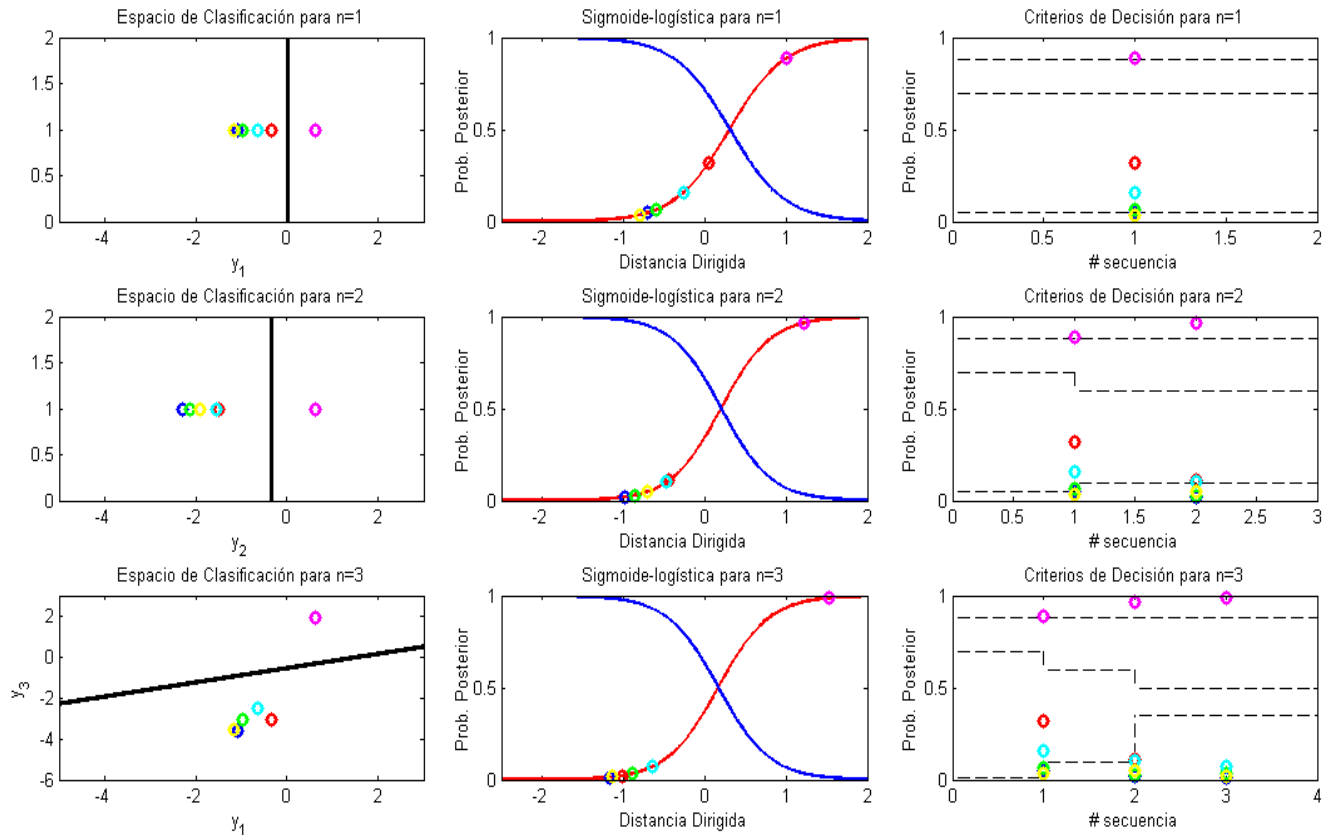


Figura 4.10. Proceso de selección para el caso de las columnas de la tercera letra del conjunto de prueba. De izquierda a derecha se muestra el mapeo de los vectores de distancia acumulada al  $n$ -ésimo espacio de clasificación; las distancias dirigidas con respecto al hiperplano correspondiente junto con el  $n$ -ésimo modelo de la sigmoide-logística y el mapeo de las distancias a valores de probabilidad posterior. Al finalizar la primera serie de estimulación el elemento morado correspondiente con la sexta columna es identificado como infrecuente, no obstante, debido a que el renglón es seleccionado hasta el término de la tercera serie, es necesario continuar con el proceso.

Al finalizar la segunda secuencia de estimulación el elemento morado se ha separado aún más del resto de los objetivos y, al presentar un valor mayor que  $maxPost$  sigue cumpliendo con el primer criterio; sin embargo, al finalizar la segunda secuencia de estimulación todavía no se ha conseguido identificar el renglón correspondiente, por lo que no es posible seleccionar el carácter y es necesario mantener el proceso. Después de la tercera secuencia de estimulación la evidencia que muestra el elemento morado es aún mayor y, debido a que presenta un valor de  $p(C_{inf}|\hat{y}_n)$  mayor a  $maxPost$  (0.88), de nuevo cumple con el primer criterio. Al mismo tiempo el primer renglón de la matriz de caracteres es etiquetado como infrecuente, por lo que es posible detener el proceso de estimulación y seleccionar la sexta columna, la cual es representada por el elemento morado. El carácter situado en la posición indicada por el segundo renglón y la sexta columna de la matriz de símbolos es la letra “L” que, como puede observarse a partir de la

tabla 3.3, corresponde con el tercer elemento del conjunto de prueba. Esto significa que al término de la tercera secuencia, el sistema ha conseguido tomar una decisión correcta y puede detener el proceso de estimulación.

## 4.2. Validación del Sistema

Al finalizar la clasificación de los elementos del conjunto de prueba se obtiene una precisión del 100%, puesto que los tres elementos que lo integran fueron identificados correctamente. Por otro lado, para seleccionarlos fue necesario presentar seis secuencias de estimulación, uno para el primer carácter, dos para el segundo y tres para el último. De manera que, en promedio, el número de letras identificadas por minuto equivale<sup>4</sup> a 7.06 letras/min o, visto desde otra perspectiva, se requieren en promedio de 2 secuencias de estimulación para identificar cada símbolo. En contraste, un sistema convencional configurado para presentar cuatro secuencias de estimulación para la selección de cada uno de los caracteres, es capaz de identificar 4.62 letras/min; lo que representa aproximadamente el 65% de la rapidez alcanzada por la metodología propuesta.

Estos resultados junto con los obtenidos para cada una de las nueve pruebas restantes de la validación cruzada se muestran en la tabla 3.9. Asimismo, con el fin de comparar el desempeño del sistema con el método convencional, se incluyen los resultados generados con la aplicación P300\_Speller® a partir de los mismos datos a través de un sistema configurado para presentar cuatro secuencias de estimulación.

La desviación estándar de los valores promedio generados a partir de la prueba de validación cruzada para los indicadores de desempeño del sistema propuesto refleja poca dispersión de los resultados; con lo que se confirma la independencia entre la clasificación y la partición de los datos a partir de la cual se obtienen los conjuntos de entrenamiento y prueba del sistema. La precisión se mantuvo en 90% con el método propuesto y en 100% con el convencional; no obstante en cuanto a la rapidez de los resultados, se observa que la metodología propuesta mejora en más de un 35% puesto que, a diferencia del sistema convencional que consigue deletrear de manera constante 4.62 caracteres por minuto, bajo el enfoque descrito se obtiene un resultado promedio de 7.2 letras/min. El aumento en la tasa de deletreo está sujeto a la disminución en el número de secuencias de

---

<sup>4</sup> El tiempo que tarda el sistema en identificar cada elemento se calcula a partir de la siguiente expresión:

$$t_{\text{deletreo}} = \text{pau}_{\text{ini}} + (\text{ser})(\#\text{obj})(\text{dur}_{\text{Est}} + \text{pau}_{\text{Est}}).$$

Donde  $\text{pau}_{\text{ini}}$  representa la pausa inicial (4 s);  $\text{ser}$  es el número de series de estimulación empleadas en el proceso de selección;  $\#\text{obj}$  es el número de objetivos de la matriz de caracteres (12, 6 renglones y 6 columnas) y  $\text{dur}_{\text{Est}}$  y  $\text{pau}_{\text{Est}}$  corresponden, respectivamente, con la duración de los estímulos (0.625 s) y con la pausa entre estímulos (1.25 s).

estimulación empleadas para seleccionar cada caracter, el cual en promedio fue de únicamente 2 sec/letra; en contraste con el sistema convencional que, en este caso fue configurado para presentar 4 sec/letra.

Vía	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letrac/min	sec/letra	Precisión (%)	letras/min	sec/letra
1	100	8.57	1.33	100	4.62	4
2	100	7.06	2	100	4.62	4
3	100	8.57	1.33	100	4.62	4
4	100	5.58	4	100	4.62	4
5	100	7.06	2	100	4.62	4
6	66.67	7.74	1.67	100	4.62	4
7	100	7.06	2	100	4.62	4
8	33.33	7.74	1.67	100	4.62	4
9	100	6.49	2.33	100	4.62	4
10	100	6	2.67	100	4.62	4
<b>Promedio</b>	<b>90 ± 22.5</b>	<b>7.2 ± 1</b>	<b>2 ± 0.54</b>	<b>100 ± 0</b>	<b>4.62 ± 0</b>	<b>4 ± 0</b>

Tabla 4.1. Resultados del análisis de validación cruzada de 10 vías.

### 4.3. Resultados de Todos los Usuarios

En la tabla 4.2 se presentan los resultados promedio, junto con la desviación estándar correspondiente, generados a partir de los datos de los catorce usuarios para sistemas con un valor de  $\hat{N} = 2; 3; 4; 5; 6$  y  $8$ . Asimismo, con el fin de comparar el desempeño con el alcanzado por un sistema convencional, se muestran los resultados obtenidos con la aplicación P300\_Speller® para los correspondientes valores de  $N=2; 3; 4; 5; 6$  y  $8$ . Los mismos resultados pueden observarse gráficamente a partir de la figura 4.11.

Validación, todos los usuarios						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	75.34 ± 19.69	7.25 ± 0.28	1.92 ± 0.11	74.58 ± 19.76	7.06 ± 0	2 ± 0
3	87.7 ± 12.37	6.59 ± 0.4	2.38 ± 0.23	86.51 ± 14.26	5.58 ± 0	3 ± 0
4	94.49 ± 6.2	6.55 ± 0.58	2.48 ± 0.39	94.49 ± 6.2	4.62 ± 0	4 ± 0
5	94.95 ± 6.54	6.1 ± 0.63	2.86 ± 0.52	92.55 ± 6.29	3.93 ± 0	5 ± 0
6	98.7 ± 3.3	6.14 ± 0.78	2.77 ± 0.63	96.47 ± 6.12	3.43 ± 0	6 ± 0
8	99.29 ± 2.67	6.04 ± 0.93	2.99 ± 0.76	96.68 ± 7.04	2.73 ± 0	8 ± 0

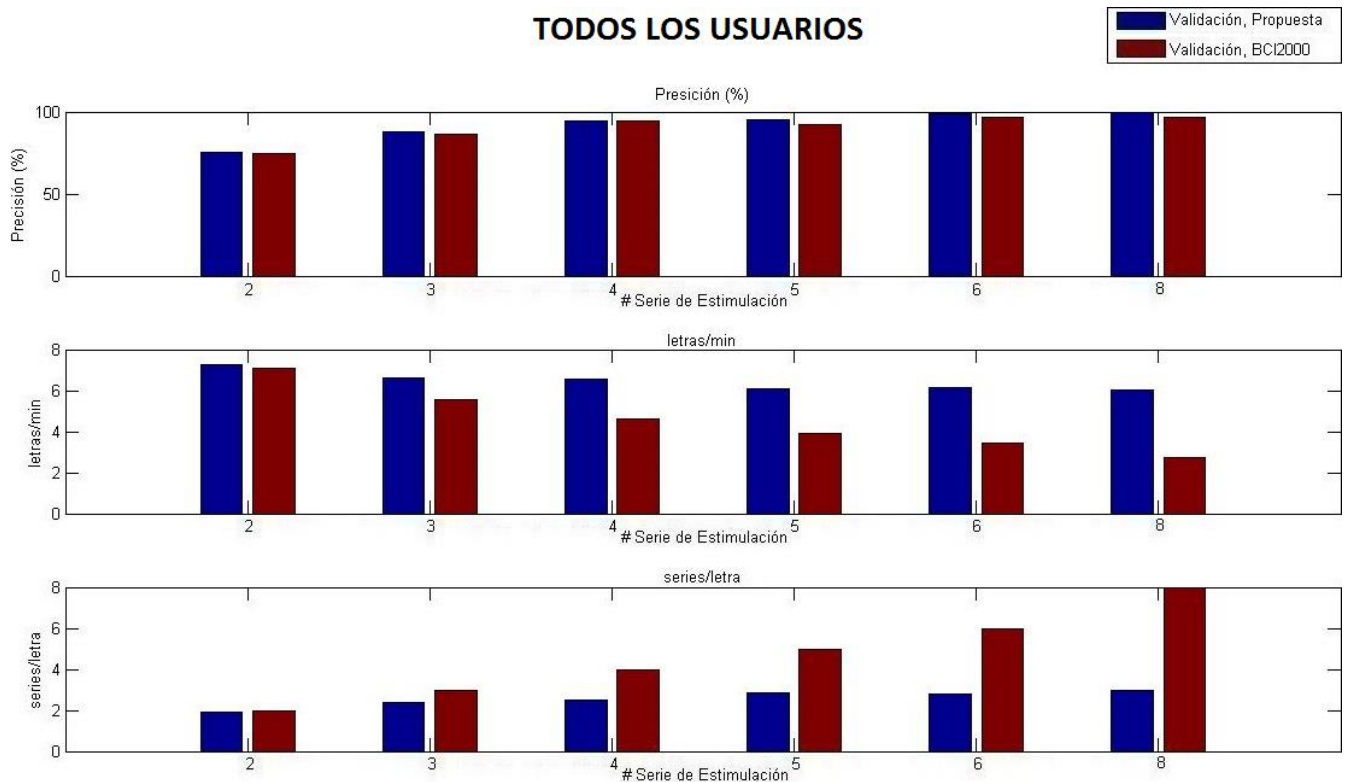


Figura 4.1. Resultados promedio de los 14 usuarios.

#### 4.4. Discusión

El hecho de ajustar el número de series de estimulación y emplear sólo una o dos para seleccionar aquellos objetivos que muestran mucha evidencia de pertenecer a la clase infrecuente, permite ahorrar tiempo e invertirlo en aquellos que son más difíciles de etiquetar. Por esta razón el sistema propuesto presenta, en promedio, resultados mucho más rápidos que el sistema convencional.

A partir de los valores de desviación estándar obtenidos durante las pruebas de validación cruzada, los cuales indican poca dispersión de los resultados, se puede concluir que el desempeño del método se mantiene sin importar la partición de los datos con los que se forman los conjuntos de entrenamiento y prueba del sistema.

Un fenómeno interesante que se observa a partir de los resultados alcanzados por el método convencional, es que no siempre mejora su precisión al incrementar el número de series de estimulación con las que se seleccionan los caracteres. Esto sucede cuando la promediación coherente no consigue resaltar los patrones de interés debido a que las respuestas que se suman presentan una peor razón señal a ruido.

## Capítulo 5

### 5. Conclusiones

Establecer el número necesario de series de estimulación de manera dinámica a partir de la evidencia que muestran las respuestas de pertenecer a la clase infrecuente, representa un esquema viable para controlar este tipo de aplicaciones. Los resultados obtenidos demuestran que bajo este enfoque de clasificación se consigue, en comparación con otros métodos convencionales, disminuir considerablemente el tiempo promedio de deletreo sin descuidar la precisión.

A partir de los resultados obtenidos se observa que el sistema desarrollado bajo la nueva propuesta alcanza, para todos los valores de  $\hat{N} = N$ , resultados ligeramente más precisos que el método convencional; no obstante, en cuanto a la rapidez del deletreo, se consigue una notable mejora. Para el caso de  $\hat{N} = N = 2; 3; 4; 5; 6$  y  $8$ , el tiempo promedio de selección de los caracteres disminuyó, respectivamente, en un 2.62%; 15.33%; 29.47%; 35.57%; 44.14% y 54.8%.

Con el sistema convencional se alcanza una buena precisión (arriba del 90%) a partir de la cuarta serie de estimulación ( $N=4$ ); lo mismo sucede con el método propuesto (para el correspondiente valor de  $\hat{N} = 4$ ), sólo que el tiempo que le lleva seleccionar cada carácter disminuye, en promedio, alrededor del 30%. Incluso con un valor de  $\hat{N} = 8$ , el sistema propuesto es casi 25% más rápido que el método convencional para un valor de  $N=4$ , con la ventaja de que resulta aproximadamente 5% más preciso.

Este nuevo esquema presenta dos cambios fundamentales en comparación con otros métodos que le permiten ajustar de forma automática el número de series de estimulación: por un lado, para cada conjunto de vectores de distancia acumulada generado al término de cada serie de estimulación, se instrumentó un clasificador con el fin de diferenciar las repuestas de tipo infrecuente de las tipo frecuente; y por otra parte, se construyeron los modelos de las funciones sigmoide logística que permiten mapear los resultados correspondientes a valores de probabilidad posterior. A partir de estas dos

modificaciones se puede analizar, al término de cada serie de estimulación, la evidencia que presentan los objetivos de pertenecer a la clase infrecuente y, con base en ello, tomar una decisión que determine si es necesario seguir estimulando al usuario o, si se ha conseguido identificar el símbolo de su elección y, por lo tanto, es posible detener el proceso. En contraste, los métodos convencionales utilizan el mismo número de series de estimulación para seleccionar cada carácter, sin importar la evidencia que muestran las diferentes respuestas. De manera que el operador del sistema se encuentra ante la disyuntiva de elegir pocas series de estimulación y, aun a costa de comprometer la precisión de los resultados, obtener una buena tasa de deletreo; o, presentar un mayor número de series y evitar errores a pesar de la lentitud que esto conlleva considerando además que, gran parte de las respuestas generadas, representan información redundante. Este último párrafo exhibe el compromiso que existe entre el tiempo que le lleva a un sistema de este tipo tomar una decisión y la precisión correspondiente; mismo que, bajo el esquema propuesto, es eliminado al conseguir que con base en las características de los datos, se ajuste el número de series de estimulación y se presenten tantas como sean necesarias.

Finalmente, resulta muy interesante observar la tendencia que presentan los parámetros de desempeño conforme aumenta el número de series de estimulación; en el caso del tiempo de deletreo se consigue un valor que se estabiliza alrededor de 6 letras/min y, en el caso del número de series para seleccionar cada carácter, un valor de 3 series/letra; esto demuestra que existe cierta consistencia en las características de las respuestas y que, el sistema, consigue ajustarse satisfactoriamente para evitar la adquisición de información redundante.

## Referencias

- [1]. Wolpaw J., Birbaumer N., McFarland D., Pfurtscheller G. & Vaughan T. *Brain-Computer Interfaces for Communication and Control*, Clinical Neurophysiology, 113, 767-791, 2002.
- [2]. Birbaumer N. *Breaking the Silence: Brain-Computer Interfaces (BCI) for Communication and Motor Control*, Psychophysiology, 43, 517-532, 2006.
- [3]. Mason S. G., Bashashati A., Fatourechi M., Navarro K.F. and Birch G.E. *A Comprehensive Survey of Brain Interface Technology designs*, Annals of Biomedical Engineering, 35(2):137-169, 2007.
- [4]. Gert Pfurtscheller, Reinhold Scherer and Chista Neuper. *EEG-Based Brain-Computer Interface*, pp. 316-317.
- [5]. Birbaumer N., Wolpaw, J., *BCI2000: A General-Purpose Brain-computer Interface System*, IEEE Trans. Biomed. Eng.,51:1034-1043, 2004.
- [6]. Chiappa, Silvia. *Analysis and Classification of EEG Signals using Probabilistic Models for Brain Computer Interfaces*. Tesis para la Obtención del grado de Doctora en Ciencias. Escuela Politécnica Federal de Lausana, Suiza (2006).
- [7]. Curran E.A., Stokes M. J. B. *Learning to Control Brain Activity: a Review of the Production and Control of EEG Components for Driving Brain-Computer Interface (BCI) Systems*. Brain and Cognition, pages 326-336, 2003.
- [8]. Baztarrica Ochoa, Jorge: *EEG Signal Classification for Brain Computer Interface Applications*. Escuela Politécnica Federal de Lausana, Suiza (2002).
- [9]. George L. and Lécuyer A. *An overview of Research on "passive" Brain-Computer Interfaces for Implicit Human-Computer Interaction*, In International Conference on Applied Bionics and Biomechanics (ICABB), 2010.
- [10]. Pfurtscheller G. & Neuper C. *Motor Imagery and Direct Brain-Computer Communication*, proceedings of the IEEE, 89(7):1123-1134, 2001
- [11]. S. Lin, Y. Tsai y C. Liou,: *Conscious Mental Tasks and their EEG Signals*. (1993).
- [12]. Raja Parasuvaman, Matthew Rizzo. *Neuroergonomics, The Brain at Work*. Oxford University Press. New York, 2007.
- [13]. Sutter E. *The Brain Response Interface: Communication through Visually Induced Electrical Brain Responses*. Journal of Microcomputer Applications, (1992).
- [14]. Vialatte F.B., Dauwels J. and Cichoki A. *Steady-State Visually Evoked Potentials: Focus on Essential Paradigms and Future Perspectives*, Progress in Neurobiology, 2010, 90:418-438.



- [15]. M. Middendorf, G. McMillan, G. Calhoun, K. S. Jones: *Brain-computer Interfaces Based on Steady-state Visual Evoked Response*. IEEE Transactions on Rehabilitation Engineering. (2000).
- [16]. Elshout J. and García Molina G, *Review of Brain-Computer Interfaces Based on the P300 Evoked Potential*, Technical Report PR-TN 2009/00066 Koninklijke Philips Electronics, 2009.
- [17]. Fazel-Rezai R., Allison B., Guger C., Sellers E.W., Klein S.C. and Kübler A. *P300 Brain Computer Interface: Current Challenges and Emerging Trends*, Frontiers in Neuroengineering, vol 5, 2012.
- [18]. Guger C., Daban S., Sellers E.W., Holzner C., Krausz G., Carabalona R., Gramatica F. and Edlinger G. *How many People are Able to Control a P300-Based Brain Computer Interface (BCI)?*, Neuroscience Letters 462:94-98, 2009.
- [19]. Lee T.W., Yu Y.W., Wu H.C., Chen T.J. *Do resting Brain Dynamics Predict Oddball evoked-potential*, BMC Neurosci, 24:12:121, 2011.
- [20]. Serby H., Yom-Tov E. and Inbar G.F. *An Improved P300-Based Brain-Computer Interface*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13(1):89-98, 2005.
- [21]. L. A. Farwell, E. Donchin: *Talking off the Top of the Head: toward a Mental Prosthesis Utilizing Event-related Brain Potentials*. Electroencephalography and Clinical Neurophysiology. (1988).
- [22]. Donchin E. *The Mental Prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface*, IEEE Transactions on Rehabilitation Engineering, 8(2):174-179, 2000.
- [23]. Rompelman O, Ros HH., *Coherent averaging technique: a tutorial review. Part 1: Noise reduction and the equivalent filter*. J Biomed Eng, 1986.
- [24]. Lotte F., Congedo M., Lécuyer A., Lamarche F. and Arnaldi B. *A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces*, Journal of Neural Engineering, 4, R1-R13, 2007.
- [25]. Muller K. R., Anderson C.W., and Birch G.E. *Linear and Nonlinear Methods for Brain-Computer Interface*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11(2):165-169, 2003.
- [26]. C. M. Bishop: *Pattern Recognition and Machine Learning*. Springer. (2006)
- [27]. Draper, N. R., and H. Smith. *Applied Regression Analysis*. Hoboken, NJ: Wiley-Interscience, 1998. pp. 307–312
- [28]. Platt J. *Probabilistic outputs for support vector machines and comparison to regularized likelihood methods*, in *Advances in Large Margin Classifiers*, A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds. Cambridge, MA: MIT Press, 2000.
- [29]. Lin H.T., Lin C.J., Weng R.C. *A Note on Platt's Probabilistic Outputs for Support Vector Machines*, Machine Learning, Volume 68, No. 3, 2007.
- [30]. *Cross Validation*
- [31]. Ledesma C., Bojorges E.R., Gentiletti G., Bougrain L., Saavedra C., Yañez O, *P300-Speller Public-Domain Database*, <http://akimpech.izt.uam.mx/p300db>
- [32]. Schalk G., McFarland D., Hinterberger T., Birbaumer N., Wolpaw, J. *BCI2000: A General-Purpose Brain-computer Interface System*, IEEE Trans. Biomed. Eng., 51:1034-1043, 2004.
- [33]. Jing J., Brendan Z., Sellers E., Clemens B., Petar H., Xingyu W. and Christa N. *An adaptive P300-based control system*, Neural Eng. 8, 2011.

## Anexo 1

### Resultados de cada Usuario

Los resultados obtenidos a partir de los registros de cada usuario durante las pruebas de desempeño del sistema se muestran en las siguientes secciones. Las pruebas realizadas corresponden con las descritas en el capítulo III y consisten en la evaluación de la metodología a partir de un conjunto de datos no vistos, así como de un análisis de validación cruzada para demostrar la independencia entre los resultados y la partición con la que se generan los conjuntos de entrenamiento y prueba del sistema. Con el fin de comparar el desempeño alcanzado con la metodología propuesta y con el método convencional, se muestran también los resultados generados a partir la aplicación P300\_Speller® sobre los mismos conjuntos de datos.

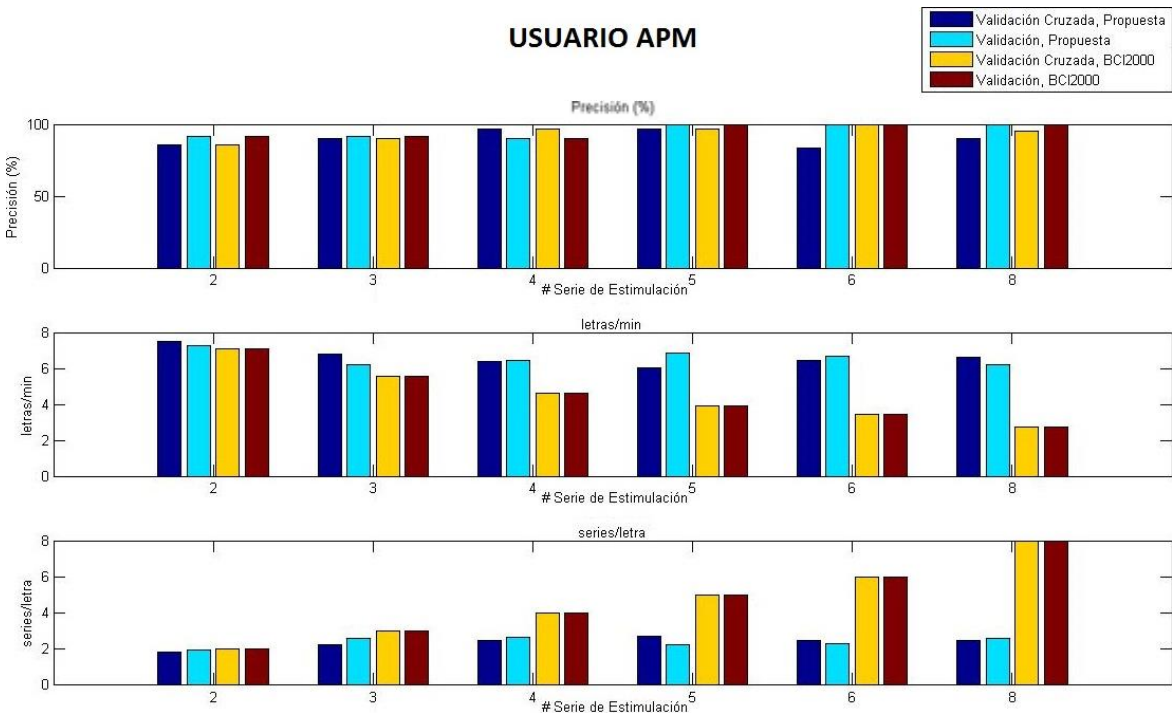
En la primera tabla de cada sección, se presenta la lista de palabras deletreadas por el usuario junto con el correspondiente número de series de estimulación emitidas. Las siguientes dos tablas muestran, respectivamente, los resultados de las pruebas de validación cruzada y la evaluación de los sistema para valores de  $N = \hat{N} = 2; 3; 4; 5; 6$  y  $8$ , estos resultados se muestran en una gráfica de barras al final de cada sección.

#### 1. Usuario APM

Usuario APM	
Palabra	# Series
CARINO	15
SUSHI	15
SUSHI	15
COMIDA	15
COCINA	15
CARPA	15
GALLO	7
NEGRO	5
LIBRO	4
APPLE	3
TECLA	3
FOCA	2
HOGAR	2

Pruebas de Validación Cruzada, Usuario APM						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	86 ± 13.5	7.51 ± 0.31	1.78 ± 0.15	86 ± 13.5	7.06 ± 0	2 ± 0
3	90 ± 17.48	6.8 ± 0.63	2.18 ± 0.36	90 ± 17.48	5.58 ± 0	3 ± 0
4	96.67 ± 10.54	6.37 ± 0.54	2.43 ± 0.35	96.67 ± 10.54	4.62 ± 0	4 ± 0
5	96.67 ± 10.54	6.03 ± 0.71	2.7 ± 0.55	96.67 ± 10.54	3.93 ± 0	5 ± 0
6	83.33 ± 23.57	6.44 ± 0.86	2.43 ± 0.59	100 ± 0	3.43 ± 0	6 ± 0
8	90 ± 21.08	6.61 ± 1.43	2.45 ± 1.04	95 ± 15.81	2.73 ± 0	8 ± 0

Validación, Usuario APM						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	91.67	7.27	1.92	91.67	7.06	2
3	91.67	6.2	2.58	91.67	5.58	3
4	90	6.43	2.6	90	4.62	4
5	100	6.85	2.22	100	3.93	5
6	100	6.69	2.25	100	3.43	6
8	100	6.22	2.57	100	2.73	8

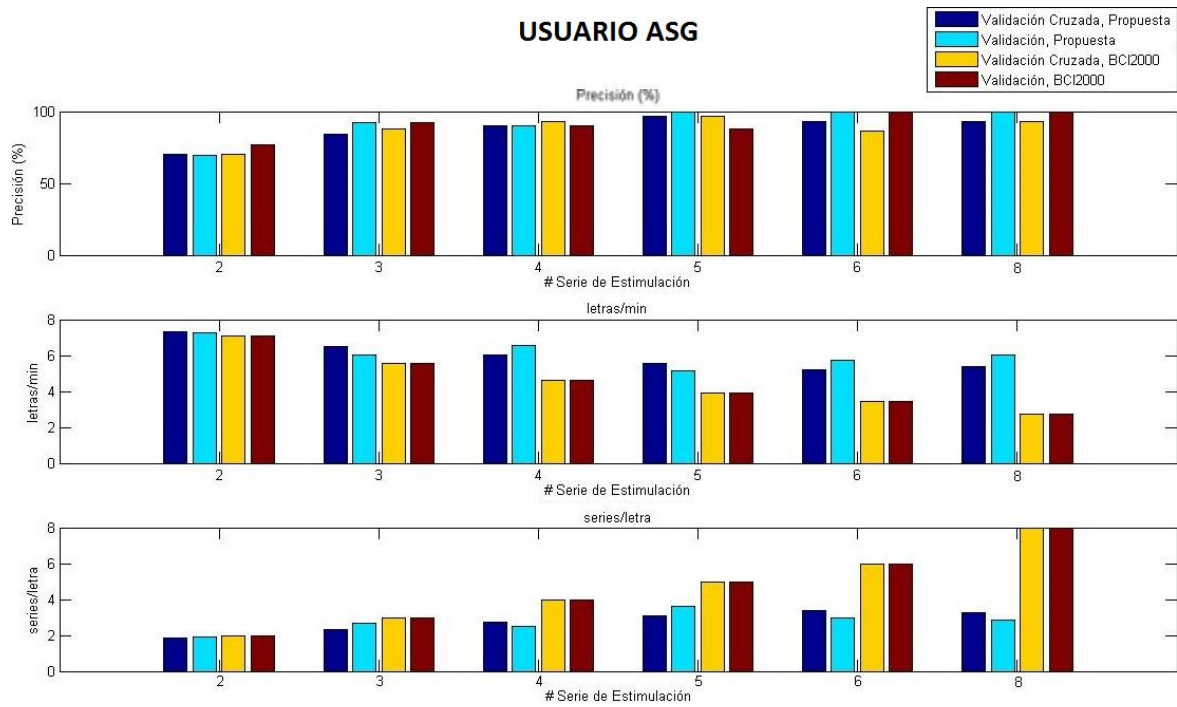


## 2. Usuario ASG

Usuario ASG	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
CENAR	15
COLOR	15
DULCES	15
NUBES	6
ESQUEMAS	3
ALEGRIA	3
GALLETA	4

Pruebas de Validación Cruzada, Usuario ASG						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	70 ± 19.44	7.34 ± 0.28	1.86 ± 0.14	70 ± 19.44	7.06 ± 0	2 ± 0
3	84 ± 20.66	6.51 ± 0.49	2.34 ± 0.3	88 ± 16.87	5.58 ± 0	3 ± 0
4	90 ± 22.5	6.05 ± 1.03	2.73 ± 0.7	93.33 ± 21.08	4.62 ± 0	4 ± 0
5	96.67 ± 10.54	5.57 ± 0.61	3.07 ± 0.54	96.67 ± 10.54	3.93 ± 0	5 ± 0
6	93.33 ± 14.06	5.19 ± 0.48	3.4 ± 0.49	86.67 ± 17.21	3.43 ± 0	6 ± 0
8	93.33 ± 14.06	5.37 ± 0.67	3.27 ± 0.68	93.33 ± 14.06	2.73 ± 0	8 ± 0

Validación, Usuario ASG						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	69.23	7.25	1.92	76.92	7.06	2
3	92.31	6.04	2.69	92.31	5.58	3
4	90	6.53	2.5	90	4.62	4
5	100	5.15	3.63	87.5	3.93	5
6	100	5.71	3	100	3.43	6
8	100	6.05	2.86	100	2.73	8

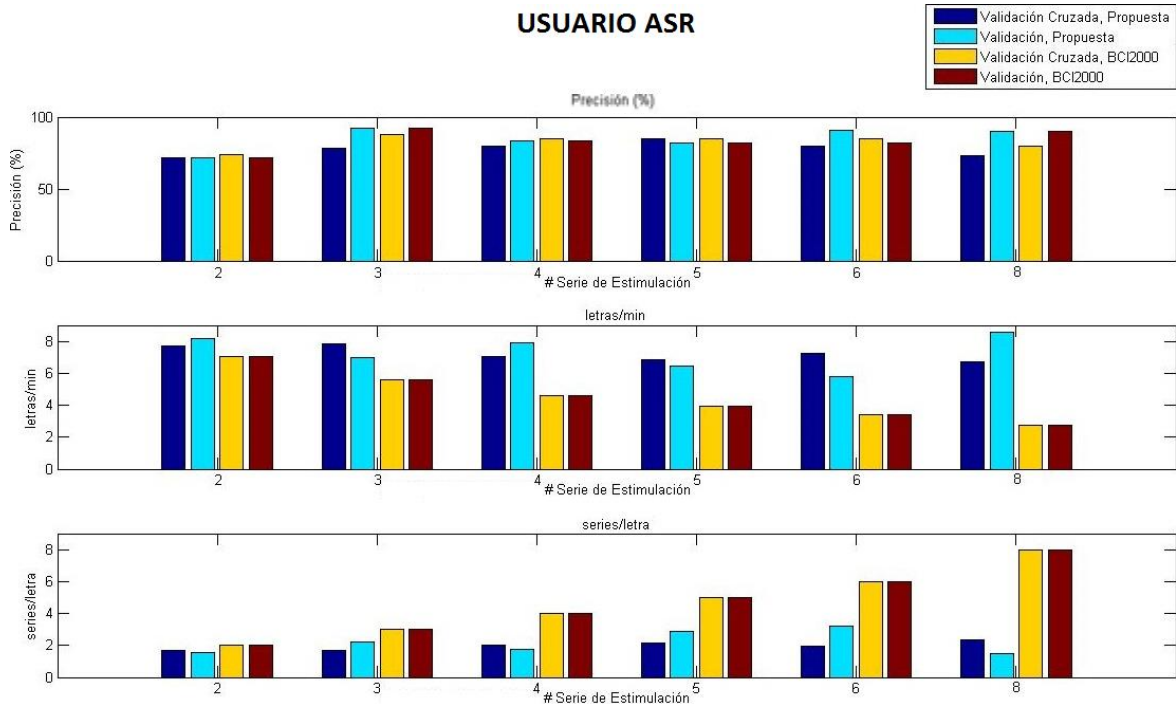


### 3. Usuario ASR

Usuario ASR	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
SUENO	15
NACHO_	15
THAT_IS_OK	15
PERROS	8
LIBRO	6
MOUSE	4
COMIDA	3
LEDEZMA	2

Pruebas de Validación Cruzada, Usuario ASR						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	72 ± 21.5	7.73 ± 23.19	1.68 ± 0.19	74 ± 23.19	7.06 ± 0	2 ± 0
3	78 ± 19.89	7.83 ± 0.79	1.66 ± 0.37	88 ± 10.33	5.58 ± 0	3 ± 0
4	80 ± 22.97	7.06 ± 0.85	2.05 ± 0.47	85 ± 24.15	4.62 ± 0	4 ± 0
5	85 ± 12.91	6.88 ± 0.84	2.15 ± .49	85 ± 12.91	3.93 ± 0	5 ± 0
6	80 ± 0.72	7.23 ± 0.95	1.98 ± 0.55	85 ± 17.48	3.43 ± 0	6 ± 0
8	73.33 ± 30.63	6.75 ± 1.45	2.33 ± 0.85	80 ± 28.11	2.73 ± 0	8 ± 0

Validación, Usuario ASR						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	71.43	8.15	1.57	71.43	7.06	2
3	92.31	6.96	2.23	92.31	5.58	3
4	83.33	7.9	1.75	83.33	4.62	4
5	81.82	6.46	2.91	81.82	3.93	5
6	90.91	5.77	3.18	81.82	3.43	6
8	90	8.59	1.5	90	2.73	8

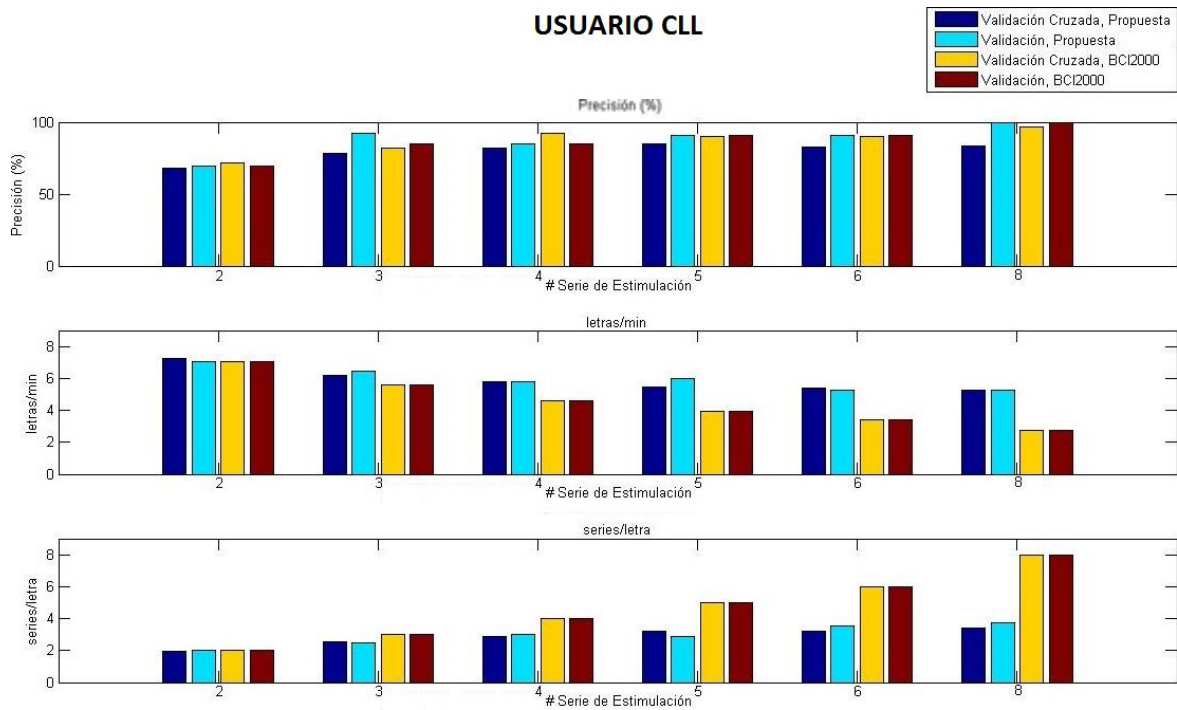


## 4. Usuario CLL

Usuario CLL	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
NARANJA	15
CUARENTENA	15
ROSA	15
RELOJ	6
PANTALLA	6
MADERA	4
ARBOL	4

Pruebas de Validación Cruzada, Usuario CLL						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	68 ± 16.87	7.22 ± 0.2	1.92 ± 0.1	72 ± 21.5	7.06 ± 0	2 ± 0
3	78 ± 23.94	6.17 ± 0.39	2.56 ± 0.28	82 ± 19.89	5.58 ± 0	3 ± 0
4	82 ± 14.76	5.78 ± 0.42	2.86 ± 0.35	92 ± 10.33	4.62 ± 0	4 ± 0
5	85 ± 21.08	5.46 ± 0.7	3.18 ± 0.57	90 ± 12.91	3.93 ± 0	5 ± 0
6	82.5 ± 23.72	5.39 ± 0.6	3.23 ± 0.55	90 ± 12.91	3.43 ± 0	6 ± 0
8	83.33 ± 23.57	5.25 ± 0.79	3.4 ± 0.75	96.67 ± 10.54	2.73 ± 0	8 ± 0

Validación, Usuario CLL						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	69.23	7.06	2	69.23	7.06	2
3	92.31	6.46	2.46	84.62	5.58	3
4	84.62	5.79	3.01	84.62	4.62	4
5	90.91	5.97	2.91	90.91	3.93	5
6	90.91	5.23	3.55	90.91	3.43	6
8	100	5.27	3.75	100	2.73	8



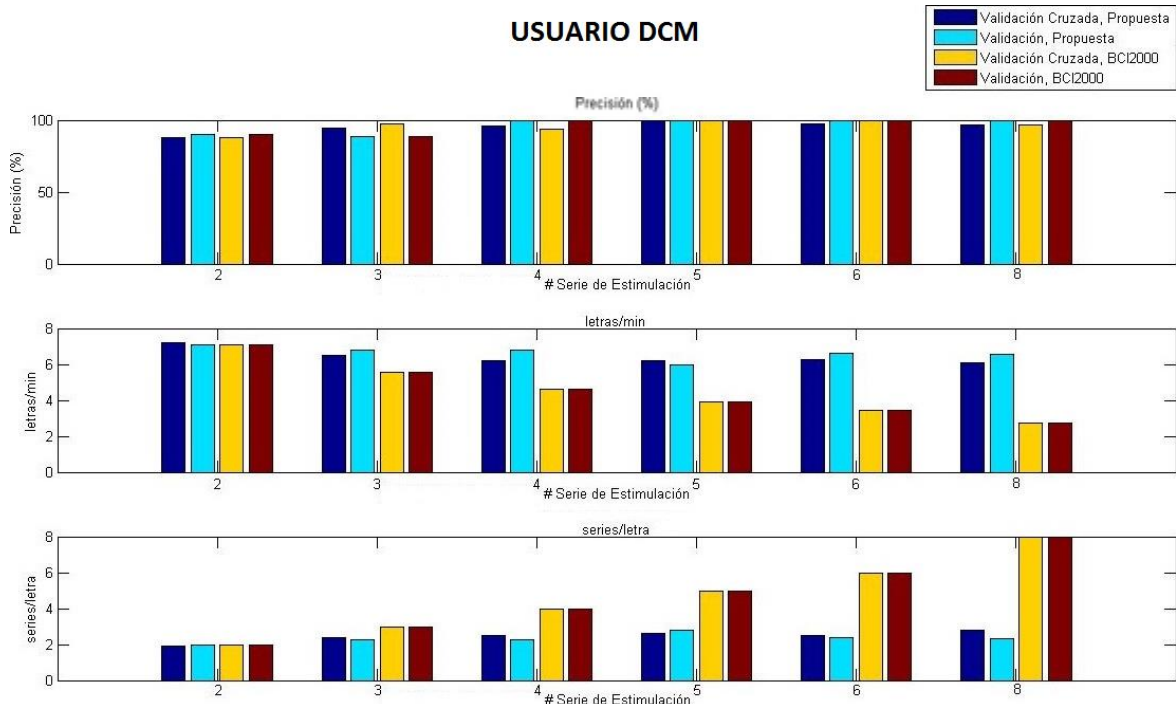
### 5. Usuario DCM

Usuario DCM	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
SO_BORDE_TU	15
CONEJITO	15
HIERRO	15
EL_PALACIO	7
IMAC_G5PRO	5
UAM_LERMA	4
XANUM_TITLANI	3
TENGO_HABRE	2



Pruebas de Validación Cruzada, Usuario DCM						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	87.5 ± 10.21	7.21 ± 0.17	1.93 ± 0.09	87.5 ± 10.21	7.06 ± 0	2 ± 0
3	94.29 ± 7.38	6.48 ± 0.44	2.36 ± 0.29	97.14 ± 6.02	5.58 ± 0	3 ± 0
4	96 ± 8.43	6.23 ± 0.39	2.52 ± 0.27	94 ± 9.66	4.62 ± 0	4 ± 0
5	100 ± 0	6.22 ± 0.95	2.6 ± 0.67	100 ± 0	3.93 ± 0	5 ± 0
6	97.5 ± 7.91	6.28 ± 0.55	2.5 ± 0.37	100 ± 0	3.43 ± 0	6 ± 0
8	96.67 ± 10.54	6.08 ± 1.14	2.77 ± 0.96	96.67 ± 10.54	2.73 ± 0	8 ± 0

Validación, Usuario DCM						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	90	7.06	2	90	7.06	2
3	88.88	6.82	2.28	88.88	5.58	3
4	100	6.77	2.27	100	4.62	4
5	100	6	2.77	100	3.93	5
6	100	6.62	2.36	100	3.43	6
8	100	6.57	2.33	100	2.73	8



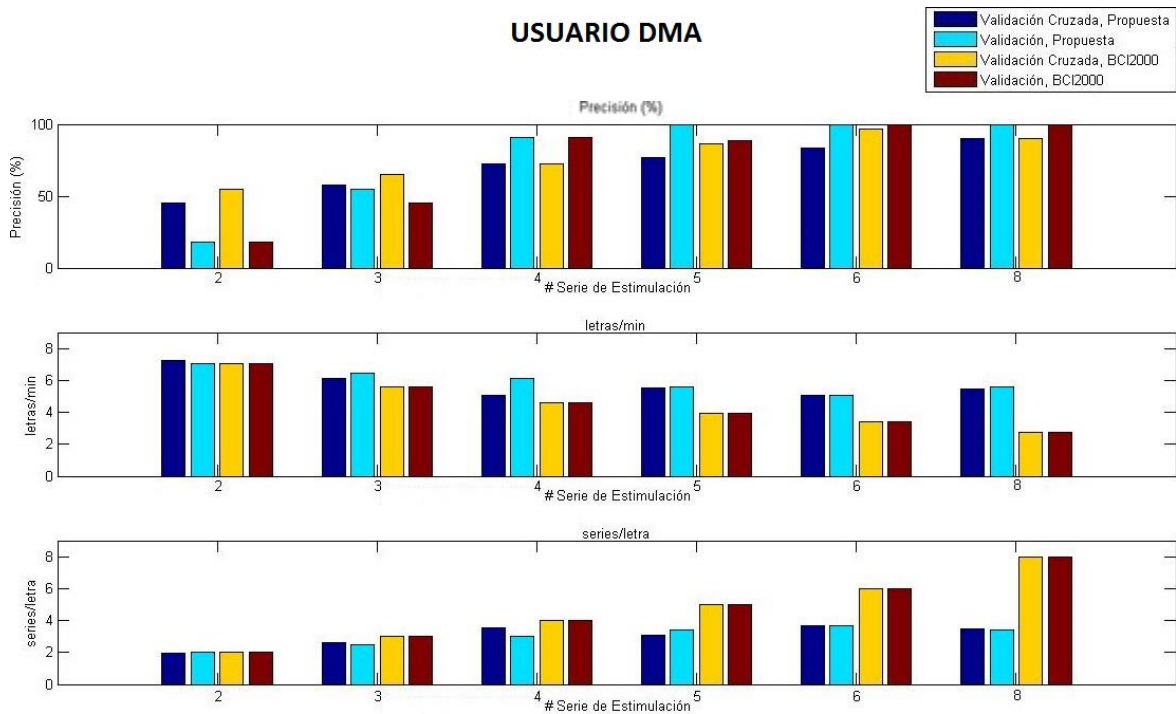
## 6. Usuario DMA

Usuario DMA	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
BCI	15
CUBO	15
PIXEL	15
WAKERLY	7
NOTICI	7
SILENCIO	4

Pruebas de Validación Cruzada, Usuario DMA						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	45 ± 19.72	7.22 ± 0.36	1.93 ± 0.19	55 ± 19.72	7.06 ± 0	2 ± 0
3	57.5 ± 23.72	6.14 ± 0.61	2.6 ± 0.39	65 ± 21.08	5.58 ± 0	3 ± 0
4	72.5 ± 21.89	5.07 ± 0.5	3.53 ± 0.48	72.5 ± 21.89	4.62 ± 0	4 ± 0
5	76.67 ± 27.44	5.55 ± 0.73	3.1 ± 0.61	86.67 ± 17.21	3.93 ± 0	5 ± 0
6	83.33 ± 23.57	5.06 ± 1.07	3.67 ± 0.9	96.67 ± 10.54	3.43 ± 0	6 ± 0
8	90 ± 31.62	5.467 ± 1.58	3.45 ± 1.36	90 ± 21.08	2.73 ± 0	8 ± 0

Validación, Usuario DMA						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	18.18	7.06	2	18.18	7.06	2
3	54.55	6.48	2.46	45.46	5.58	3
4	90.91	6.15	3	90.91	4.62	4
5	100	5.56	3.44	88.89	3.93	5
6	100	5.06	3.67	100	3.43	6
8	100	5.57	3.43	100	2.73	8

### USUARIO DMA



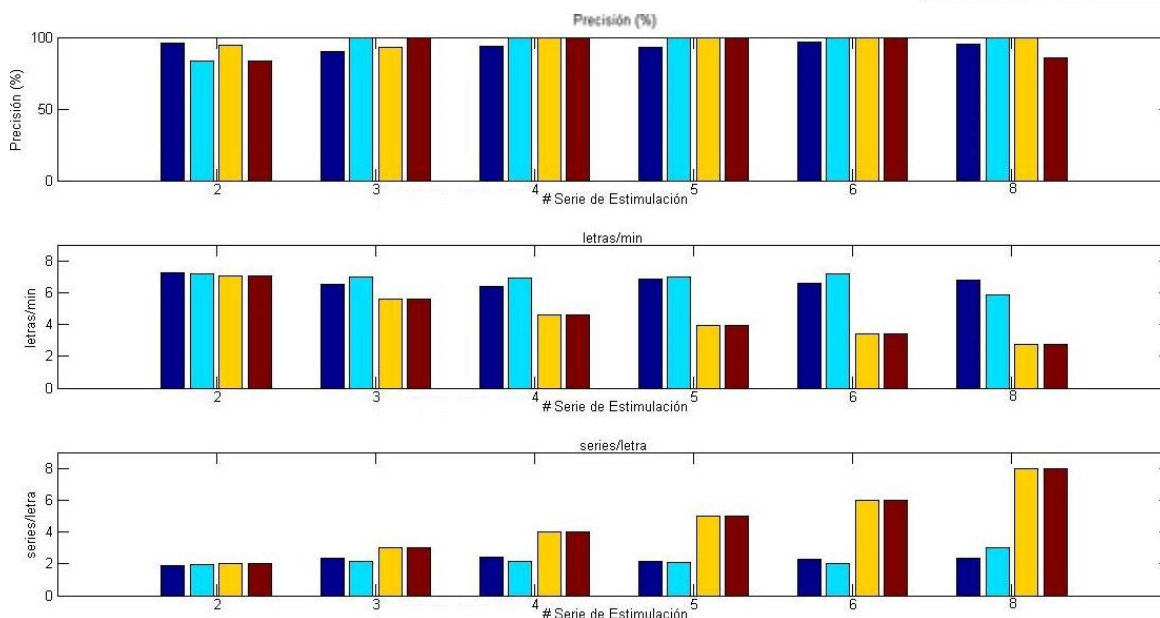
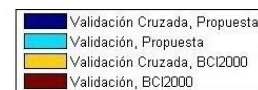
## 7. Usuario ELC

Usuario ELC	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
CASCODO	15
QUOTE	15
PRINCE	15
INSIDE	7
SERVER	7
OPERATOR	4
PRACTICAS	4
WESLEY	3
SANIDAD	3
STARTED	2
DISENAR	2
JUEGO	1
BAMBINO	1

Pruebas de Validación Cruzada, Usuario ELC						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	95.71 ± 6.9	7.23 ± 0.2	1.91 ± 0.1	94.29 ± 9.99	7.06 ± 0	2 ± 0
3	90 ± 14.06	6.51 ± 0.37	2.33 ± 0.24	93.33 ± 11.65	5.58 ± 0	3 ± 0
4	94 ± 9.66	6.39 ± 0.49	2.42 ± 0.33	100 ± 0	4.62 ± 0	4 ± 0
5	93.33 ± 14.06	6.84 ± 0.73	2.17 ± 0.48	100 ± 0	3.93 ± 0	5 ± 0
6	96.67 ± 10.54	6.6 ± 0.64	2.3 ± 0.4	100 ± 0	3.43 ± 0	6 ± 0
8	95 ± 15.81	6.76 ± 1.33	2.35 ± 1.06	100 ± 0	2.73 ± 0	8 ± 0

Validación, Usuario ELC						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	83.33	7.2	1.94	83.33	7.06	2
3	100	7	2.13	100	5.58	3
4	100	6.91	2.15	100	4.62	4
5	100	7.01	2.11	100	3.93	5
6	100	7.18	2	100	3.43	6
8	100	5.84	3	85.71	2.73	8

**USUARIO ELC**



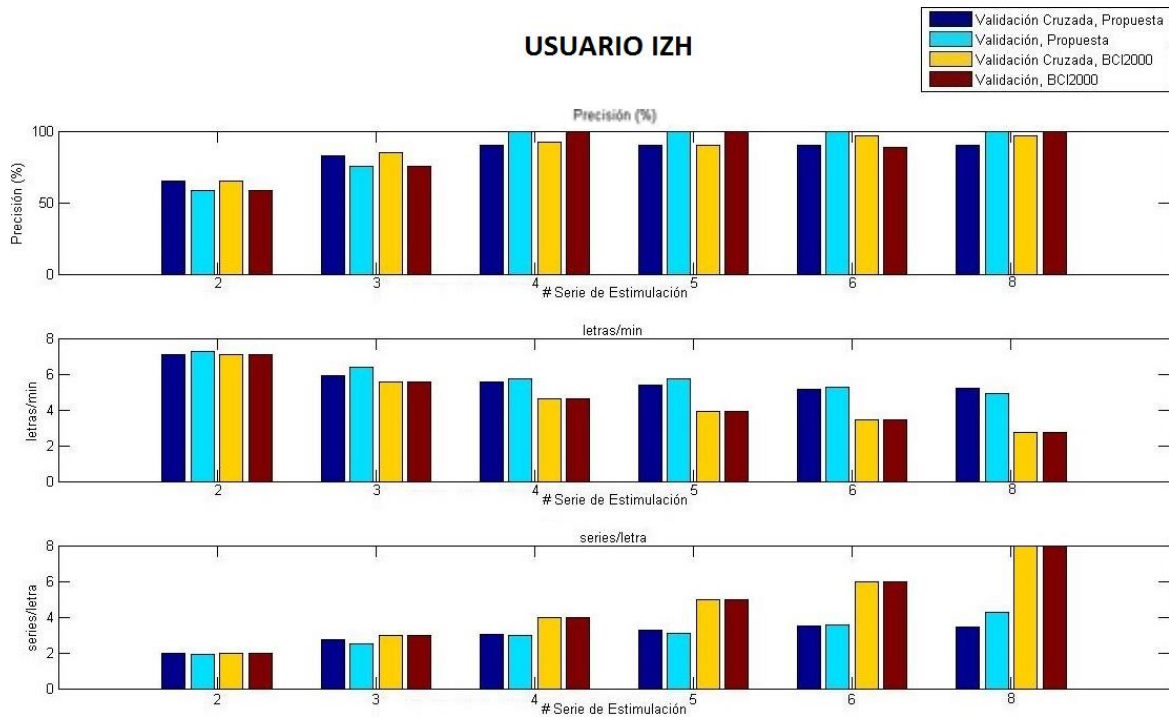
## 8. Usuario IZH

Usuario IZH	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
1987JUN19	15
YOU	15
BIOMEDICA	15
HOLA	7
APPLE	5
APPLE	5
AGUA	3

Pruebas de Validación Cruzada, Usuario IZH						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	65 ± 17.48	7.06 ± 0	2 ± 0	65 ± 17.48	7.06 ± 0	2 ± 0
3	82.5 ± 16.87	5.94 ± 0.34	2.73 ± 0.25	85 ± 17.48	5.58 ± 0	3 ± 0
4	90 ± 12.91	5.55 ± 0.42	3.05 ± 0.35	92.5 ± 12.08	4.62 ± 0	4 ± 0
5	90 ± 12.91	5.39 ± 0.83	3.28 ± 0.74	90 ± 12.91	3.93 ± 0	5 ± 0
6	90 ± 22.5	5.13 ± 0.67	3.5 ± 0.63	96.67 ± 10.54	3.43 ± 0	6 ± 0
8	90 ± 16.1	5.23 ± 0.8	3.43 ± 0.85	96.67 ± 10.54	2.73 ± 0	8 ± 0

Validación, Usuario IZH						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	58.33	7.27	1.92	58.33	7.06	2
3	75	6.41	2.5	75	5.58	3
4	100	5.72	3	100	4.62	4
5	100	5.73	3.09	100	3.93	5
6	100	5.24	3.56	88.89	3.43	6
8	100	4.93	4.25	100	2.73	8

### USUARIO IZH

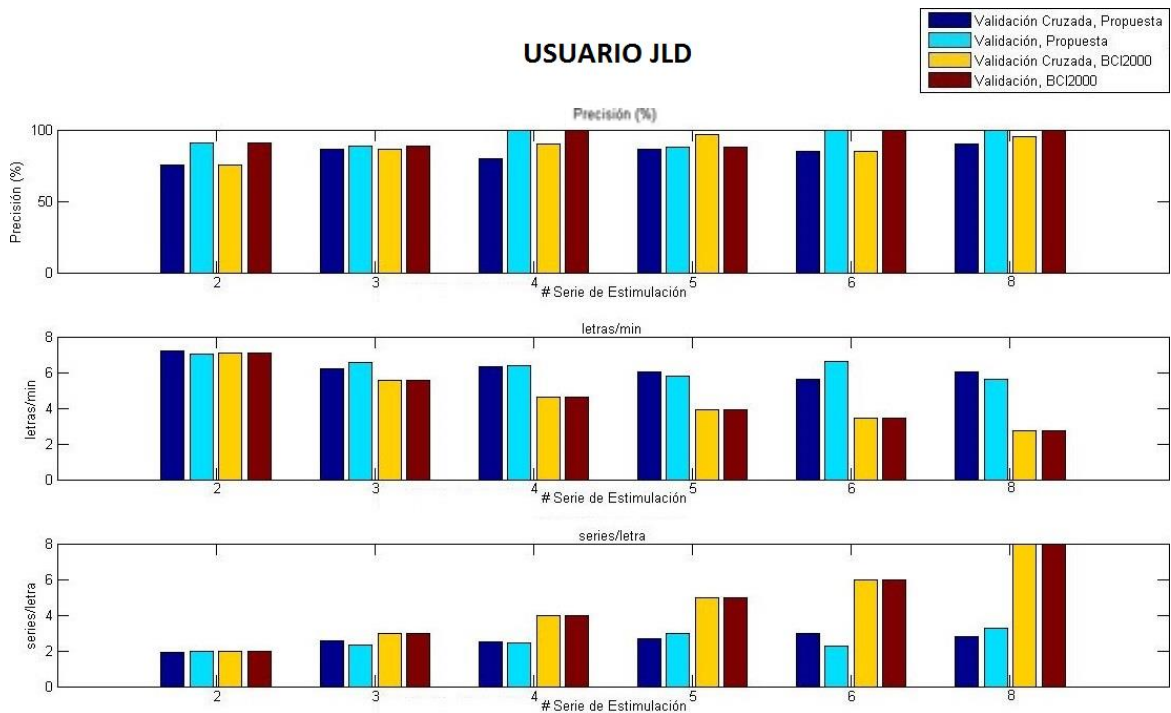


### 9. Usuario JLD

Usuario JLD	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
HOLA	15
FEO	15
PAULA	15
OK	7
AGUA	5
WATER	4
UAM	3
CLAUDIA	2

Pruebas de Validación Cruzada, Usuario JLD						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	75 ± 16.67	7.21 ± 0.24	1.93 ± 0.12	75 ± 16.67	7.06 ± 0	2 ± 0
3	86.67 ± 23.31	6.22 ± 0.49	2.53 ± 0.32	86.67 ± 23.31	5.58 ± 0	3 ± 0
4	80 ± 28.11	6.31 ± 0.72	2.5 ± 0.48	90 ± 16.1	4.62 ± 0	4 ± 0
5	86.67 ± 17.21	6.03 ± 0.75	2.7 ± 0.51	96.67 ± 10.54	3.93 ± 0	5 ± 0
6	85 ± 24.15	5.63 ± 0.56	3 ± 0.47	85 ± 24.15	3.43 ± 0	6 ± 0
8	90 ± 21.08	6.02 ± 1.12	2.8 ± 0.89	95 ± 15.81	2.73 ± 0	8 ± 0

Validación, Usuario JLD						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	90.91	7.05	2	90.91	7.06	2
3	88.89	6.57	2.33	88.89	5.58	3
4	100	6.4	2.44	100	4.62	4
5	87.5	5.81	3	87.5	3.93	5
6	100	6.64	2.29	100	3.43	6
8	100	5.61	3.29	100	2.73	8



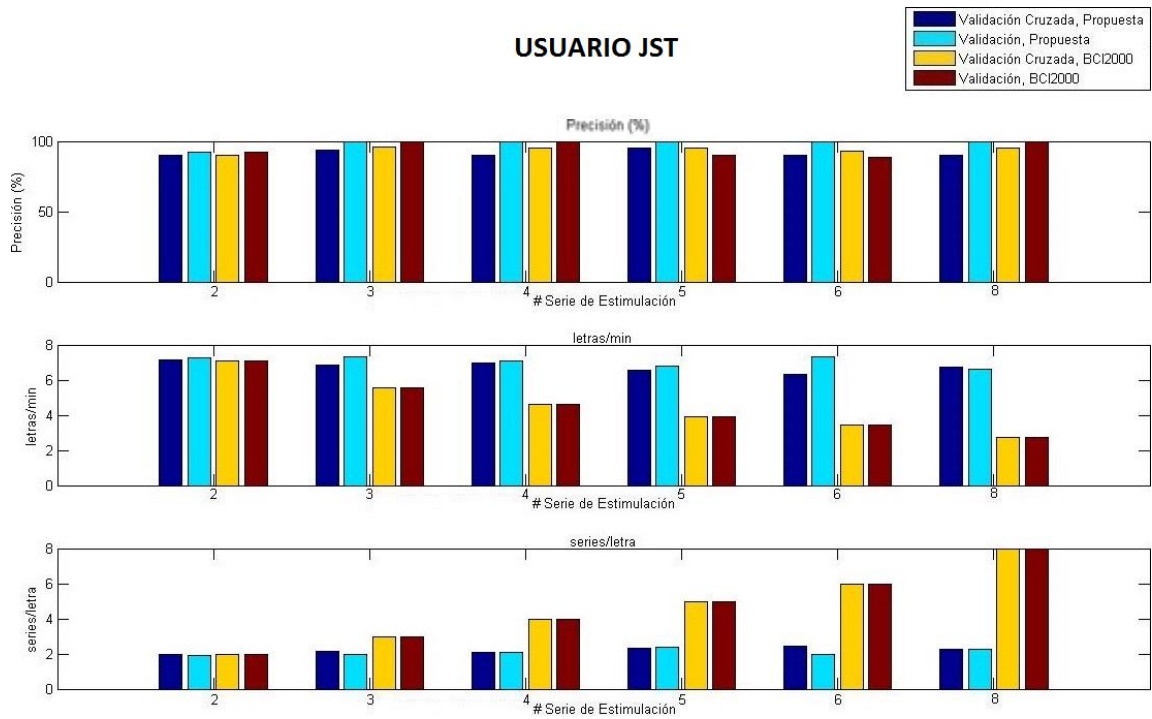
## 10. Usuario JST

Usuario JST	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
ZUKY	15
MAYTE	15
AZUL	15
MANCHA	7
PACHUS	7
NACHA	5
MARRO	4
GONZO	3
TREJO	2

Pruebas de Validación Cruzada, Usuario JST						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	90 ± 14.14	7.14 ± 0.16	1.96 ± 0.08	90 ± 14.14	7.06 ± 0	2 ± 0
3	94 ± 13.5	6.87 ± 0.46	2.12 ± 0.27	96 ± 8.43	5.58 ± 0	3 ± 0
4	90 ± 12.91	6.95 ± 0.43	2.08 ± 0.24	95 ± 10.54	4.62 ± 0	4 ± 0
5	95 ± 10.54	6.54 ± 0.53	2.33 ± 0.33	95 ± 10.54	3.93 ± 0	5 ± 0
6	90 ± 16.1	6.32 ± 0.49	2.47 ± 0.32	93.33 ± 14.05	3.43 ± 0	6 ± 0
8	90 ± 21.08	6.72 ± 0.82	2.25 ± 0.54	95 ± 15.81	2.73 ± 0	8 ± 0

Validación, Usuario JST						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	92.31	7.25	1.92	92.31	7.06	2
3	100	7.33	2	100	5.58	3
4	100	7.07	2.09	100	4.62	4
5	100	6.78	2.4	90	3.93	5
6	100	7.3	2	88.89	3.43	6
8	100	6.64	2.29	100	2.73	8



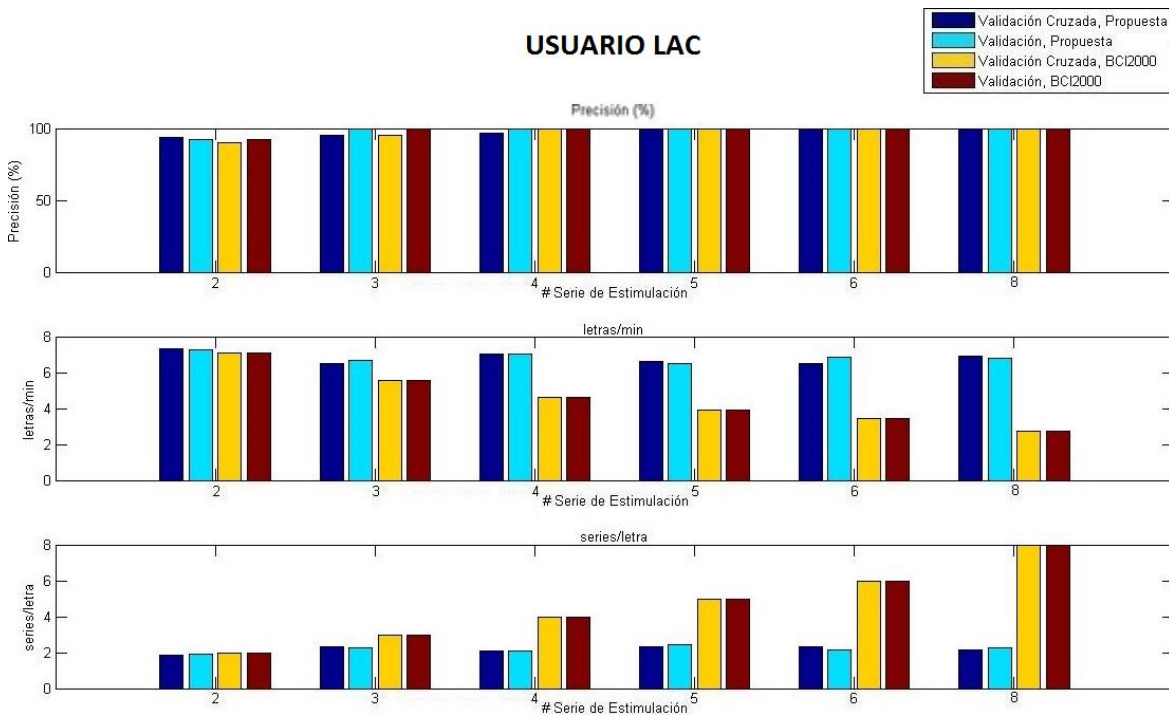


### 11. Usuario LAC

Usuario LAC	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
LAURA	15
DANZA	15
CASA	15
ARBOL	7
PERRO	5
GATO	4
CLAUDIA	3
MIGUEL	2
LIBRO	2
COMPUTADORA	2

Pruebas de Validación Cruzada, Usuario LAC						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	94 ± 13.5	7.34 ± 0.34	1.86 ± .17	90 ± 14.14	7.06 ± 0	2 ± 0
3	95 ± 10.54	6.49 ± 0.48	2.35 ± 0.29	95 ± 10.54	5.58 ± 0	3 ± 0
4	96.67 ± 10.54	7.05 ± 0.95	2.07 ± 0.52	100 ± 0	4.62 ± 0	4 ± 0
5	100 ± 0	6.64 ± 0.84	2.3 ± 0.53	100 ± 0	3.93 ± 0	5 ± 0
6	100 ± 0	6.5 ± 0.25	2.33 ± .16	100 ± 0	3.43 ± 0	6 ± 0
8	100 ± 0	6.9 ± 0.89	2.15 ± 0.53	100 ± 0	2.73 ± 0	8 ± 0

Validación, Usuario LAC						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	92.31	7.25	1.92	92.31	7.06	2
3	100	6.66	2.27	100	5.58	3
4	100	7.02	2.1	100	4.62	4
5	100	6.52	2.44	100	3.93	5
6	100	6.87	2.13	100	3.43	6
8	100	6.79	2.29	100	2.73	8

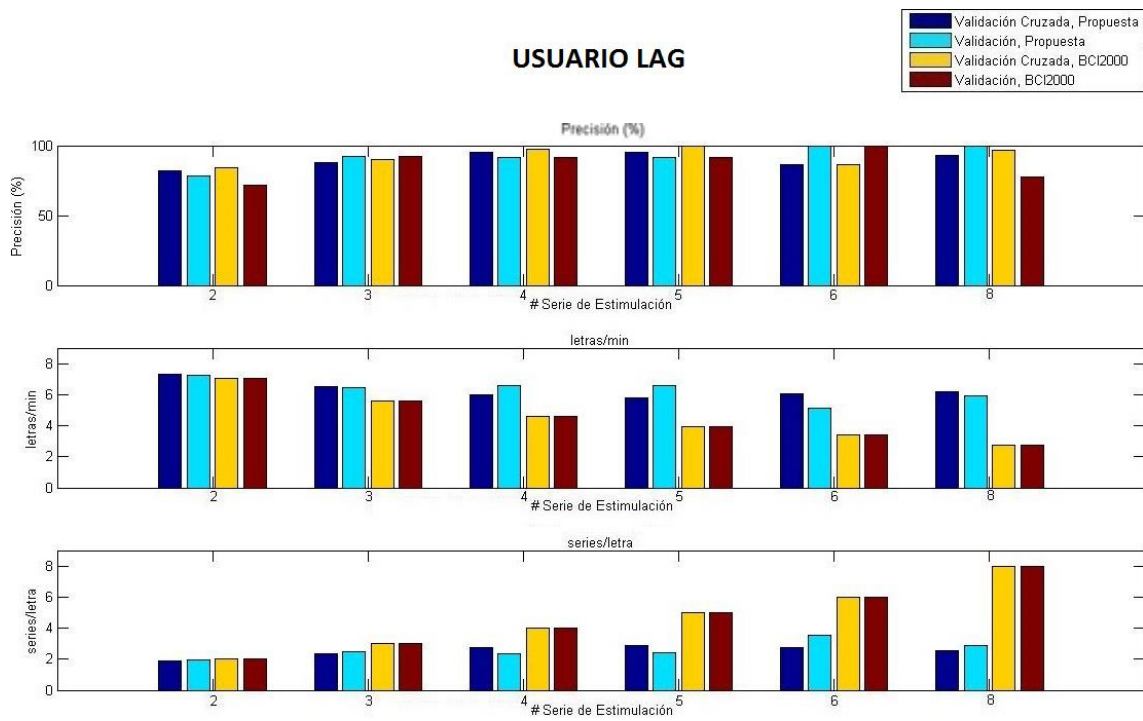


## 12. Usuario LAG

Usuario LAG	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
DORMIR	15
QUIERO_COCA	15
HAMBRE	15
QUIERO_VER_TELE	5
CASA	3
PERRO	2
AUTO	1

Pruebas de Validación Cruzada, Usuario LAG						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	82 ± 22.01	7.3 ± 0.29	1.88 ± 0.14	84 ± 20.66	7.06 ± 0	2 ± 0
3	88 ± 16.87	6.51 ± 0.44	2.34 ± 0.3	90 ± 10.54	5.58 ± 0	3 ± 0
4	95 ± 10.54	5.96 ± 0.48	2.73 ± 0.36	97.5 ± 7.91	4.62 ± 0	4 ± 0
5	95 ± 10.54	5.79 ± 0.67	2.88 ± 0.45	100 ± 0	3.93 ± 0	5 ± 0
6	86.67 ± 17.21	6.04 ± 0.9	2.73 ± 0.7	86.67 ± 17.21	3.43 ± 0	6 ± 0
8	93.33 ± 14.05	6.22 ± 0.74	2.57 ± 0.52	96.67 ± 10.54	2.73 ± 0	8 ± 0

Validación, Usuario LAG						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	78.57	7.24	1.93	71.43	7.06	2
3	92.31	6.46	2.46	92.31	5.58	3
4	91.67	6.57	2.33	91.67	4.62	4
5	91.67	6.58	2.42	91.67	3.93	5
6	100	5.16	3.56	100	3.43	6
8	100	5.95	2.89	77.78	2.73	8

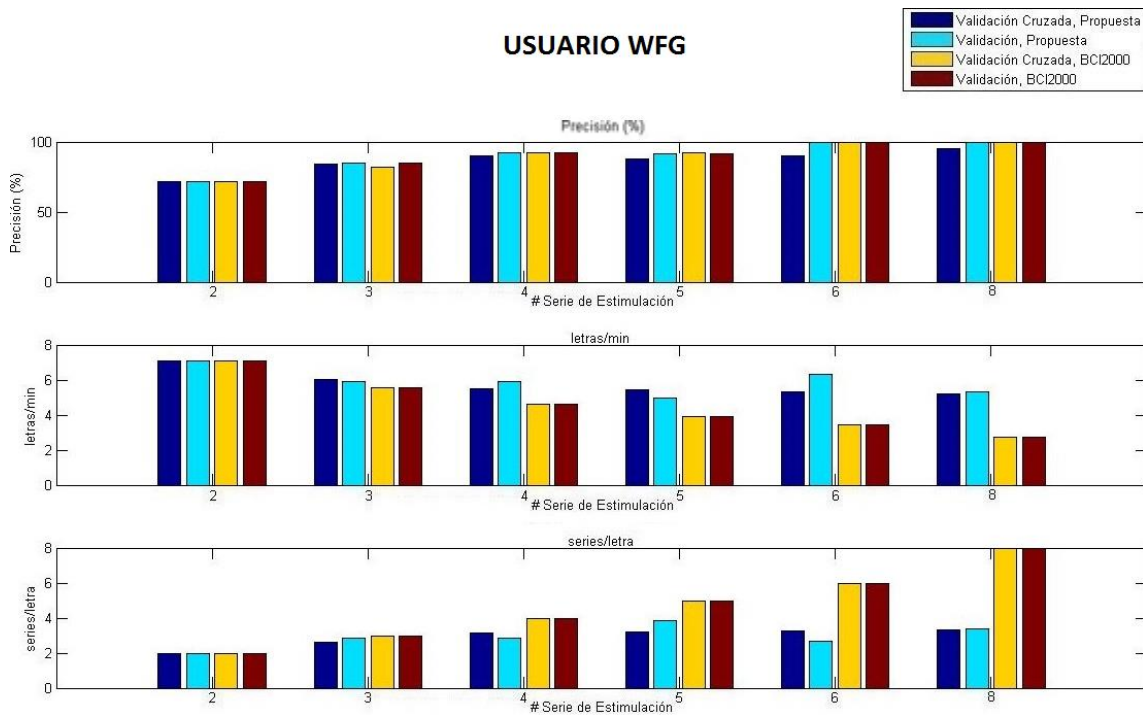


### 13. Usuario WFG

Usuario WFG	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
UAM_I	15
INGENIERIA	15
BIOMEDICA	15
UAM_I	15
INGENIERO	5
SENAL	4
ECG	3
EMG	2
MEDICINA	1

Pruebas de Validación Cruzada, Usuario WFG						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	72 ± 13.98	7.1 ± 0.13	1.98 ± 0.06	72 ± 13.98	7.06 ± 0	2 ± 0
3	84 ± 8.43	6.05 ± 0.35	2.64 ± 0.25	82 ± 11.35	5.58 ± 0	3 ± 0
4	90 ± 14.14	5.49 ± 0.52	3.12 ± 0.45	92 ± 10.33	4.62 ± 0	4 ± 0
5	87.5 ± 17.68	5.43 ± 0.62	3.2 ± 0.62	92.5 ± 12.08	3.93 ± 0	5 ± 0
6	90 ± 12.91	5.31 ± 0.44	3.28 ± 0.43	100 ± 0	3.43 ± 0	6 ± 0
8	95 ± 10.54	5.21 ± 0.18	3.35 ± 0.18	100 ± 0	2.73 ± 0	8 ± 0

Validación, Usuario WFG						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	71.43	7.06	2	71.43	7.06	2
3	84.62	5.89	2.85	84.62	5.58	3
4	92.31	5.93	2.85	92.31	4.62	4
5	91.67	4.98	3.83	91.67	3.93	5
6	100	6.3	2.7	100	3.43	6
8	100	5.35	3.4	100	2.73	8



## 14. Usuario XCL

Usuario XCL	
Palabra	# Series
CALOR	15
CARINO	15
SUSHI	15
SUSHI	15
GATO	15
PEZ	15
PERRO	15
CASA	5
LORO	3
LIBRO	2

Pruebas de Validación Cruzada, Usuario XCL						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	93.33 ± 14.06	7.2 ± 0.29	1.93 ± 0.14	96.67 ± 10.54	7.06 ± 0	2 ± 0
3	90 ± 16.1	6.26 ± 0.42	2.5 ± 0.28	93.33 ± 14.06	5.58 ± 0	3 ± 0
4	83.33 ± 23.57	6.13 ± 0.94	2.67 ± 0.67	93.33 ± 14.06	4.62 ± 0	4 ± 0
5	93.33 ± 14.06	5.91 ± 0.49	2.77 ± 0.39	96.67 ± 10.54	3.93 ± 0	5 ± 0
6	100 ± 0	5.46 ± 0.47	3.15 ± .47	100 ± 0	3.43 ± 0	6 ± 0
8	100 ± 0	5.68 ± 0.52	2.95 ± 0.44	100 ± 0	2.73 ± 0	8 ± 0

Validación, Usuario XCL						
N=Ñ	Metodología Propuesta			P300_Speller®		
	Precisión (%)	letras/min	series/letra	Precisión (%)	letras/min	series/letra
2	77.78	7.34	1.89	66.67	7.06	2
3	75	7	2.13	75	5.58	3
4	100	6.44	2.57	100	4.62	4
5	85.71	5.98	2.86	85.71	3.93	5
6	100	6.22	2.57	100	3.43	6
8	100	5.14	4	100	2.73	8

### USUARIO XCL

