

Optimización de paquetes de onditas para la clasificación de señales

Leandro D. Vignolo[†]
 Asesor: Diego H. Milone[‡]

Grupo de Investigación en Señales e Inteligencia Computacional,
 Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral

[†] ldvignolo@yahoo.com.ar, [‡] d.milone@ieee.org

Resumen— A partir del procesamiento basado en paquetes de onditas es posible obtener un conjunto redundante de coeficientes que caracterizan una señal. Para realizar una tarea de clasificación es necesario encontrar un subconjunto de estos coeficientes que resulte óptimo para distinguir entre las distintas clases de señales. Los métodos de computación evolutiva son útiles en la solución de este tipo problemas de optimización. En este trabajo se presenta un algoritmo genético con el cual se puede obtener el conjunto de coeficientes que permite discriminar mejor entre las diferentes clases de señales. Este algoritmo genético se utiliza en combinación con un método de cuantización vectorial con aprendizaje para determinar la bondad de cada posible solución. Los resultados para un conjunto de fonemas del español permiten apreciar las ventajas del método propuesto.

Palabras Clave— algoritmos genéticos, cuantización vectorial, paquetes de onditas, clasificación.

I. INTRODUCCIÓN

Los sistemas de reconocimiento automático del habla requieren de una etapa de preproceso que ponga en evidencia las características distintivas de cada fonema permitiendo mejorar los resultados (Rabiner and Juang, 1993). Para ello existen diferentes técnicas de preprocesamiento como el análisis espectral por bandas, los métodos de predicción lineal y el cepstrum (Rabiner and Schafer, 1978).

La transformada ondita es una herramienta con características especialmente útiles para el análisis de señales no estacionarias (Mallat, 1989). Para el caso discreto, el análisis multiresolución realizado por esta transformada puede ser implementado mediante una estructura de bancos de filtros (Vetterli and Herley, 1992). La *transformada paquetes de onditas* (TPO) (Hess-Nielsen and Wickerhouser, 1996) es una generalización de la descomposición realizada

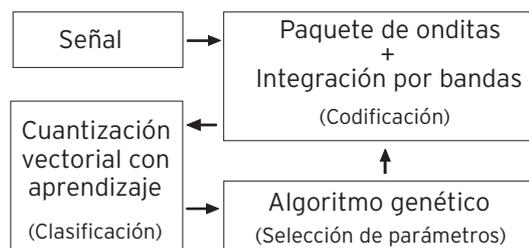


Figura 1: Esquema general del método propuesto

por la transformada ondita discreta que ofrece un rango más amplio de posibilidades para el análisis de señales. Para el cálculo de esta transformada se requiere la selección de una base ortogonal particular dentro de la familia de bases disponibles (o bancos de filtros). Sin embargo en aplicaciones de clasificación no resulta imprescindible trabajar con una base ortogonal. Si se quita entonces la restricción de ortogonalidad se obtiene en general un conjunto redundante de coeficientes, el cual es conveniente optimizar antes de utilizarlo para alimentar un clasificador.

A la proyección de la señal de interés en términos de *todos* los elementos disponibles dentro de la familia de bases paquetes de onditas se la denominará aquí como *transformada paquetes de onditas completa* (TPOC). El objetivo de este trabajo es presentar un método para la selección de las componentes de la TPOC más relevantes para la clasificación empleando un *algoritmo genético* (AG). Al método propuesto se lo denominará como *selección genética de coeficientes de la TPOC* (AG-TPOC). Para medir la bondad de las soluciones durante la búsqueda el AG utiliza un clasificador basado en la cuantización vectorial de la representación en términos de los coeficientes. El esquema del método propuesto se presenta en la Fig. 1.

Este artículo se organiza de la siguiente manera. A continuación se presenta una breve descripción de las propiedades de los AG, la transformada paquetes de onditas y los métodos de aprendizaje con cuantización vectorial. Luego se introduce el método propuesto para la selección de los componentes de la TPOC.

Posteriormente se presentan los resultados obtenidos y se plantean las conclusiones.

A. Transformada ondita y paquetes de onditas

Lo que hace interesantes a las bases de onditas es que, a diferencia de las que utilizan senos y cosenos, los elementos de la base se encuentran simultáneamente localizados en la frecuencia y en el tiempo. Esto resulta particularmente atractivo en el caso de las señales de voz que presentan comportamientos tanto estacionarios, como transitorios.

Una ondita puede definirse de forma simplificada como una función con valor medio igual a cero, norma unitaria y centrada en la vecindad de 0 (Mallat, 1999):

$$\psi(t) \in L^{\mathbb{R}}; \int_{-\infty}^{\infty} \psi(t)dt = 0; \|\psi(t)\| = 1 \quad (1)$$

A partir de ésta, se obtiene por escalado y traslación el átomo tiempo-frecuencia:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \quad (2)$$

Para calcular la transformada discreta, se evalúa en las escalas $s = a^j$ con $a = 2^{\frac{1}{v}}$, lo que hace que en cada intervalo $[2^j, 2^{j+1}]$ haya v valores intermedios. La función ondita resulta:

$$\psi_j[n] = \frac{1}{\sqrt{a^j}} \psi\left(\frac{n}{a^j}\right) \quad (3)$$

La transformada ondita discreta (TOD) será entonces:

$$Wf[n, a^j] = \sum_{m=0}^{N-1} f[m] \psi_j^*[m-n] \quad (4)$$

donde $a^j \in [2N^{-1}, K^{-1}]$ y K es el soporte de ψ (es distinta de 0 en el intervalo $[-\frac{K}{2}, \frac{K}{2}]$). Es importante notar que en la transformada (4) la función base no está especificada.

En el análisis multiresolución, basado en la transformada ondita discreta, una señal se descompone en aproximación y detalle. Luego la aproximación se vuelve a descomponer en un segundo nivel de aproximación y detalle, y el proceso se repite. Se puede considerar a la transformada paquetes de onditas como una extensión a este análisis multiresolución que permite mayor flexibilidad en la selección de las bandas de frecuencia. Utilizando el mismo razonamiento se pueden descomponer también los componentes de alta frecuencia (detalles) tal como en el caso de los componentes de baja frecuencia (aproximaciones). Esto produce lo que se llama árbol de descomposición de la TPO (Fig. 2).

Una vez seleccionado el árbol adecuado, el análisis mediante paquetes de onditas permite representar

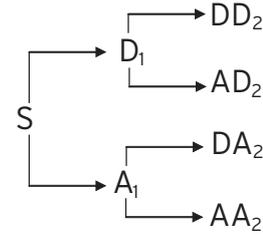


Figura 2: Árbol de descomposición de la TPO

la información en un plano tiempo-escala más flexible. Para la selección de este árbol es posible aprovechar conocimiento *a priori* de las características de una señal para obtener una representación eficiente en el dominio transformado. Por otro lado, la familia de árboles de descomposición mediante la transformada paquete de onditas ofrece una gran cantidad de combinaciones para descomponer una misma señal. Elegir una de éstas posibles combinaciones para una aplicación particular representa un problema interesante, el cual se resuelve comúnmente para el caso ortogonal utilizando diversos criterios para elegir la descomposición más conveniente. Para el caso de compresión de señales se utilizan criterios basados en la “entropía” de la energía normalizada, lo que se denomina como *mejor base ortogonal* (Coifman and Wickerhauser, 1992). Otra posibilidad, más cercana al problema considerado, es la denominada *base discriminante local* que proporciona una base ortogonal adecuada para la clasificación de señales (Saito, 1994). Estos criterios parten del supuesto de que se debe obtener una base ortogonal. Sin embargo, para el caso en estudio no existe una regla que asegure que una representación en términos de un conjunto de elementos redundantes no resulte más beneficiosa para la clasificación. Por esta razón se requiere un método que explore un rango más amplio de posibilidades.

B. Algoritmos genéticos

John Holland (1975) introdujo los algoritmos genéticos como un nuevo proceso de optimización que imita el proceso natural de evolución de los seres vivos. Esta técnica provee la flexibilidad y robustez necesarias para encontrar soluciones en espacios de búsqueda complejos (Goldberg, 1989).

Usualmente un AG simple consiste en tres operadores: selección, operación genética y reemplazo (Tang *et al.*, 1996). La población está formada por un grupo de individuos cuya información se codifica en cromosomas, y de ella se eligen los candidatos para la solución de un problema. El desempeño de cada individuo dentro de la población se mide mediante un valor de aptitud que se obtiene calculando la función objetivo en base a la información genética decodificada (fenotipo). Esta función es la que simula a la presión selectiva del ambiente en la evolución

natural.

Un grupo particular de individuos (padres) se selecciona para generar la descendencia mediante las operaciones genéticas de cruzamiento y mutación. La población actual es luego reemplazada por su descendencia, para lo cual existen diferentes estrategias. El ciclo de un AG es repetido hasta que se alcanza un criterio de finalización deseado (por ej: número máximo de generaciones, valor máximo de aptitud, etc). Luego del proceso de evolución, el mejor individuo de la población final puede resultar en la solución buscada para el problema.

C. Aprendizaje por cuantización vectorial

El aprendizaje por cuantización vectorial supone una extensión supervisada del aprendizaje competitivo en la cual los prototipos están etiquetados (Kohonen, 1995). Por ésto, además de considerar la cercanía de un prototipo se puede evaluar la clase de éste e imponer, por lo tanto, correcciones de premio (acercamiento) o castigo (alejamiento). El objetivo final puede resumirse como sigue: dado un conjunto de entrenamiento original, se trata de construir un conjunto de referencia mediante la técnica de cuantización vectorial con aprendizaje (LVQ: del inglés *Learning Vector Quantization*); una vez construido, los patrones a clasificar se etiquetarán utilizando la regla del vecino más proximo (1-NN, del inglés *Nearest Neighbour*) con el conjunto de referencia.

El proceso de construcción del conjunto de referencia puede descomponerse en dos pasos:

1. Inicialización del conjunto de prototipos:

$$\mathbf{m}_i(0), i = 1, 2 \dots N_p \quad (5)$$

donde N_p es el tamaño del conjunto. El procedimiento recomendado es la selección de prototipos con ciertas restricciones que aseguren que éstos se encuentran dentro del agrupamiento correspondiente a su clase. Esta estrategia de inicialización acelera la convergencia del aprendizaje.

2. Aprendizaje o corrección de los prototipos. Consiste en actualizar de forma iterativa los prototipos del conjunto de referencia actual en cada paso hasta que se consiga estabilidad o se cumpla algún criterio de convergencia adicional. Esto significa para cada tiempo t se corrige el prototipo $\mathbf{m}_c(t)$ más cercano al patrón $\mathbf{x}(t)$ según la expresión:

$$\mathbf{m}_c(t+1) = \mathbf{m}_c(t) + s(t)\alpha(t)[\mathbf{x}(t) - \mathbf{m}_c(t)] \quad (6)$$

donde $\alpha(t)$ es la velocidad de aprendizaje; $s(t)$ es igual a +1 cuando el patrón $\mathbf{x}(t)$ pertenece a la misma clase que el prototipo $\mathbf{m}_c(t)$ y -1 cuando pertenece a una clase diferente.

Algoritmo LVQ optimizado

En este caso cada prototipo del conjunto de referencia tiene su propia velocidad de aprendizaje, por lo que la función común $\alpha(t)$ se sustituye por N_p funciones $\alpha_i(t)$ para $i = 1, 2 \dots N_p$. Para cada paso de corrección, se actualiza $\alpha_c(t)$ de mediante la expresión:

$$\alpha_c(t) = [\alpha_c(t-1)]/[1 + s(t)\alpha_c(t-1)] \quad (7)$$

Teniendo en cuenta la desigualdad $0 < \alpha_i(t) < 1$ puede verse que $s(t)$ hace que $\alpha_c(t)$ decrezca cuando se clasifica correctamente y crezca cuando se clasifica incorrectamente.

II. DESARROLLO

A. Materiales

Los archivos de voz utilizados en este trabajo son un subconjunto de 600 frases de la base de datos Albayzyn (Moreno *et al.*, 1993). La base de datos fue segmentada fonéticamente con un reconocedor basado en *modelos ocultos de Markov*, por lo que para cada uno de los fonemas grabados, se dispone de la información de las muestras en que comienza y finaliza dentro del archivo de voz junto a la etiqueta correspondiente.

Se procesaron los archivos de voz utilizando una implementación de la TPO que fue diseñada en base a los algoritmos correspondientes a la TOD (Press *et al.*, 1992). Se trabajó con segmentos de 256 muestras para cada fonema, lo que implica que cada uno de los segmentos analizados representa 32 msec (para una frecuencia de muestreo de 8 kHz). Éste es un tamaño de ventana frecuentemente utilizado en procesamiento de voz. Para todos los casos se escogieron 128 muestras (16 msec) a cada lado del punto medio del fonema.

Se aplicaron los filtros correspondientes a la TPO sucesivamente decimando la señal hasta obtener seis niveles de descomposición. Teniendo en cuenta que se trabajó con señales de 256 muestras, esto significa que en último nivel del árbol se tienen 64 nodos de 4 muestras cada uno. Lo que implica que el árbol binario de paquetes de onditas tiene 127 nodos y 1792 coeficientes.

Para reducir las dimensiones del espacio de búsqueda se realiza una integración por bandas de los coeficientes de la transformada. Esta operación consiste formar grupos de coeficientes dentro de cada banda y sumar sus cuadrados. De esta forma, resulta un coeficiente de energía por cada grupo que se defina. En la Tabla 1 se puede observar el esquema utilizado en este trabajo.

Los valores resultantes de esta integración por bandas son luego normalizados de forma tal que todos los coeficientes de energía de un patrón dado estén entre 0 y 1. Esto es, si $w_p[k]$ es uno de los coeficien-

tes de energía del patrón p , entonces el coeficiente normalizado según este criterio será:

$$\hat{w}_p[k] = w_p[k] / \max_{\forall i} \{w_i[k]\} \quad (8)$$

Una vez obtenidos los valores normalizados de los coeficientes de la transformada se procede a crear los archivos de entrenamiento y validación para utilizar con el clasificador.

B. Diseño de la función de aptitud

La función de aptitud del AG propuesto consiste en un clasificador que permite evaluar los individuos proporcionando como resultado un porcentaje de reconocimiento, que se utiliza como valor aptitud de los mismos. El método de clasificación utilizado está basado en el método de cuantización vectorial con aprendizaje y fue elegido principalmente por razones de costo computacional.

La función de aptitud debería poder darnos una medida de lo bien que separa las diferentes clases la representación que la composición del cromosoma ha sugerido.

Considerando que se cuenta con los valores de los coeficientes de la transformada paquetes de onditas para cada fonema y las etiquetas correspondientes, se puede utilizar un método de cuantización vectorial con aprendizaje que permita conocer lo bien que se pueden clasificar las señales con cada uno de los distintos conjuntos de coeficientes.

C. Implementación de método AG-TPOC

El método propuesto consiste en un modelo de evolución simple con cromosomas binarios. Más precisamente, se trata de un algoritmo genético en el cual cada individuo representa una combinación diferente de los elementos provenientes de la integración por bandas de los coeficientes de la transformada paquetes de onditas.

Cada gen o alelo que compone un cromosoma representa uno de los 208 coeficientes. Si posee un valor 0 indica que éste elemento no debe tenerse en cuenta para representar una señal. Por el contrario, un valor 1 indica que sí debe tenerse en cuenta. Para la inicialización se completa cada una de las posiciones de cada cromosoma de la población con valores 0 o 1 elegidos aleatoriamente. La representación puede ser redundante, por lo que no existen restricciones respecto a qué elementos existen. Para la selección de los individuos más aptos se utiliza el método de la ruleta.

Para optimizar la función de aptitud, la parametrización de las señales se realiza sólo una vez y en los archivos de entrenamiento y validación se almacenan todos los coeficientes. En el momento de evaluar un individuo, sólo se consideran los coeficientes que indique el respectivo cromosoma.

Los individuos se deben escoger de manera que las combinaciones de coeficientes que representen permi-

Nivel	Nodos	Grupos por nodo	Coefficientes por nivel
1	2	8	16
2	4	8	32
3	8	4	32
4	16	2	32
5	32	1	32
6	64	1	64

Total de coeficientes de energía: 208

Tabla 1: Esquema de integración aplicado al árbol de paquetes de onditas para una señal de 256 muestras.

Fonema	AG-TPOC	TOD	AG-TPOC	TOD
a	88.89 %	28.0 %	57.69 %	30.3 %
e	55.56 %	36.0 %	53.85 %	6.06 %
i	88.89 %	92.0 %	92.31 %	66.67 %
o	77.78 %	76.0 %	61.54 %	48.48 %
u	100 %	38.0 %	53.85 %	66.67 %
t	-	-	0.00 %	72.73 %
b	-	-	92.31 %	66.67 %
d	-	-	0.00 %	12.12 %
p	-	-	84.61 %	9.09 %
Total	82.22 %	54.0 %	55.12 %	42.09 %

Tabla 2: Porcentajes de acierto para los dos experimentos realizados.

tan separar lo mejor posible las diferentes clases de patrones correspondientes a distintos fonemas. Con este fin se propone la utilización del método LVQ Optimizado como función objetivo, tomando como valor de aptitud el resultado de la clasificación. Para esto se particionan los datos en dos grupos, uno para entrenamiento y otro diferente para realizar la validación de los resultados del clasificador.

Las implementaciones del AG, el clasificador y la TPOC se realizaron en el lenguaje C++.

III. RESULTADOS

Para la elección de la familia de onditas no existen criterios claros que destaquen una frente a otras en una aplicación particular (Rufiner, 1996). Por esta razón se decidió experimentar con las más difundidas, entre las cuales podemos citar las de Meyer, Daubechies, Symmlets, Coiflets y Splines (Daubechies, 1992). En base a los experimentos realizados con las diferentes familias se optó por utilizar la familia de onditas Coiflet por ser ésta la que permitió obtener mejores resultados en la mayoría de los casos.

En el algoritmo genético se utilizó el método de la ruleta para la selección de los individuos en cada generación y se empleó la estrategia de reemplazo total con elitismo en la cual se conserva el mejor individuo para la generación siguiente. Se implementaron las operaciones genéticas de mutación y cruza sim-

	t	b	d	o	u	p	a	e	i
t	0.00	11.54	0.00	0.00	3.84	76.92	0.00	0.00	7.69
b	0.00	92.31	0.00	3.84	0.00	3.84	0.00	0.00	0.00
d	0.00	15.38	0.00	19.23	3.84	11.54	7.69	30.77	11.54
o	0.00	3.85	0.00	61.54	11.53	0.00	11.54	11.54	0.00
u	0.00	19.23	0.00	15.38	53.85	0.00	3.85	0.00	7.69
p	0.00	11.54	0.00	0.00	0.00	84.61	0.00	0.00	3.85
a	0.00	3.85	0.00	30.77	0.00	0.00	57.69	7.69	0.00
e	0.00	0.00	0.00	23.08	0.00	3.85	0.00	53.85	19.23
i	0.00	3.85	0.00	0.00	0.00	0.00	0.00	3.85	92.31

Tabla 3: Matriz de confusión para la mejor solución encontrada con el método propuesto AG-TPOC.

ple. El tamaño de la población fue de 100 individuos y las probabilidades de cruce y mutación fueron de 0.9 y 0.05 respectivamente.

El primer experimento se realizó sólo con las 5 vocales del idioma español. En este caso se utilizó un conjunto de referencia (codebook) de 50 vectores, es decir, de aproximadamente 10 vectores para cada fonema. Para el entrenamiento del clasificador en la evaluación de cada individuo se realizaron 8 épocas de entrenamiento empleando 415 patrones y una validación cruzada con otros 45 patrones. La razón de aprendizaje inicial para cada prototipo se ajustó en 0,05 habiendo comprobado que este valor permite la estabilización de los vectores del conjunto de referencia. La mejor solución se obtuvo luego de 11 generaciones del AG y el porcentaje total de reconocimiento conseguido a partir de ésta fue de 82.22%. En la Tabla 2 se pueden apreciar los porcentajes de acierto para cada vocal.

En el segundo experimento realizado se incluyeron, además de las 5 vocales, los fonemas /b/, /d/, /p/ y /t/. Éstos cuatro fonemas pertenecen a la clase de los oclusivos (Quilis, 1993) y se eligieron porque sus similares características los hacen particularmente difíciles de clasificar. Por ejemplo, la única diferencia que existe entre el fonema /p/ que corresponde a una consonante bilabial sorda y el fonema /b/ que corresponde a una consonante bilabial sonora, es su modo de articulación.

En el caso de los 9 fonemas se utilizó un codebook de 144 vectores (aproximadamente 16 vectores por fonema) y para el entrenamiento se realizaron 5 épocas empleando 1548 patrones. Para efectuar la validación se utilizaron 234 patrones. La mejor solución, que fue encontrada luego de 21 generaciones, permitió obtener un reconocimiento total de 55.12%. En la Tabla 3 se pueden apreciar los porcentajes de acierto para cada fonema y también los porcentajes de equivocación entre cada uno de ellos.

Para comparar los resultados se realizaron experimentos de clasificación de los mismos grupos fonéticos pero parametrizando la señal con la TOD y utilizando los mismos valores para los parámetros del entrenamiento y el clasificador. Como se puede apre-

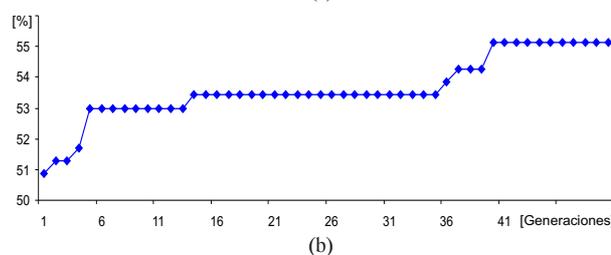
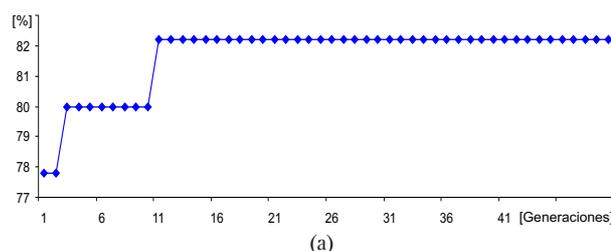


Figura 3: Porcentaje de acierto en función del nro. de generaciones. (a) Vocales. (b) Fonemas /a/, /e/, /i/, /o/, /u/, /b/, /d/, /p/ y /t/.

ciar en la Tabla 2, para ambos casos la parametrización encontrada utilizando el AG-TPOC permite obtener resultados de clasificación considerablemente mejores.

IV. DISCUSIÓN

En la Tabla 3 se puede apreciar que el fonema /t/, en la mayoría de los casos (76.92%) fue clasificado como /p/. Este error es razonable porque los fonemas plosivos tienen sus características espectrales más distintivas en el comienzo (la plosión) y en los experimentos realizados las muestras se tomaron de la parte central de los mismos. Otra solución consiste en distinguirlos por su contexto (que no se incluyó en estas pruebas). Para el caso del fonema /d/ se tiene un problema similar.

Como se puede observar en la Fig. 3 el algoritmo genético converge en un reducido número de generaciones tanto para el caso de las vocales como para el caso de los 9 fonemas estudiados. Se han realizado experimentos de más de 250 generaciones y en todos los casos el valor máximo de aptitud se mantuvo estable a partir de las primeras 50. Existe la posibilidad

de que con un número mucho mayor de generaciones el AG encuentre un individuo con un valor de aptitud mayor. Sin embargo, el costo computacional que representa la evaluación de cada individuo no permite realizar tales experimentos en una computadora personal. Otras pruebas posibles de realizar podrían consistir por ejemplo en variar parámetros como la probabilidad de mutación.

Debe agregarse que las representaciones obtenidas varían en cada experimento y no presentan una estructura simple de analizar a partir de una gráfica, por lo que no se han incluido en este trabajo. En trabajos futuros se evaluará la posibilidad de obtener conclusiones generales acerca de la estructura de la representación óptima obtenida y en relación con las características de la señal utilizada.

V. CONCLUSIÓN Y TRABAJOS FUTUROS

En este trabajo se ha presentado una técnica que permite seleccionar los componentes relevantes de la transformada paquete de ondas completa con el objetivo de clasificar señales. De esta manera se puede encontrar una representación "óptima" que permite alimentar un sistema de reconocimiento automático del habla, como alternativa al enfoque clásico basado en la TOD.

Se han implementado los algoritmos correspondientes y se ha validado dicha implementación. La comparación con una parametrización ampliamente utilizada muestra que este método cumple con el objetivo de encontrar una buena representación permitiendo obtener muy buenos resultados en la tarea de clasificación de fonemas. Esto se debe a que la representación encontrada permite separar eficientemente las diferentes clases de señales logrando que la tarea del clasificador sea más sencilla.

En trabajos futuros se realizarán los ajustes necesarios para la utilización de la mejor representación obtenida en un sistema de reconocimiento automático del habla basado en modelos ocultos de Markov y se considerará la realización de pruebas de validación cruzada con ruido.

Otro trabajo pendiente consiste en realizar una implementación del AG para ejecutarlo en computadoras paralelas, de manera que la población se distribuya entre varios procesadores para que se evalúen simultáneamente reduciendo críticamente el tiempo de proceso. Esto permitiría además utilizar funciones de aptitud más complejas y directamente relacionadas con las utilizadas en sistemas de reconocimiento del habla.

REFERENCIAS

Coifman, R. and M. V. Wickerhauser (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory* **38**(2), 713–718.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics. Philadelphia, PA, USA.

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.

Hess-Nielsen, N. and M.V. Wickerhauser (1996). Wavelets and time-frequency analysis. *Proceedings of the IEEE* **84**(4), 523–540.

Holland, J.H. (1975). *Adaptation in Natural and Artificial System*. University of Michigan Press. Ann Arbor.

Kohonen, T. (1995). *The Self-Organizing Map*. Springer-Verlag. New York.

Mallat, S.G. (1989). A theory of multiresolution of signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.* **11**(7), 674–693.

Mallat, S.G. (1999). *A Wavelet Tour of signal Processing*. second ed.. Academic Press.

Moreno, A., D. Poch, A. Bonafonte, E. Lleida, J. Llis-terri, J. Mariño and C.Ñadeu (1993). Albayzin speech database design of the phonetic corpus. Technical report. Universitat Politècnica de Catalunya (UPC), Dpto. DTSC.

Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery (1992). *Numerical Recipes in C, The Art of Scientific Computing*. second ed.. Cambridge University Press.

Quilis, A. (1993). *Tratado de Fonología y Fonética Españolas*. Biblioteca Románica Hispánica. Editorial Gredos. Madrid.

Rabiner, L.R. and B.H. Juang (1993). *Fundamentals of Speech Recognition*. Prentice Hall. NJ.

Rabiner, L.R. and R.W. Schafer (1978). *Digital Processing of Speech Signals*. Prentice Hall. NJ.

Rufiner, H. (1996). Comparación entre análisis ondas y fourier aplicados al reconocimiento automático del habla. Master's thesis. Universidad Autónoma Metropolitana. Iztapalapa.

Saito, N. (1994). Local feature extraction and its applications using a library of bases. PhD thesis. Yale University. New Haven, USA. Director-Ronald R. Coifman.

Tang, K.S., K.F. Man, S. Kwong and Q. He (1996). Genetic algorithms and their applications. *IEEE Signal Processing* **13**(6), 22–29.

Vetterli, M. and C. Herley (1992). Wavelets and filter banks: Theory and design. *IEEE Trans. Signal Proc.* **40**(10), 2207–2232.