

## HERRAMIENTA DE SOFTWARE PARA EL ANÁLISIS DE DATOS ELECTROFORÉTICOS

J. Cuadros<sup>a</sup>, J. L. Paredes<sup>b</sup>, G. Ceballos<sup>b</sup>

<sup>a</sup>Postgrado de Ingeniería Biomédica. Universidad de los Andes, Mérida 5101, Venezuela,  
jcuadros@unet.edu.ve, <http://www.ing.ula.ve/~gibula/>

<sup>b</sup>Escuela de Ingeniería Eléctrica. Facultad de Ingeniería. Universidad de los Andes, Mérida 5101,  
Venezuela, paredesj@ula.ve, ceballos@ula.ve, <http://www.ula.ve>

**Palabras clave:** Electroforesis Capilar, Procesamiento de Señales electroforéticas, wxWidgets, wxDevCpp, transformada Wavelet.

**Resumen.** La electroforesis capilar (EC) es una técnica de separación que, basada en la química analítica, proporciona un mapa detallado (señal en el tiempo) de las sustancias que conforman una muestra bioquímica, representando las sustancias por picos donde la amplitud de los picos está relacionada con la concentración de las sustancias en la muestra. Su valor radica en la capacidad de separar macromoléculas de interés en las áreas de biomédica, biofarmacéutica, alimentos, así como en el área de control ambiental. Debido a la poca reproducibilidad inherente en las señales electroforéticas, el análisis de los electroferogramas está basado en procedimientos visuales y manuales, tanto para el reconocimiento de patrones como para la medición de la concentración de las sustancias. En consecuencia, un método automatizado para la visualización y análisis de electroferogramas resulta ser altamente deseable. En este trabajo se presenta una herramienta para el procesamiento y análisis de señales provenientes de electroforesis capilar (EC). En dicha herramienta los datos electroforéticos son previamente acondicionados en el dominio wavelet. Dicho acondicionamiento involucra corrección de línea base, detección de zona de actividad, eliminación de saltos abruptos y reducción de componentes ruidosos. Se empleó para su desarrollo herramientas de software libre y de código abierto como wxWidgets (herramienta de visualización), ofreciendo una interfaz amigable, de fácil manejo, intuitiva, que permite satisfacer los requerimientos informáticos a los médicos e investigadores en la interpretación de los electroferogramas. Para ello cuenta con una serie de opciones tanto de procesamiento como de visualización, integradas en pestañas verticales, barras horizontales (menú de opciones y herramientas de acceso rápido) y áreas de graficación donde se muestran los registros originales y los procesados. Adicionalmente, la herramienta cuenta con un manual de usuario elaborado empleando técnicas de multimedia donde se muestra el funcionamiento de la aplicación. La programación modular de la herramienta facilita el mantenimiento del software implementado así como el desarrollo a futuro por parte de terceros quienes podrán incluir nuevos módulos de procesamiento y/o visualización ó modificar los existentes sin mayores complicaciones. Esta herramienta corre en Windows 95/98/2000/NT/XP en un computador con un mínimo de recursos, ofreciendo la ventaja adicional de poder ser instalado en distintos sistemas operativos como Windows® y UNIX/Linux.

## 1 INTRODUCCIÓN

La electroforesis capilar (EC) es una técnica de separación que, basada en la química analítica de interés, proporciona un mapa detallado de las sustancias que conforman la muestra. El registro resultante es un electroferograma en el cual se representa cada sustancia por una forma de onda similar a una campana de Gauss en donde la amplitud de cada pico está relacionada con la concentración de cada sustancia (Ceballos, 2005). Su valor radica en la alta resolución al separar macromoléculas de interés en las áreas de biomédica (en el campo de las proteínas, péptidos, ADN, análisis de líquidos de perfusión, monitoreo de drogas, marcadores genéticos tumorales y neurobioquímicos, y pericias forenses), biofarmacéutica (control de calidad de productos farmacéuticos y biotecnológicos, quimioterápicos y de estructura quiral), alimentos (fraccionamiento y cuantificación de aminoácidos, hidratos de carbono, ácidos orgánicos, aditivos y contaminantes) así como en el área de control ambiental (identificación de contaminantes y sus metabolitos, pesticidas, metales pesados e hidrocarburos). Debido a la poca reproducibilidad inherente en el proceso de EC, actualmente el análisis de los registros electroforéticos está basado en procedimientos asistidos, tanto para el reconocimiento de patrones como para la medición de la concentración de las sustancias. En consecuencia, un método automatizado para la visualización y análisis del electroferograma resulta ser altamente deseable.

En este trabajo se presenta una herramienta para el análisis y acondicionamiento de registros electroforéticos conformada por dos módulos. El primero, denominado “Módulo de Acondicionamiento ELF (Electroforético)”, permite el pre-acondicionamiento y visualización de electroferogramas basado en la Transformada de Wavelet, el segundo módulo denominado “Módulo de Reconocimiento de Patrones ELF” tiene como finalidad proveer al especialista una serie de herramientas que le permita realizar el reconocimiento de patrones en electroferogramas acondicionados (generados por el primer módulo) mediante técnicas basadas en programación dinámica en el dominio wavelet. La aplicación está enmarcada dentro de un proyecto de desarrollo de herramientas de software libre para brindar soporte visual a los médicos e investigadores en la interpretación de los electroferogramas, proporcionar componentes reutilizables para cualquier aplicación o sistema que lo requiera, así como también servir de base para continuar estudios en el área de electroforesis capilar. La aplicación implementa, usando software libre, los algoritmos de preprocesamiento propuestos en Ceballos, Paredes y Hernández (2008), como lo son: limpieza, corrección de línea base, aproximación por gaussianas y reconocimiento de patrones en el dominio wavelet usando programación dinámica, bajo un ambiente de interfaz gráfica amigable.

La entrada al módulo de “Acondicionamiento ELF” es un archivo que contiene la señal temporal del registro electroforético generado por un equipo de electroforesis capilar. Una vez que el electroferograma es cargado y procesado puede ser visualizado en el dominio temporal, el usuario puede cambiar los parámetros de procesamiento tales como el nivel de descomposición wavelet, umbral de denoising, parámetros asociados a la aproximación en gaussianas, entre otros. El electroferograma pre-acondicionado constituye la entrada para el segundo módulo en el cual se aplican técnicas de reconocimiento de patrones, visualizándose el patrón seleccionado, el patrón encontrado en un conjunto de registros electroforéticos pre-acondicionados, el alineamiento, el índice de similitud entre el patrón seleccionado y encontrado, y curva de variación de concentración de sustancias. Al igual que en el primer módulo, en el segundo módulo el usuario puede definir los parámetros para el reconocimiento de patrones como penalización en la inserción de espacios para los alineamientos, rigurosidad en el alineamiento, banda de ubicación esperada, entre otros.

Adicionalmente, la herramienta ofrece numerosas ventajas entre las que se pueden citar: puede ser ejecutada en equipos personales con un mínimo de requerimientos, es multiplataforma y posee una interfaz gráfica amigable al usuario. Esto se logra gracias al carácter híbrido de la aplicación pues se emplean para su implementación la herramienta de software libre wxWidgets, por lo que es accesible económicamente, en comparación con otros sistemas de visualización de electroforesis capilar que requieren la adquisición de costosas licencias (wxWidgets, 2007).

## 2 MARCO TEÓRICO

La Electroforesis es una técnica de separación de compuestos químicos de alta resolución, que se basa en la migración de las moléculas según su masa y carga eléctrica, como resultado de la acción de un campo eléctrico (Pérez, 2000). La técnica de electroforesis es una de las más ampliamente empleadas en laboratorios de Biotecnología (en cualquiera de sus campos y/o líneas de investigación), ya que no sólo se usa como técnica preparativa sino como técnica principal. Además, sus aplicaciones contemplan todo un amplio abanico de áreas o campos de acción: salud, medicina forense, criminalística, antropología, evolución, estudios de parentesco, en la industria en sus diferentes áreas: alimenticia, ambiental, farmacéutica, de cosméticos, entre otras. Esta técnica es particularmente importante por sus implicaciones en el Proyecto Genoma Humano (CCS, 2007).

Se utiliza en dos modalidades distintas: Electroforesis Convencional en gel y Electroforesis Capilar (EC). En la primera, las separaciones se llevan a cabo sobre una capa delgada y plana de un gel semisólido y poroso que contiene un tampón acuoso en el interior de sus poros, dando origen a una imagen donde se visualizan las sustancias en forma de zonas teñidas. En la Electroforesis Capilar la separación se lleva a cabo en un tubo capilar muy delgado. Cuando se aplica un campo eléctrico elevado a este tubo, la muestra migra a lo largo del capilar a la vez que sus componentes se separan y eluyen de la columna a intervalos diferentes. Esta técnica presenta mejoras sustanciales respecto a la electroforesis convencional en gel, principalmente en cuanto a los pequeños volúmenes de sustancia necesaria para el estudio (en el orden de los nanolitros) y la rapidez con que se obtienen los resultados del análisis.

Una de las ventajas de la EC sobre otras técnicas de separación es que requiere de una instrumentación relativamente simple: una fuente de alto voltaje, un capilar cuyos extremos se encuentran sumergidos junto con dos electrodos en dos viales que contienen una solución amortiguadora, un fotodetector y un sistema de adquisición de datos, tal como se ilustra en la Figura 1.

Los volúmenes de la muestra a analizar son extraordinariamente pequeños, tomando valores entre 0,1 a 10 nl. Los equipos de Electroforesis Multicapilar (Electroforesis Capilar, 2007), por otro lado, poseen la capacidad de llevar a cabo el proceso de electroforesis en varios capilares simultáneamente, permitiendo el análisis de mayor cantidad de muestras en el tiempo y por ende la generación de grandes volúmenes de datos.

Los datos obtenidos en una corrida de Electroforesis Capilar se representan en un Electroferograma, que es una gráfica de intensidad (de emisión de UV, o fluorescencia) en función del tiempo (en minutos). La duración de una corrida suele ser entre 9 y 45 minutos y tiene el aspecto que se observa en la Figura 2. Cada pico corresponde a una sustancia en particular. La ubicación de una sustancia en el tiempo depende de su relación carga/masa, las sustancias ionizadas positivamente (izquierda) son detectadas por el sensor antes que las sustancias ionizadas negativamente (derecha). La concentración de una determinada sustancia se calcula midiendo la altura del pico respectivo o el área encerrada por la correspondiente curva. En un experimento es común realizar varias corridas obteniéndose varios electroferogramas con los cuales se puede observar y medir las variaciones de concentración

de determinadas sustancias bajo diferentes condiciones de estudio, por lo que generalmente existen patrones de picos, como el que se muestra en la [Figura 2](#), que pueden repetirse con cierto grado de similaridad en los electroferogramas producto de un mismo experimento.

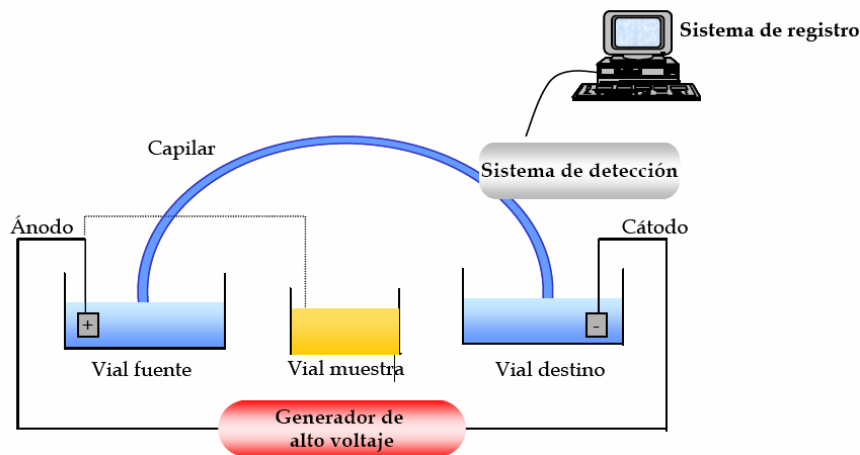


Figura 1. Esquema de un sistema de Electroforesis Capilar.

Existen varios factores que afectan la reproducibilidad de los resultados de los análisis, por ejemplo: la forma física que tienen los extremos de los capilares, el acondicionamiento del capilar antes de ser utilizado, la calidad del fluorocromógeno, así como las condiciones en que se mantienen los capilares cuando no se utilizan. Su efecto se refleja en la calidad de inyección, tiempos de migración distintos para una misma sustancia, así como presencia de picos con largas colas en los resultados. Experimentalmente se ha comprobado que cuando se utilizan capilares de sílice fundida, un acondicionamiento inicial al pH alcalinos es suficiente para obtener un valor constante de movilidad electroosmótica y una reproducibilidad satisfactoria de los tiempos de migración ([Electroforesis Capilar, 2007](#)).

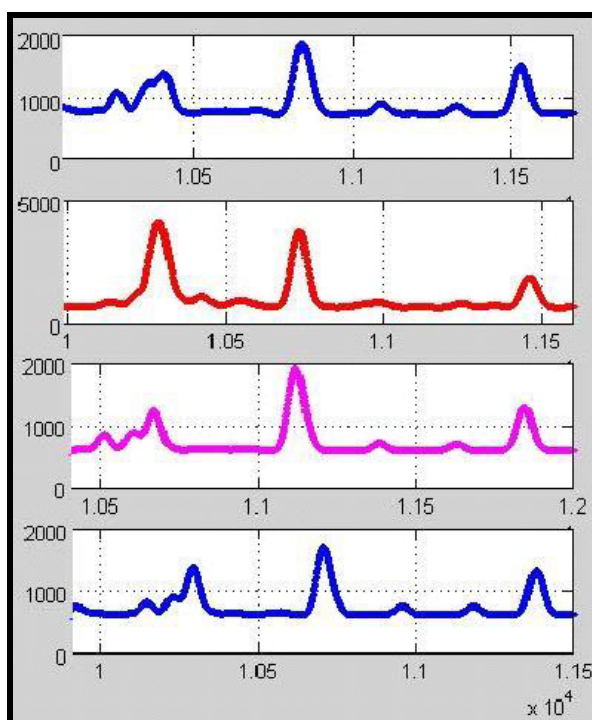


Figura 2. Patrón en varios electroferogramas.

### 3 RECONOCIMIENTO DE PATRONES EN ELECTROFORESIS CAPILAR

El análisis de grandes volúmenes de datos electroforéticos requiere la implementación de técnicas de reconocimiento que permitan, de manera rápida y confiable, detectar la presencia ó no de determinada sustancia para su posterior medición. Partiendo del concepto más elemental, el reconocimiento de patrones tiene como finalidad la clasificación de objetos en un número de categorías o de clases. Dependiendo de la aplicación, estos objetos pueden ser imágenes o formas de onda de una señal o cualquier tipo de medidas que necesiten ser clasificadas (Roa, 2002). En nuestra aplicación en particular, se está interesado en reconocer patrones (picos sucesivos) de sustancias presentes en la muestra a analizar.

En general, el reconocimiento de patrones como los observados en la Figura 2, se realiza de manera visual por parte del especialista. Observe, en la Figura 2, que la similaridad se ve gravemente afectada por la poca reproducibilidad inherente a la EC y la variación importante de la concentración de una o varias sustancias (picos) en el patrón, por lo que el reconocimiento de patrones en este tipo de señales es un problema de investigación que presenta innumerables retos. Note que los picos no miden exactamente lo mismo en todos los electroferogramas, es decir, no son de la misma altura o no encierran la misma área, no se encuentran ubicados en los mismos instantes de tiempo y no están separados unos de otros por la misma diferencia temporal en los electroferogramas. Estas características hacen que el proceso de reconocimiento de patrones en forma automática, sea más complicado que en otros tipos de señales con menor variabilidad, fallando las técnicas tradicionales de reconocimiento.

## 4 MATERIALES Y MÉTODOS

### 4.1 Preprocesamiento con Wavelets

En este trabajo se usa la transformada wavelet discreta como una herramienta de procesamiento de señales para hacer análisis multiresolución (reducir la dimensión de los datos sin perder la información relevante para el reconocimiento de patrones en los electroferogramas), para detectar eventos temporales y para la supresión de ruido. En su forma más general, una señal discreta  $x(k)$ , se puede representar como una combinación lineal de funciones contraídas en tiempo y desplazadas a lo largo del tiempo (Burrus, Gopinath y Guo, 1998), como se observa en la Eq. (1):

$$x(k) = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k} \Phi_{j_0,k}(t) + \sum_{j=j_0}^{N-1} \sum_{k=0}^{2^j-1} d_{j,k} \Psi_{j,k}(t) \quad \text{con } j, k \in \mathbb{Z}^+ \quad (1)$$

Donde  $C_{j_0,k}$  son los coeficientes de escala o de aproximación a menor resolución y  $d_{j,k}$  son los coeficientes wavelets o de detalles.  $\Phi_{j_0,k}(t)$  Es la función escalar desplazada al instante  $k$ ,  $\Psi_{j,k}(t)$  es la wavelet madre (Mallat, 1989).

En este trabajo se empleó la wavelet Symlet4. La familia de wavelets Symlet surge de modificaciones sobre la familia db (Daubechies) para obtener un mayor grado de simetría en la forma de onda de la wavelet, la Symlet4 se caracteriza por poseer un momento de desvanecimiento para la función wavelet de cuarto orden, es de soporte compacto y biortogonal.

Aunque la escogencia de una familia de wavelet pudiera considerarse que depende de la aplicación en estudio, en este trabajo se adoptó la wavelet probada en Weidong, Xiaoyan, Xiurong y Erkang (2003), donde mostraron tener un buen desempeño para el análisis de Electroferogramas.

Para esta aplicación se usa la transformada wavelet a fin de disminuir la resolución del electroferograma así como para su limpieza y posterior procesamiento. En particular, para hacer la detección de patrones usando programación dinámica se usarán los coeficientes de aproximación en el nivel 4 de descomposición, en el cual se preserva en gran medida la forma de la señal original y a la vez una reducción de la dimensión de los datos en un factor de  $2^4$  y un suavizamiento en la forma de la señal, producido por el proceso de filtrado pasabajo sucesivo implícito en la descomposición Wavelet.

Es posible probar que la correspondencia temporal entre el número de coeficiente de aproximación del nivel 4 y la muestra asociada de la señal original es dado por la Eq. (2):

$$P_{orig} = k2^j - \lambda \sum_{i=1}^j 2^i \quad \text{para } k \geq 2\lambda \quad (2)$$

donde  $P_{orig}$  corresponde al número de muestra en la señal original (posición temporal en el electroferograma),  $k$  es la posición o número de coeficiente de aproximación o de detalle en el nivel de descomposición  $j$  y  $\lambda$  viene dada por la Eq. (3):

$$\lambda = \begin{cases} 1 & \text{Para coeficientes de aproximación} \\ 2 & \text{Para coeficientes de detalle} \end{cases} \quad (3)$$

## 4.2 Codificación a caracteres

Luego de aplicar Wavelets en el electroferograma (Ceballos, Paredes y Hernández, 2008), para reducir la dimensión de los datos, eliminar ruido y limitar la zona de búsqueda, se aplica programación dinámica para la búsqueda del patrón. En este trabajo se realizó una transformación de cada valor numérico (coeficiente de aproximación) a un carácter para posteriormente aplicar los algoritmos de alineamiento local similar a como se hace en Bioinformática (en base a cadenas de caracteres). La codificación propuesta se define así: Sea  $Y(n) \in \mathfrak{R}$  los coeficientes de aproximación de la señal a un nivel  $j$ -enésimo de descomposición Wavelet, donde  $\mathfrak{R}$  denota el conjunto de los números reales. Sea  $\Omega = \{M, L, P, Car(i)\}$  para  $i=1,2,\dots,\gamma$  el conjunto de caracteres a ser usado en la codificación, donde  $Car$  es un vector cuyos elementos son caracteres y  $\gamma$  es el número de niveles de cuantización para la altura de los picos. La función de codificación a caracteres  $G$  se define en la Eq. (4):

$$G: \mathfrak{R} \rightarrow \Omega \left\{ \begin{array}{ll} G(Y(n)) = M & \text{si } (Y'(n) > 0) \& (Y'(n+1) > 0) \\ G(Y(n)) = L & \text{si } (Y'(n) \leq 0) \& (Y'(n+1) \geq 0) \\ G(Y(n)) = P & \text{si } (Y'(n) < 0) \& (Y'(n+1) < 0) \\ G(Y(n)) = Car\left(\left\lfloor \gamma \frac{Y(n)}{m_{PP}} \right\rfloor\right) & \text{si } (Y'(n) > 0) \& (Y'(n+1) \leq 0) \end{array} \right. \quad (4)$$

donde  $\&$  representa el operador lógico AND,  $m_{PP}$  es el tope de altura de pico para el conjunto de datos, en el presente trabajo se utilizó una altura máxima de 4000 pues se observó que los valores de los electroferogramas no sobrepasaban dicho valor y se consideraron 8 niveles de

cuantización ( $\gamma = 8$ ). El símbolo  $\lfloor x \rfloor$  denota el mayor entero menor o igual a  $x$  y  $Y'(n)$  representa la derivada discreta. En este estudio, se explota la información relacionada a las pendientes de la señal del cuarto nivel de aproximación wavelet así como una cuantización de múltiples niveles asociados a los picos del electroferograma (Ceballos, Paredes y Hernández, 2008).

### 4.3 Algoritmo de Alineamiento local con heurística

Para aplicar el algoritmo de alineamiento local, es necesario codificar a secuencias de caracteres el patrón a buscar y los electroferogramas en los que se realizará la búsqueda, de acuerdo a la metodología de codificación descrita en la sección anterior y detallada en Ceballos, Paredes y Hernández (2008).

Los parámetros propios del algoritmo de alineamiento local de Smith & Waterman (1981), *PenalInsEspacio*, *PenalExtEspacio* y la matriz de sustitución se definieron de la siguiente manera: *PenalInsEspacio* = - 8 *PenalExtEspacio* = - 7

#### Matriz de Sustitución:

	M	P	L	A	B	C	D	E	F	G	H
M	7	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
P	-2	7	-2	-2	-2	-2	-2	-2	-2	-2	-2
L	-2	-2	7	-2	-2	-2	-2	-2	-2	-2	-2
A	-2	-2	-2	24	22	20	18	16	14	12	10
B	-2	-2	-2	22	24	22	20	18	16	14	12
C	-2	-2	-2	20	22	24	22	20	18	16	14
D	-2	-2	-2	18	20	22	24	22	20	18	16
E	-2	-2	-2	16	18	20	22	24	22	20	18
F	-2	-2	-2	14	16	18	20	22	24	22	20
G	-2	-2	-2	12	14	16	18	20	22	24	22
H	-2	-2	-2	10	12	14	16	18	20	22	24

Como se puede observar en la matriz de sustitución; donde M, P, L corresponde a la codificación en carácter de la señal wavelet si el crecimiento de la señal es positivo, negativo ó un punto mínimo local, respectivamente, mientras que los caracteres {A, B, C, D, E, F, G, H} corresponde a una cuantización uniforme de los picos del electroferograma, se puede observar que la coincidencia de un pico con otro de tamaño similar en el alineamiento se premia con un máximo puntaje (24 en este caso), mientras que si las alturas difieren se va disminuyendo (proporcionalmente a la diferencia) el puntaje asignado al apareamiento de los picos.

Adicionalmente, se puede notar en la matriz de sustitución que a los caracteres pertenecientes a *Car* se les asigna un *Puntaje* de coincidencia según la Eq. (5) :

$$Score(Car(i), Car(j)) = Mscore - |i - j| \frac{(Mscore - mscore)}{\gamma - 1} \quad (5)$$

donde *Mscore* es la máxima puntuación, asignada a la coincidencia de picos con amplitudes muy similares (en este caso 24), *mscore* es el puntaje asignado a la coincidencia de un pico de

mínima amplitud con uno de máxima amplitud (10, en este caso) y  $\gamma$ , que este caso toma el valor 8, es el número de niveles de cuantización de los picos del electroferograma.

#### 4.4 Aproximación mediante gaussianas

Para realizar la aproximación mediante suma de gaussianas de las señales electroforéticas, se prestará atención en el cálculo de los parámetros de las gaussianas iniciales al proceso de optimización. El método de optimización iterativo (mínimos cuadrados no lineal) para obtener los parámetros finales será el de Región de Confianza (Ceballos, Paredes y Hernández, 2008).

Cada gaussiana, representativa de un pico o sustancia en particular en el electroferograma, se encuentra caracterizada por 3 parámetros a saber,  $\mu$  es el parámetro de localización,  $\sigma$  el de dispersión y  $A$  el de amplitud, tal como se muestra en la Figura 3. La aproximación de una señal mediante la combinación lineal de  $n_g$  gaussianas está caracterizada por los parámetros  $\mu_i$ ,  $\sigma_i$  y  $A_i$  para  $i = 1, \dots, n_g$ , métodos de optimización iterativos para determinar estos parámetros, como el utilizado en este trabajo (Región de Confianza), requieren de parámetros iniciales a partir de los cuales comenzar el proceso iterativo. Es de hacer notar que mientras más exactos sean los valores iniciales, es decir más cercanos a los que se obtendrán del proceso iterativo, menor número de iteraciones se necesitarán para obtener los parámetros finales y por tanto menor tiempo de cálculo se requerirá para la aproximación mediante gaussianas, por ello en este trabajo se utilizó como método para el cálculo de los parámetros iniciales el análisis de la segunda derivada (Ceballos, 2005).

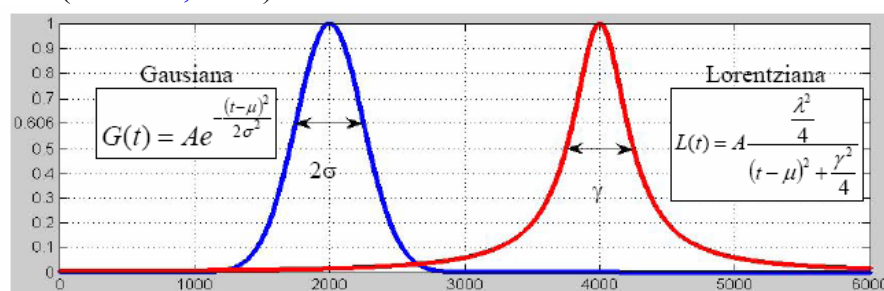


Figura 3. Parámetros de localización, dispersión y amplitud de una gaussiana

## 5 MÓDULO DE PROCESAMIENTO Y MÓDULO DE RECONOCIMIENTO

El primer módulo Acondicionamiento ELF, permite el procesamiento y visualización de registros electroforéticos. Consiste en dar como entrada a la herramienta un electroferograma proveniente de un equipo electroforético. Una vez que el electroferograma es cargado y procesado (con los parámetros de procesamiento seleccionados por el usuario), se visualiza la línea base, la señal original sin ruido, así como las señales tipo gaussiana cuya superposición lineal se aproximan al registro original. Luego que el electroferograma es acondicionado puede ser procesado por el segundo módulo (Reconocimiento de Patrones), visualizándose el patrón seleccionado, el patrón encontrado, el alineamiento, el índice de similitud entre el patrón seleccionado y encontrado, curva de variación de la concentración de una determinada sustancia para diferentes electroferogramas. Al igual que en el primer módulo, el usuario puede definir los parámetros para el reconocimiento de patrones como penalización en la inserción de espacios para los alineamientos, rigurosidad en el alineamiento, banda de ubicación esperada, entre otros. Estos resultados se pueden guardar a disco así como el entorno de trabajo (parámetros con los cuales se procesó el electroferograma).



El primer módulo consta de tres etapas fundamentales. En primer lugar el pre-procesamiento de los datos utilizando análisis multiresolucional basado en la Transformada Wavelet, herramienta mediante la cual se reduce la dimensión de los datos, se eliminan componentes ruidosos, y se eliminan zonas innecesarias para el análisis de los electroferogramas. En segundo lugar está el reconocimiento de patrones utilizando Programación Dinámica mediante el Algoritmo de Alineamiento Local ([Smith & Waterman, 1981](#)) para lo cual se hace necesaria la codificación de las señales a caracteres y se implementan algunas optimizaciones a fin de determinar el mejor alineamiento entre los registros electroforéticos. Y en tercer lugar, la aproximación de los electroferogramas mediante suma de gaussianas, para mejorar la precisión en la medición de las alturas de los picos, encontrar una representación aproximada del electroferograma usando un modelo paramétrico que permita la detección de señales con tiempos de migración parecidos.

## 6 DISEÑO DE LA HERRAMIENTA

La selección de la herramienta de visualización se realizó de acuerdo a las exigencias requeridas para el desarrollo de la aplicación, destacándose entre ellas:

- Interfaz amigable con el usuario.
- Rápido desarrollo.
- Independencia de plataforma.
- Estructura orientada a objetos (OO).
- Licencias de las librerías de software que no restrinjan o condicionen los términos que tendrá la licencia del software a desarrollar.

Por tal razón, se decidió utilizar wxWidgets como herramienta de visualización, en su versión 2.8.0, que es la versión estable mas reciente. Se examinaron otras herramientas tales como GTK ([Gtk, 2007](#)), Qt ([Qt, 2007](#)), Fltk ([Fltk, 2007](#)), siendo wxWidgets la herramienta que satisface las exigencias del diseño. En cuanto al lenguaje de programación se decidió utilizar C++, pues permite que la aplicación no dependa de una plataforma en particular. En lo que se refiere al entorno de desarrollo para la implementación del sistema, se seleccionó wxDevCpp ([wxDevCpp, 2007](#)), debido a que ofrece una interfaz amigable para el programador, funcionalidad para la depuración de programas y permite trabajar en un ambiente orientado a objetos, el cual es el área de trabajo de la herramienta de visualización seleccionada.

La biblioteca wxWidgtes, soporta varios sistemas operativos, a saber: Windows® / Unix / Mac, entre otros, pero con la particularidad que de acuerdo a la plataforma en que se ejecute la aplicación, ésta tomará la apariencia del sistema operativo donde se esté ejecutando. Posee funcionalidades para muchos aspectos aparte de la construcción de interfaces gráficas, como gráficos 2D, 3D con OpenGL, Bases de Datos (ODBC), Redes, Impresión, Hilos, entre otras ([wxWidgets, 2007](#)).

## 7 RESULTADOS

La aplicación desarrollada a lo largo de este trabajo, es una herramienta para el procesamiento de los registros electroforéticos orientada a objetos, diseñada con la intención de satisfacer las necesidades de sus usuarios mediante el uso de una interfaz sencilla y flexible a sus requerimientos. Para ello cuenta con una serie de opciones tanto de procesamiento como de visualización, integradas en pestañas verticales, barras horizontales (título, panel, herramientas) y un área de graficación donde se muestran los registros originales y los procesados, ver [Figura 4](#) para mayores detalles.

## 7.1 Primer Módulo: Acondicionamiento

La barra de Menú (superior izquierda) de la [Figura 4](#) permite al usuario manejar todo lo concerniente a **Archivos** (Abrir, Guardar, Imprimir y Vista Previa), **Editar** (Deshacer, Rehacer, cambiar color de fondo del área de graficación (*background*), definir color y tipo de línea), **Ver** (Habilitar o deshabilitar las áreas de trabajo), **Gráfica** (herramienta de mano: paneo, acercarse (*zoom in*), alejarse (*zoom out*), definir leyendas, entre otros), **Configuración** (archivo leído, archivo acondicionado), así como también consultar la **Ayuda** de la aplicación. Adicionalmente existe una barra de acceso rápido (parte inferior de la barra de Menú) a las operaciones más frecuentes como por ejemplo: abrir y guardar archivos, vista previa, impresión, herramienta de selección, herramienta de mano, Alejarse/Acercarse a la gráfica, rehacer, deshacer, ayuda, entre otros. En la parte superior izquierda de la ventana se muestra el área de graficación, en ella se visualiza el electroferograma original (parte superior) en la escala tiempo vs amplitud, el electroferograma procesado (parte inferior), así como un área informativa (debajo del área donde se visualiza el electroferograma procesado) donde se muestran reportes del estatus de la aplicación en la medida que va procesando los electroferogramas.

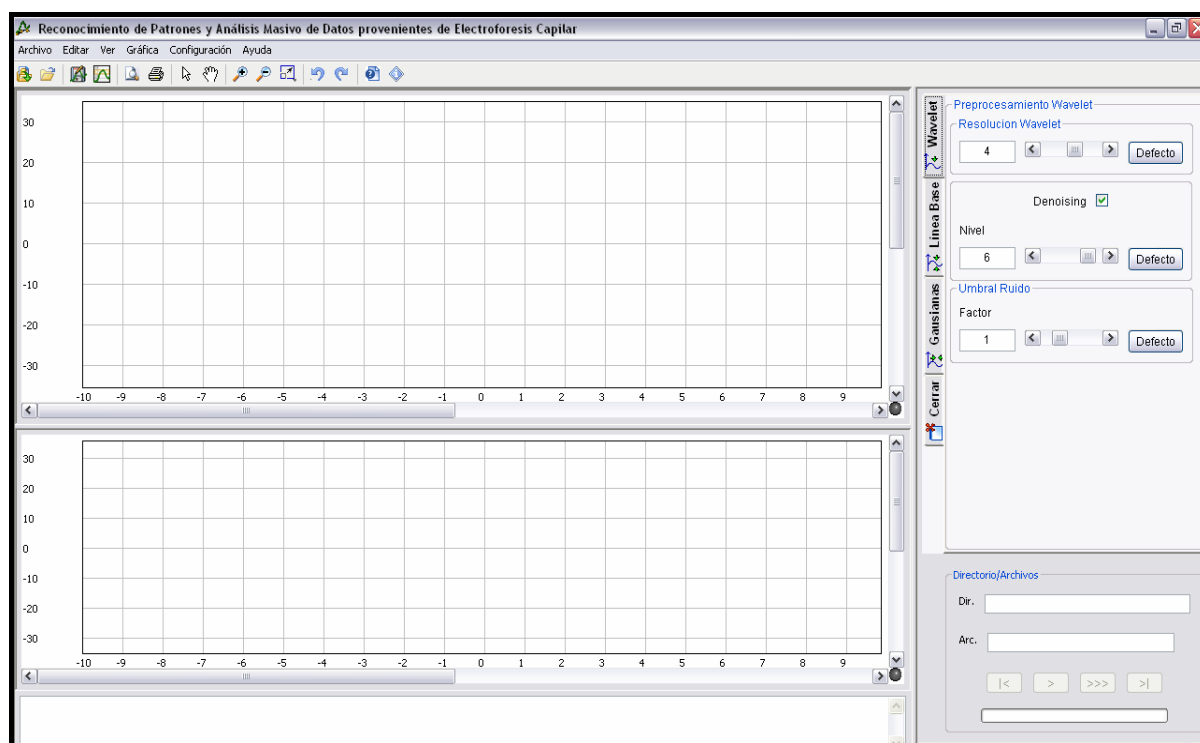


Figura 4. Módulo para el Procesamiento y Visualización de Registros Electroforéticos

En la [Figura 6](#) se muestra un ejemplo típico de cómo se ve el electroferograma original y el procesado, así como también un reporte del estatus de la aplicación. Cabe destacar que el color del background, color y tipo de línea son fácilmente configurados por el usuario.

En el sector superior derecho de la ventana aparecen un conjunto de pestañas a través de las cuales el usuario puede definir los parámetros más apropiados para iniciar el procesamiento de los electroferogramas.

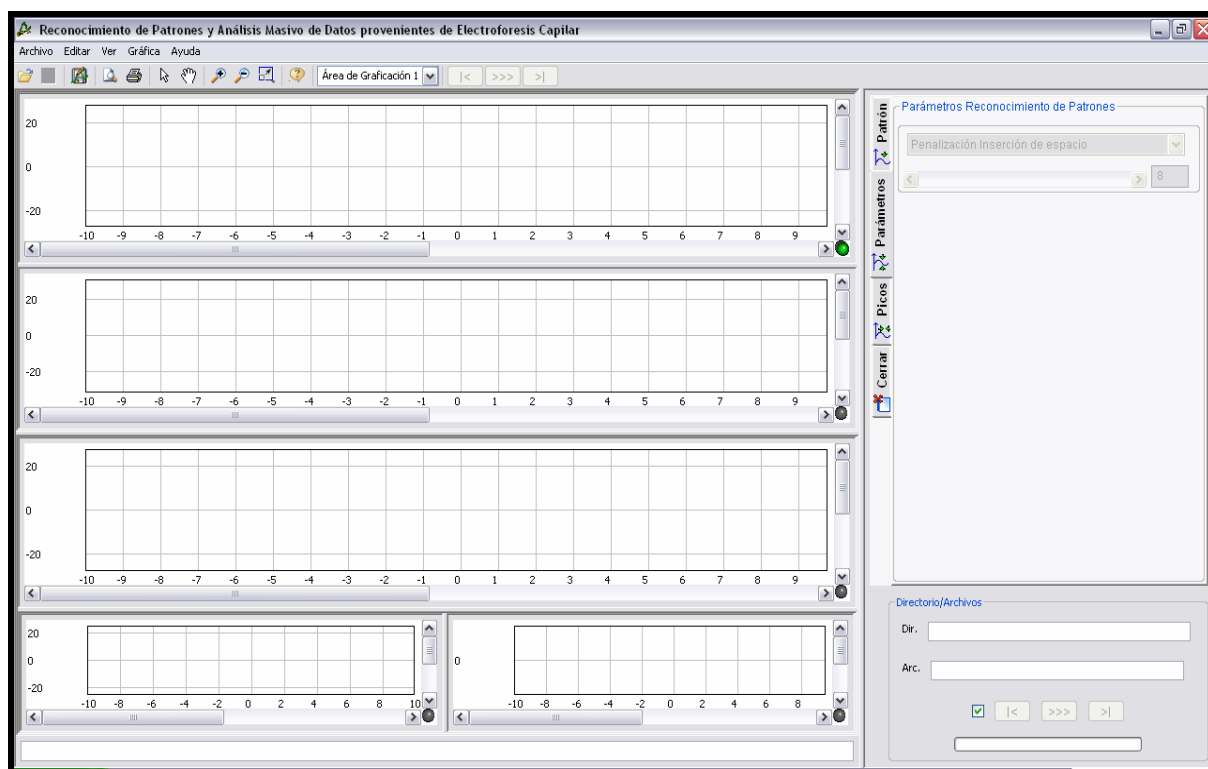


Figura 5. Módulo para el Reconocimiento de Patrones en Registros Electroforéticos

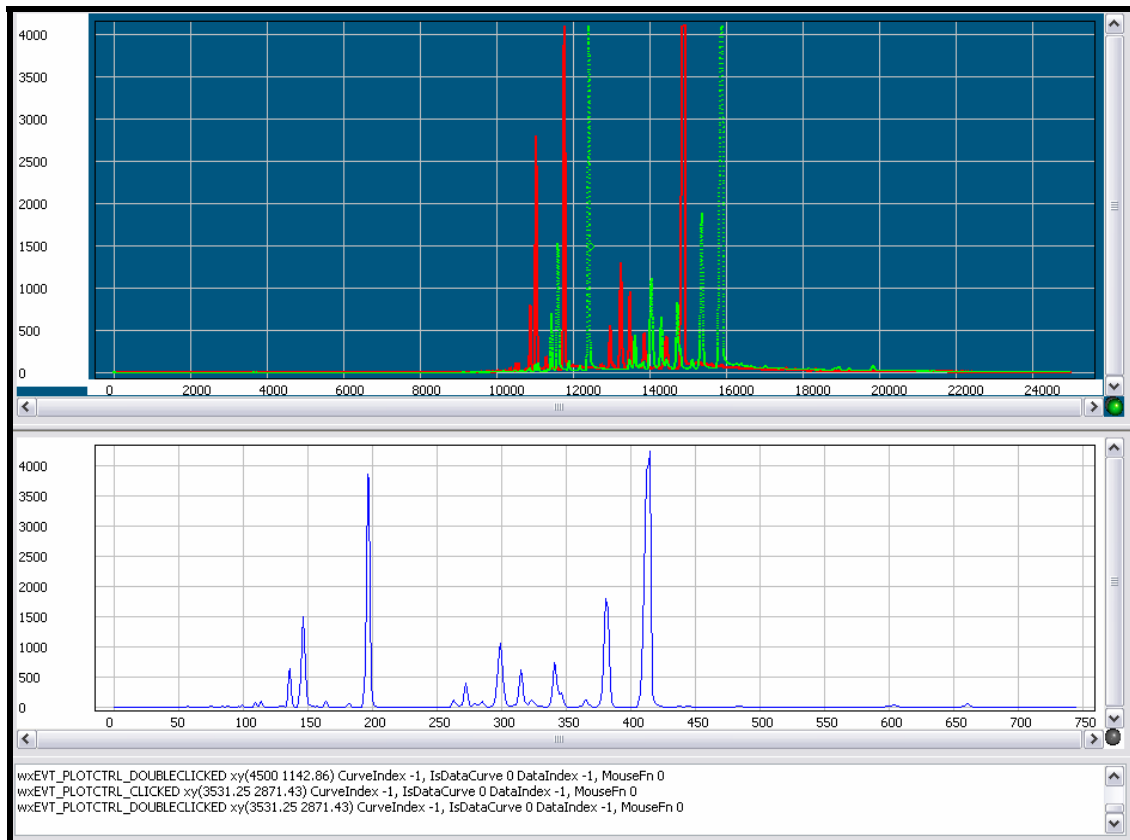


Figura 6. Visualización de un electroferograma original, su resultado y el área de comentarios, cambiando el color de fondo del área de graficación, tipo y color de línea.

La Figura 7(a) muestra la pestaña de parámetros Wavelet, en ella el usuario define: el nivel de resolución al cual se realizará la descomposición wavelet y en el cual se va a hacer el reconocimiento de patrones, habilita/deshabilita la limpieza del ruido en el electroferograma, selecciona el nivel de resolución en el cual se va a hacer la limpieza y especifica el factor que controla la cantidad de ruido a ser eliminado de la señal (Ceballos, Paredes y Hernández, 2008).

A través de la pestaña Línea Base, mostrada en la Figura 7(b), el usuario puede: habilitar la corrección automática de línea base, variar parámetros de control en el procedimiento automático descrito en Ceballos, Paredes y Hernández (2008). Adicionalmente, permite hacer una corrección asistida sobre la línea base escogiendo manualmente aquellos puntos del electroferograma que pertenecen a la misma.

La pestaña de parámetros Gaussianas mostrada en la Figura 7(c) permite: visualizar la representación de electroferograma mediante suma de Gaussianas, definir el nivel de resolución Wavelet en el cual se realizará la aproximación mediante Gaussianas, y establecer el umbral del electroferograma por debajo del cual no se realizará aproximación por gaussianas, eliminar gaussianas no deseadas en el resultado, entre otros.

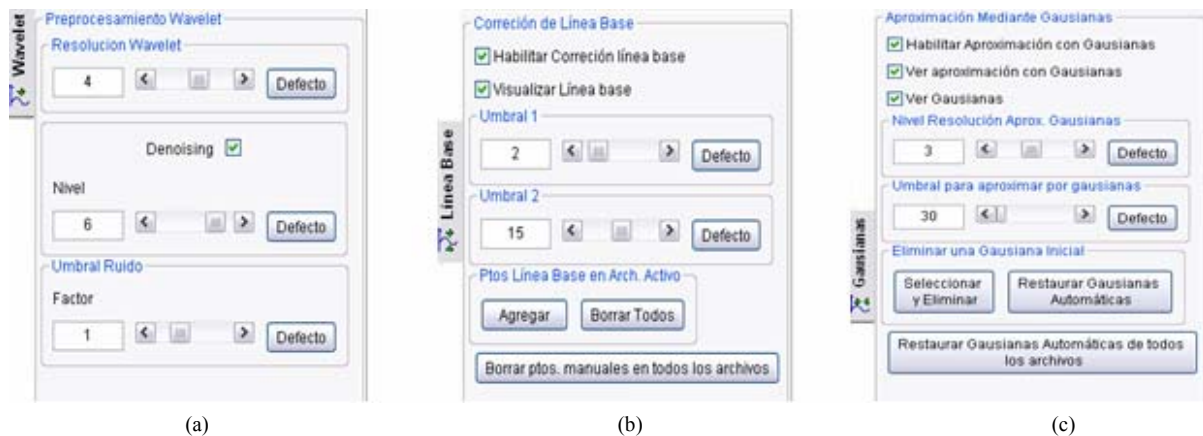


Figura 7. Pestañas de configuración de parámetros (a) Parámetros de Procesamiento Wavelet, (b) Parámetros de Procesamiento Línea Base, (c) Parámetros de Procesamiento Gaussianas.

Como valor agregado a la aplicación, ésta puede ser ejecutada en equipos personales con un mínimo de requerimientos (procesador Pentium IV, de 3 GHz y 1Gb de memoria RAM), usando versiones de Windows® igual ó mayor a Windows® 95, ó usando otros sistemas operativos como Linux ó Unix. Esto se logra gracias al carácter híbrido de la aplicación pues se emplean para su implementación algunas herramientas de software libre y algunas de código abierto como wxWidgets (herramienta de visualización), por lo que es accesible económicamente, en comparación con los sistemas de visualización de los registros electroforéticos existentes en el presente.

## 7.2 Segundo Módulo: Reconocimiento de patrones

Al igual que el módulo anterior, el módulo Reconocimiento ELF (ver la Figura 5) cuenta con una barra de Menú (superior izquierda) la cual le permite al usuario manejar todo lo concerniente a **Archivos** (Abrir, Guardar, Imprimir y Vista Previa), **Editar** (definir el color de fondo del área de graficación (*background*), color y tipo de línea), **Ver** (Habilitar o deshabilitar vista del área de trabajo), **Gráfica** (herramienta de mano: paneo, acercarse (*zoom in*), alejarse (*zoom out*), establecer leyendas sobre los gráficos, entre otros), así como también

consultar la **Ayuda** de la aplicación. Adicionalmente existe una barra de acceso rápido (debajo de la barra de Menú) a las operaciones más frecuentes similar a la aplicación descrita anteriormente.

En la parte superior izquierda de la ventana de la [Figura 5](#) se muestra el área de graficación del archivo de referencia, en ella se visualiza el electroferograma acondicionado en el cual se seleccionará el patrón que se desea buscar en un conjunto de electroferogramas. Adicionalmente, se tiene el área de graficación de los archivos donde se busca el patrón, área de graficación del patrón buscado y encontrado, ambos alineados, el índice de similitud entre el patrón buscado y el patrón encontrado, y la curva de medición de picos la cual refleja la tendencia que sigue la concentración de una sustancia específica ante cambios en las condiciones del experimento, así como un área informativa del estatus de la aplicación.

En la [Figura 8](#) se muestra un ejemplo típico de cómo se ve el electroferograma acondicionado de referencia y el procesado, así como también un reporte del estatus de la aplicación.

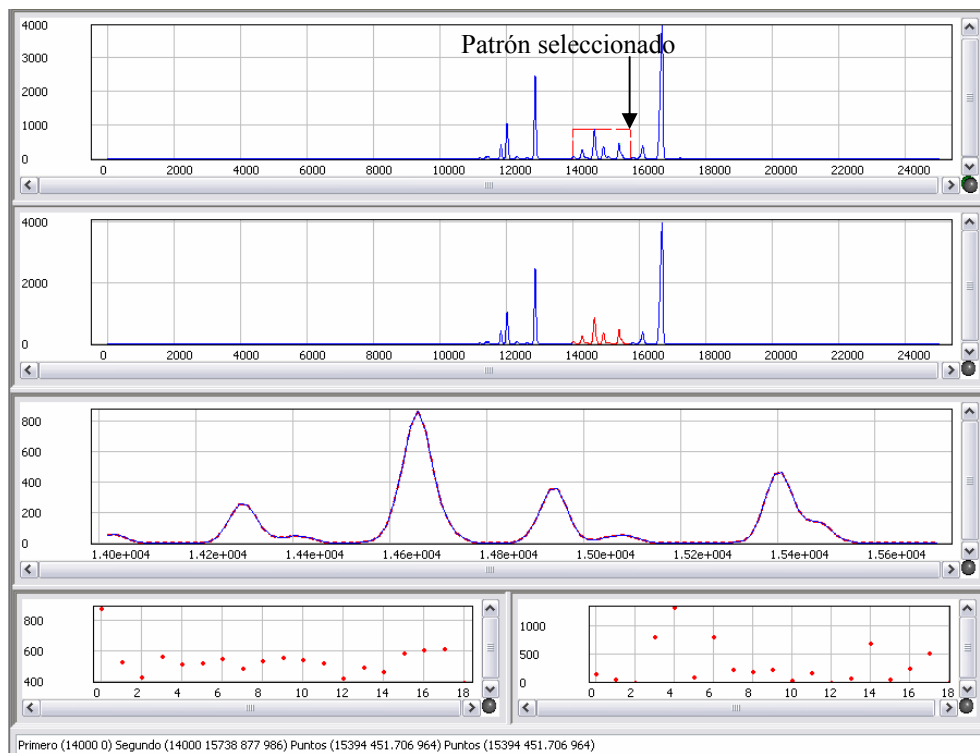


Figura 8. Visualización de un archivo electroforético de referencia donde se selecciona el patrón de interés, su resultado y el área de comentarios.

En el área superior derecha de la ventana aparecen un conjunto de pestañas a través de las cuales el usuario puede definir los parámetros más apropiados para iniciar el procesamiento de los electroferogramas.

La [Figura 9\(a\)](#) muestra la pestaña Patrón, la cual le permite al usuario seleccionar el patrón en el archivo de referencia. Así, al activarse la opción seleccionar patrón el usuario deberá escoger en el electroferograma mostrado en la parte superior, la parte del electroferograma que desea considerar como patrón.

A través de la pestaña Parámetros, mostrada en la [Figura 9\(b\)](#), el usuario puede definir: la penalización en la inserción de espacios para los alineamientos, la rigurosidad en el alineamiento, banda de ubicación esperada, entre otros parámetros.

La pestaña Picos mostrada en la [Figura 9\(c\)](#) permite: seleccionar un pico en el registro de referencia e identificarlo a través de la asignación de un nombre de una sustancia, la concentración de la sustancia asociada a dicho pico, la cual se muestra en el área de graficación inferior derecha, así como la variación que muestra dicha sustancia ante variaciones experimentales.

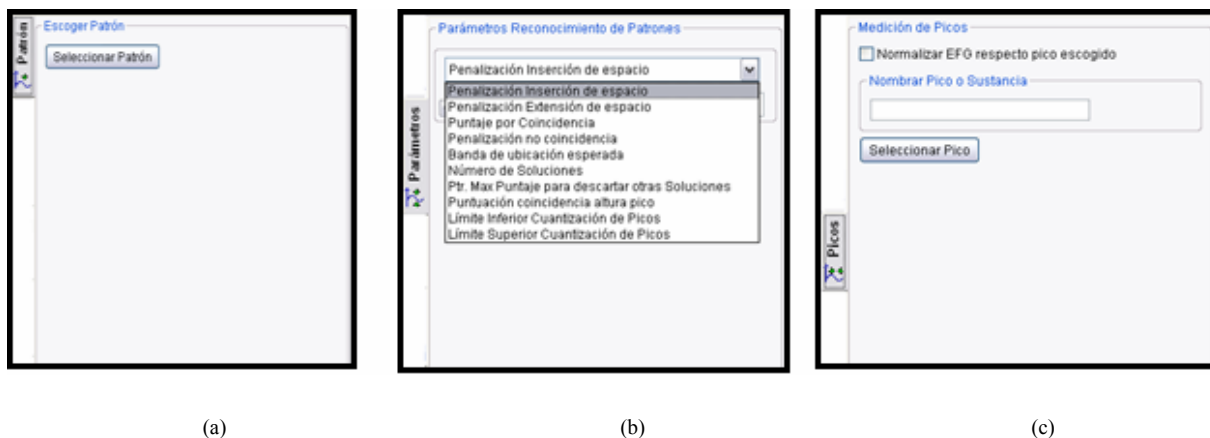


Figura 9. Pestañas (a) Patrón, (b) Parámetros de Procesamiento, (c) Picos

## 8 PRUEBA DE LA APLICACIÓN

Las pruebas se realizaron con electroferogramas provenientes de un equipo de Electroforesis Capilar (con una duración entre 9 y 45 min, a una tasa de muestreo de 41 Hz) desarrollado en el Laboratorio de Fisiología de la Conducta de la Universidad de Los Andes.

Dichos registros corresponden a muestras bioquímicas obtenidas mediante microdiálisis en el cerebro de un grupo de ratas. Se tomaron de la zona del cerebro denominada sustancia gris periacueductal, lo cual significa que, aunque los especímenes fueron sometidos a tratamientos químicos experimentales con Formalina, Suero Fisiológico y Morfina, los electroferogramas presentan patrones de picos asociados a la composición química de esa zona del cerebro en particular con modificaciones de amplitud en algunos de los picos debido a efecto de las sustancias de prueba.

## 9 DISCUSIÓN DE RESULTADOS

Para comprobar el buen desempeño de la herramienta diseñada se realizaron diferentes pruebas utilizando como referencia una aplicación desarrollada usando el software comercial MatLab®, que implementa los algoritmos presentados en detalle en [Ceballos, Paredes y Hernández \(2008\)](#). En dichas pruebas se compararon tanto los resultados numéricos como los resultados gráficos, encontrándose que no había diferencias sustanciales en los resultados, como se muestra en la [Tabla 1](#).

Nº muestra	Valor obtenido para la Línea Base usando la Herramienta desarrollada	Valor obtenido para la Línea Base con la Aplicación desarrollada en Matlab®
1	18.0838	18.083776021480173
2	18.8293	18.829308333344652
3	15.2994	15.299438323442864
4	13.0024	13.002446952051418

5	10.3008	10.300763359880357
6	8.30299	8.30298736929925
7	6.63586	6.635860478774965
8	4.93264	4.932641955568479
9	4.14109	4.141085379978974
10	2.86577	2.865772730140545

Tabla 1. Comparación de Resultados.

En la [Figura 10](#) se muestra que la gráfica generada por la herramienta desarrollada se superpone a la gráfica generada por la aplicación desarrollada en Matlab, confirmando el buen funcionamiento de la misma.

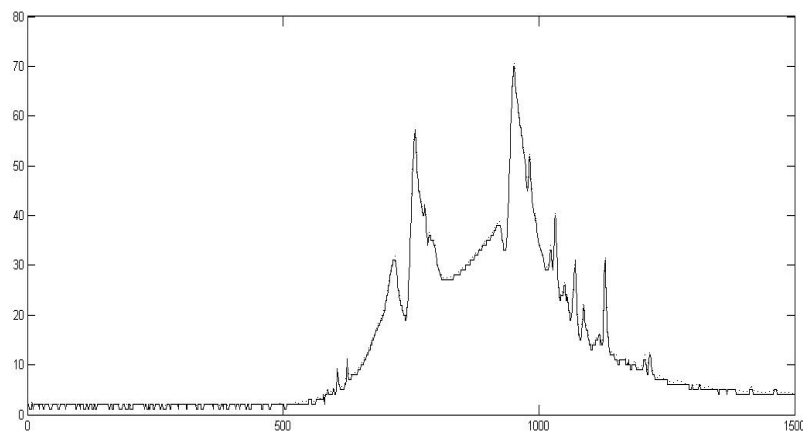


Figura 10. Electroferograma mostrado por la Aplicación desarrollada (línea punteada) y Matlab® (línea continua)

Adicionalmente a las pruebas funcionales se realizaron dos pruebas de rendimiento, usando el mismo electroferograma. Dichas pruebas permitieron obtener medidas de referencia de los recursos del sistema que requiere la herramienta. Una prueba se realizó sobre el módulo de procesamiento de la señal; la segunda prueba se realizó sobre el componente de visualización integrado con el módulo de procesamiento.

Los resultados obtenidos se muestran en la [Tabla 2](#). Para el cálculo del uso de CPU y uso de memoria se utilizó el programa Process Explorer ([Process Explorer, 2007](#)), el cual es un monitor y gestor de tareas para el sistema operativo Windows®, de distribución gratuita.

Para determinar el tiempo de ejecución se desarrolló un programa en C++ que permite cronometrar el tiempo de ejecución del programa. Debido a que la integración de ambos componentes arroja una aplicación donde las opciones de procesamiento y visualización dependen de eventos accionados por el usuario, resulta difícil determinar el tiempo de ejecución de la herramienta.

Estos resultados, que se obtuvieron utilizando un procesador Pentium IV, de 3 GHz y 1Gb de memoria RAM, muestran que el mayor porcentaje de uso del procesador se involucra en el procesamiento de la señal, aunque su tiempo de ejecución es muy corto, mientras que el mayor tiempo de ejecución lo requiere el módulo de visualización, todo ello indica que la herramienta demanda muy pocos recursos del sistema.

Módulo ó Componente	Variables		
	Uso de CP U (%)	Uso de Memo ria (KB)	Tiempo de Ejec. (seg)
Procesamiento de la Señal	100	12.276	0,0046
Componente de Visualización (.dll)	7,69	12.192	6

Tabla 2. Comparación de Resultados.

## 10 CONCLUSIONES

Se desarrolló una herramienta para el procesamiento y visualización de señales electroforéticas, bajo el paradigma de desarrollo de software libre, orientado a objetos, utilizando lenguaje C++ y la librería wxWidgets con el entorno de desarrollo wxDevCpp. Para el procesamiento de la señal se utilizó la Transformada Wavelet, la cual se calculó mediante la implementación del algoritmo propuesto por Valens (2006). Este algoritmo fue codificado en C++ debido al alto costo computacional requerido por las operaciones matemáticas involucradas en el cálculo de la transformada Wavelet. La confiabilidad de la herramienta se comprobó utilizando funciones de prueba definidas explícitamente, así como, realizando comparaciones entre los resultados obtenidos con la herramienta desarrollada en el software comercial MatLab® y la presente aplicación utilizando como patrón de comparación los mismos registros electroforéticos. Además, la herramienta puede ser ejecutada en computadores personales ya que demanda pocos recursos computacionales.

## 11 RECONOCIMIENTO

Los autores agradecen al Fondo Nacional de Ciencia, Tecnología e Investigación (FONACIT), por el soporte financiero dado bajo el proyecto N° G-2005000342 para la realización de este trabajo.

## 12 REFERENCIAS

- Burrus C., Gopinath R. y Guo H. Introduction to Wavelets and Wavelet Transforms: A Primer. Prentice Hall, pp. 268, 1998.
- CCS. <http://www.ccs.net.mx> (consultada en junio 2007)
- Ceballos, G. Reconocimiento de Patrones en electroforesis capilar utilizando Análisis Multiresolucional y Programación Dinámica. Trabajo de Grado. Universidad de Los Andes. Mérida, 2005.
- Ceballos G., Paredes J., y Hernández L. Pattern recognition in capillary electrophoresis data using dynamic programming in the wavelet domain. *Electrophoresis*, Marzo, 2008. Aceptado.
- Fltk. <http://www.fltk.org/> (consultada Enero 2007)
- Fundamentos de la Electroforesis Capilar. [www.tesisenred.net](http://www.tesisenred.net) (Consultada Junio 2007)



Gtk. <http://www.gtk.org/> (Consultada Enero 2007)

Mallat, S. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Pattern Anal. and Machine Intell. vol. 11, N° 7, pp. 674-693, 1989.

Pérez, J. Principios básicos de electroforesis capilar. Trabajo de Ascenso. Facultad de Farmacia. Universidad de los Andes. Mérida, 2000.

Process Explorer.

<http://www.microsoft.com/technet/sysinternals/utilities/ProcessExplorerer.msp>  
(Consultada en Junio 2007)

Qt. <http://www.trolltech.com/products/qt/> (Consultada Enero 2007).

Roa, F. Identificación y Análisis de Bandas en Imágenes Digitales de Geles de Electroforesis. Trabajo de Grado. Universidad Experimental del Táchira. San Cristóbal, 2002.

Smith T. and Waterman M., 1981. Identification of common molecular subsequences. J. Mol. Biol., vol. 147, pp. 195–197.

Valens, C. The fast Lifting Wavelet Transform.

<http://www.perso.wanadoo.fr/polyvalens/clemens/lifting.html> (consultada en Mayo 2006)

Weidong C., Xiaoyan C., Xiurong Y., Erkang W. Electrophoresis, Vol. 24. pp: 3124–3130, 2003.

wxDevCpp. <http://www.wxdsn.sourceforge.net/> (Consultada Enero 2007).

wxWidgets. <http://www.wxwidgets.org> (Consultada en enero 2007).