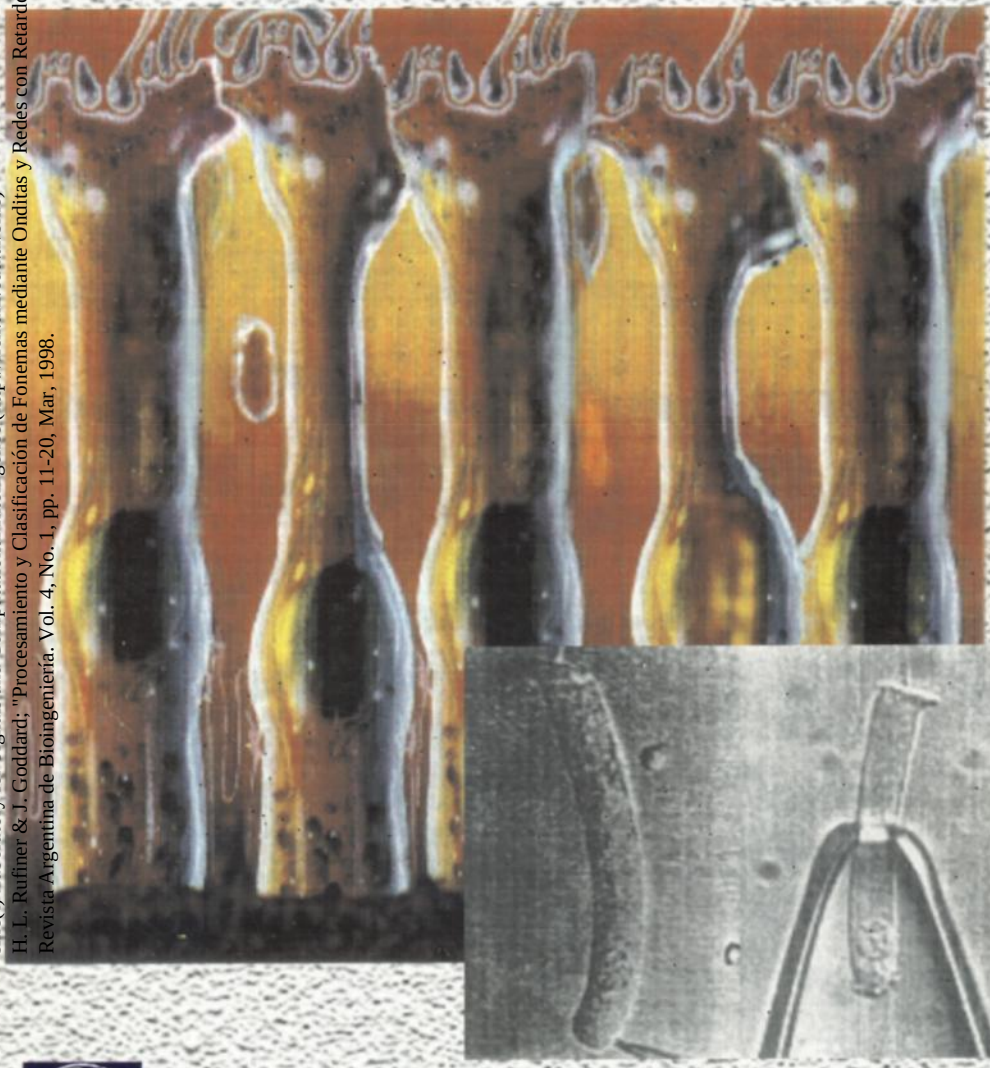
ISSN 0329-5257

Revista Argentina de Bioingeniería

Volumen 4; Nro 1, Marzo 1998



Sociedad Argentina de Bioingeniería

Revista Argentina de Bioingeniería

Marzo1998. Volumen. Número 1

Pág. Contenido

Procesamiento y Clasificación de Fonemas mediante Onditas y Redes con Retardos

Hugo Leonardo Rufiner(*), John Goddard(**)

(*)UNER, Laboratorio de Cibernética, Ruta 11 Km 10, Paraná, (3100), Entre Ríos, Argentina, lrufiner@alpha.arcride.edu.ar
(**)UAMI, Departamento de Ing. Eléctrica, Av. Michoacán y La Purisima S/N, (09340), México D.F., México, jgc@xanum.uam.mx.

Resumen

Mientras que las onditas han sido aplicadas a una variedad de problemas relacionados con el procesamiento de señales individuales, como compresión y filtrado, relativamente poco se ha hecho en casos que involucran la clasificación de varias señales. El presente trabajo tiene el fin de estudiar el procesamiento mediante onditas de ciertos fonemas del habla y su posterior clasificación con redes neuronales con retardos temporales. Los fonemas fueron extraídos de la base de datos TIMIT que ha sido especialmente construida para este tipo de problemas..

Palabras Claves: • Onditas • Redes Neuronales con Retardos • Clasificación • Reconocimiento Automático del Habla

Introducción

En poco más de diez años de existencia, el área de las onditas (en inglés *wavelets*) ha llegado a ser de suma importancia para el procesamiento de señales. Esto se debe en gran parte a su manera natural de tratar a las señales no-estacionarias. En lugar del análisis tradicional basado en la transformada de Fourier, que examina una señal a una resolución fija, la transformada de onditas posee la característica de hacerlo a distintas escalas (ó resoluciones). Esto implica un análisis más similar al realizado por los sistemas sensoriales biológicos, en particular análogo al caso del oído. En una variedad de investigaciones se han encontrado beneficios con este tipo de transformadas para tareas tales como la compresión y el filtrado de señales[1]. Sin embargo se ha hecho relativamente poco en materia de clasificación de patrones dinámicos de longitud variable, como es el caso de la clasificación de los fonemas del habla. Esto representa una tarea diferente debido a la necesidad de procesar un gran número de señales con una sola familia de onditas. Más aún, se debe tomar en cuenta el papel del tipo de clasificador utilizado.

En el presente trabajo, se estudia la aplicación de las onditas al procesamiento de señales de voz, para su posterior clasificación en fonemas. La familia de onditas utilizadas esta basada en las denominadas Symmlets[2]. Los fonemas han sido extraídos de la base de datos TIMIT, que ha sido especialmente diseñada con el propósito de crear y probar sistemas de reconocimiento automático del habla. Posteriormente se aplica una red neuronal con retardos para realizar la clasificación de los patrones generados mediante las onditas.

Se puede considerar que un sistema típico de reconocimiento automático del habla esta compuesto por las etapas mostradas en la *Figura 1*. Este trabajo esta orientado principlamente a contestar algunas preguntas acerca de la utilización de Onditas en la etapa de Preproceso.

La organización del presente trabajo es la siguiente. En primer término se describe la base de datos y los fonemas elegidos para el procesamiento y la clasificación. En las dos secciones siguientes se presentan los principales resultados sobre onditas, la transformada discreta de onditas (TDO), y las redes con retardos que se requieren para una mejor comprensión del artículo. A continuación se presentan los detalles sobre la forma de procesar los fonemas y las arquitecturas de las redes con retardos utilizadas. La sección siguiente presenta los resultados obtenidos en la clasificación y en luego se discuten estos resultados desde le punto de vista del procesamiento y la clasificación. Finalmente se presentan las conclusiones relacionadas con el trabajo.

Este trabajo esta basado parcialmente en la tesis de maestría del primer autor[3].

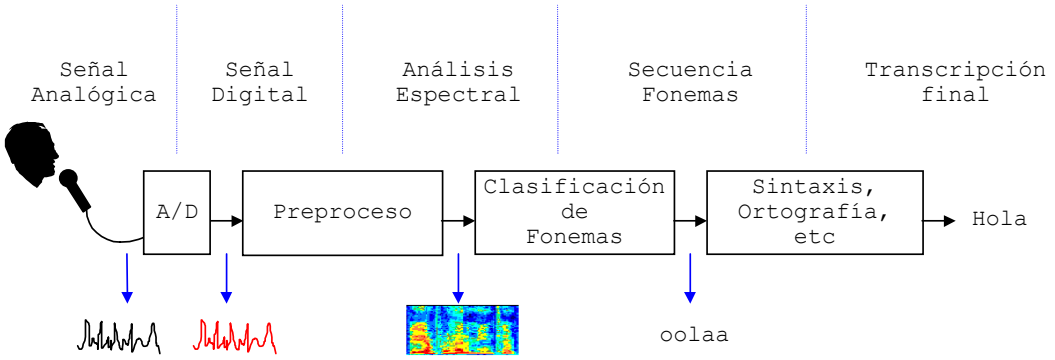


Figura 1: Componentes de un Sistema de Reconocimiento.

Los Datos

Los fonemas escogidos en este trabajo fueron tomados de la base de datos TIMIT. Este *corpus* fue construido especialmente por Texas Instruments y el Massachusetts Institute of Technology para realizar experimentos con sistemas de reconocimiento automático del habla[4]. Consiste en una serie de emisiones de voz grabadas a través de la lectura de diversos textos en inglés por un conjunto de casi 600 hablantes. TIMIT contiene un total de 6300 oraciones, 70% de los hablantes son masculinos y el 30% restante son femeninos, totalizando unos 650 Mbytes de información. Los datos fueron digitalizados a 16 KHz, 16 bits por muestra. Por el tamaño de TIMIT se enfocó el trabajo sobre un subconjunto del total de fonemas. Parte de los fonemas escogidos correspondieron a la primera consonante del conjunto denominado E-set (TI-46), conocido por su dificultad de clasificación mediante técnicas automáticas. A estos se agregaron dos vocales eligidas por su cercanía en el espacio de formantes. De esta manera el conjunto final de entrenamiento estuvo formado por los fonemas del inglés /b/, /d/, /jh/, /eh/, y /ih/ del total de los hablantes de la región 1. En la ***Figura 2*** se puede apreciar una emisión típica de la base de datos con las etiquetas correspondientes a palabras y fonemas.

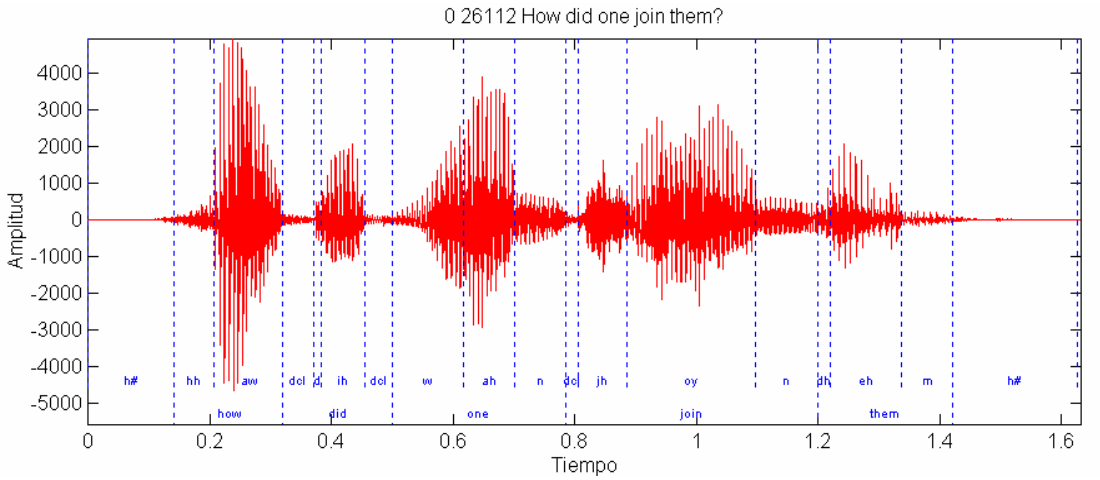


Figura 2: Señal de voz con etiquetas de palabras y fonemas

Las Onditas

El área de onditas empezó a desarrollarse a mediados de los años 80´s con el trabajo de Meyer[5]. Desde entonces ha demostrado ser una herramienta importante para el procesamiento de señales debido a su manera natural de analizar señales con discontinuidades y picos (transitorios). En esta sección mencionamos los principales resultados, para onditas ortonormales de una dimensión, necesarios para el desarrollo posterior. Excelentes referencias son Daubechies [2], Wojtaszczyk [6], y Mallat [7, 8].

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. L. Rufiner & J. Goddard; "Procesamiento y Clasificación de Fonemas mediante Onditas y Redes con Retardos"
Revista Argentina de Bioingeniería. Vol. 4, No. 1, pp. 11-20, Mar, 1998.

Una ondita puede definirse de la siguiente forma:

Definición 1: Una ondita es una función ψ en $L^2(\mathbf{R})$ que tiene la propiedad de que la familia de funciones $\psi_{j,k}(x) := 2^{j/2}\psi(2^j x-k)$ para j,k $\in$ Z, es una base ortonormal del espacio de Hilbert $L^2(\mathbf{R})$.

La definición habla sobre la existencia de una *sola* función cuyas traslaciones y dilataciones apropiadas forman una base ortonormal. La ondita más sencilla es la de Haar que tiene el valor 1 en el intervalo [0,1/2), -1 en el intervalo [1/2,1], y 0 para otros valores reales.

Aunque no es obvio, generalmente las 'buenas' onditas (por tener propiedades adicionales como regularidad) se contruyen a través de un Análisis Multiresolución (AMR), definido de la siguiente manera:

Definición 2: Una AMR es una secuencia $(Vj)_{j\in Z}$ de subespacios de $L^2(\mathbf{R})$ tal que:

1. $V_j \subset V_{j+1}$ para cualquier j$\in$Z
2. $\cup V_j$ es denso en $L^2(\mathbf{R})$ y $\cap V_j = \{0\}$
3. $\in V_j$ si y solo si $f(2^{-j}t)\in V_0$ para cualquier j$\in$Z
4. Existe una función $\phi \in V_0$, llamada la función de escala, tal que la familia $\{\phi(t-m)\}_{m\in Z}$ es una base ortonormal para V_0

Un ejemplo de una función de escala (de Haar) es la función característica del intervalo [0,1]. Por la definición 2 podemos aproximar cualquier función o señal en $L^2(\mathbf{R})$ por una función en alguno de los V_j. Se dice que es una aproximación a la resolución o escala j. Como $\{2^{j/2}\phi(2^j t-m)\}_{m\in Z}$ es una base ortonormal para V_j, esto da la posibilidad de analizar una señal con 'ventanas' de diferentes tamaños, a diferencia del análisis de Fourier. Las condiciones sobre los subespacios V_j implican que por estar en el espacio de Hilbert $L^2(\mathbf{R})$ existen subespacios $(W_j)_{m\in Z}$ tal que cada V_{j+1} es la suma directa de V_j con W_j. Las W_j tienen la interpretación de que representan la información de detalle que se requiere cuando se pasa de una aproximación a la resolución j, a una aproximación a la resolución j+1. En este caso $V_{j+1} = V_j \oplus W_j$, y continuando el proceso obtenemos $V_{j+1} = V_k \oplus W_k \oplus W_{k+1} \oplus \ldots\ldots \oplus W_{j-1} \oplus W_j$. Entonces una aproximación a una resolución más alta puede ser representada en términos de una resolución más baja con los detalles adicionales. Además, $L^2(\mathbf{R})$ es la suma directa de los W_j.

Si ϕ es la función de escala, entonces $\phi\in V_1$ y $\phi(t/2)\in V_0$, y por las condiciones de la definición 2:

$$\phi(t/2) = 2\sum_n h_n\phi(t-n)$$

de manera equivalente,

$$\phi(t) = 2\sum_n h_n\phi(2t-n) \qquad (3)$$

donde $h_n = 1/2\langle\phi, \phi_{1,n}\rangle = 1/2\int \phi(t)\phi^*(2t-n)dt$ y $\phi^*(t)$ es el conjugado de $\phi(t)$.

A (3) se la llama ecuación de escala. Para el caso de Haar, tendremos la siguiente ecuación:

$$\phi(t) = \phi(2t) - \phi(2t-1)$$

De la ecuación de escala obtenemos la siguiente ecuación tomando las transformadas de Fourier:

$$\Phi(\varsigma) = H(\varsigma/2)\Phi(\varsigma/2)$$

donde $H(\varsigma) = 2\Sigma\, h_n e^{-i\varsigma}$, que es una función 2π periódica. Resulta que H es un filtro pasabajos que además determina a la función de escala. En el caso de Haar, $H(\varsigma) = \frac{1}{2}(1+ e^{-i\varsigma})$.

Con esta notación, se presenta el siguiente teorema (c.f. Theorem 2.20 [2]) de la existencia de onditas asociadas con un AMR:

Teorema 4: Supóngase que $(Vj)_{j\in Z}$ es un AMR con la función de escala $\phi\in V_0$. Entonces la función $\psi\in W_0$ es una ondita si y sólo si $\Psi(\varsigma) = e^{i\varsigma/2}v(\varsigma)H^*(\varsigma/2+\pi)\Phi(\varsigma/2)$ para alguna función 2π-periódica $v(\varsigma)$ con $|v(\varsigma)| = 1$ en casi todas partes. Cada ondita ψ tiene la propiedad que $\{\psi_{j,n}\}_{n\in Z}$ es una base ortonormal de Wj.

Si queremos construir una ondita utilizando el teorema, podemos tomar $v(\varsigma) = 1$, dando

$$\psi(t) = 2\sum_n g_n\phi(2t-n)$$

Definiendo $G(\varsigma) = 2\sum_n g_n e^{-i\varsigma n}$ resulta en que $g_n = (-1)^n h_{1-n}$ y $G(\varsigma) = -e^{-i\varsigma} H(\varsigma+\pi)$

G es un filtro pasaaltos, y junto con H forman una pareja de filtros de cuadratura espejo.

Generalmente no existen formas analíticas cerradas para las onditas, y para describir la TDO de una señal se pueden derivar las ecuaciones de análisis y síntesis sin necesidad de escribir explícitamente la ondita [9]. Sea f(t) $\in V_k$. Para el caso de la descomposición de la señal f(t), como $V_k = V_{k-1} \oplus W_{k-1}$, se puede escribir f(t) de las siguientes maneras:

$$\sum_n c_{k,n}\phi_{k,n}(t) \text{ y } \sum_n c_{s,n}\phi_{s,n}(t) + \sum_{m=s}^{k-1}\sum_n d_{m,n}\psi_{m,n}(t) \quad \text{para } k>s.$$

De allí se puede derivar de que:

$$c_{k-1,n} = \sqrt{2}\sum_i h_{i-2n}c_{k,i} \text{ y } d_{k-1,n} = \sqrt{2}\sum_i g_{i-2n}c_{k,i}$$

Para la reconstrucción de la señal tenemos que:

$$c_{m,n} = \sqrt{2}\left(\sum_i h_{n-2i}c_{m-1,i} + \sum_i g_{n-2i}d_{m-1,i}\right)$$

En los dos casos se puede lograr el análisis y síntesis por bancos de filtros. En la *Figura 3* se muestra el esquema que corresponde a la descomposición o análisis.

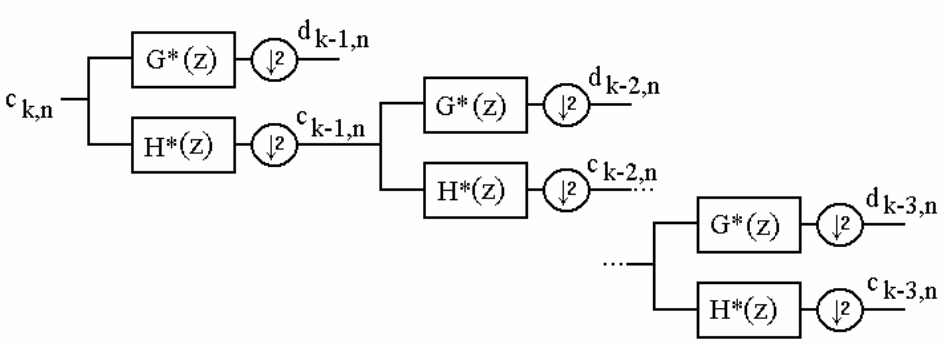


Figura 3 : Algoritmo de la TDO.

Es importante contar con filtros de repuesta finita al impulso (FIR). En este caso H y G tendrán un número finito de coeficientes distintos de cero. Esto corresponde a que la función de escala, y por ende la ondita, tienen soporte compacto. En Daubechies [2], se describen ejemplos donde esto ocurre, aunque también se muestra que la única ondita que además es simétrica es la de Haar. En el presente trabajo utilizamos una ondita de la familia llamada Symmlets, que es la ondita de soporte compacto con la menor asimetría posible.

La familia particular depende de la elección de un parámetro N que determina el número de Momentos de Desvanecimiento (Vanishing Moments):

$$\int_{-\infty}^{\infty} t^k \psi_1(t)\,dt = 0, \quad k = 0,1,\ldots,N$$

Este parámetro permite modificar la localización tiempo-frecuencia de la Ondita, lo que puede observarse a través de su evolución temporal y su espectro (que deben estar bien concentrados alrededor de un punto del dominio). Esto es también equivalente al desvanecimiento a alto orden de la transformada de Fourier en el origen :

$$\left(\frac{d}{d\xi}\right)^k \hat{\psi}_1(0) = 0, \quad k = 0,1,\ldots,N$$

lo que implica una débil forma de localización en la frecuencia debido a que la transformada de Fourier de $\psi_1(2^j t - k)$ esta mayormente concentrada alrededor de valores de $|\xi|$ del orden de 2^j.

En la *Figura 4* se observa la gráfica de varias onditas Symmlet individuales a distintas escalas y en diferentes localizaciones (con 6 momentos desvanecientes).

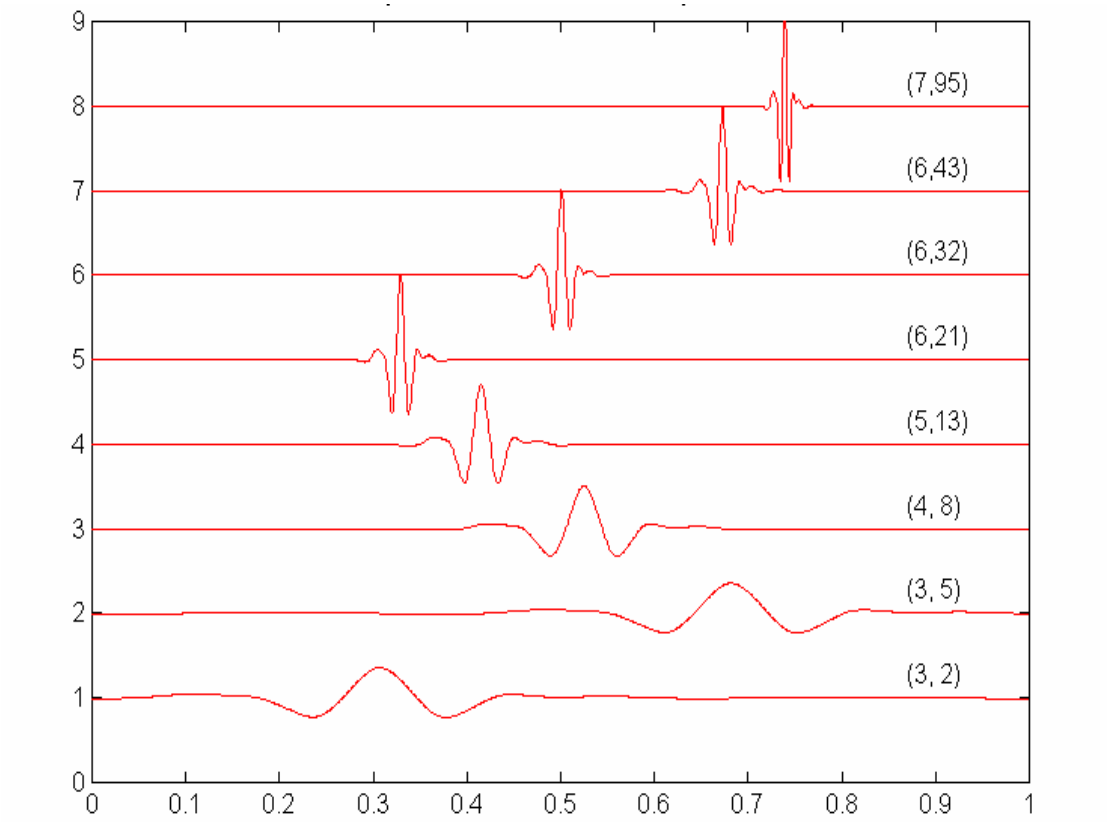


Figura 4: Onditas Symmlets a distintas escalas y localizaciones

Las Redes con Retardos

Existen muchas técnicas de clasificación que se han aplicado al caso de los fonemas. Entre ellas podríamos mencionar los Modelos Ocultos de Markov [10], las redes con retardos [11], las redes recurrentes [12], los árboles de redes neuronales [13], y otros clasificadores híbridos [14]. En este trabajo se utiliza una red con retardos con diferentes arquitecturas para realizar una clasificación de los fonemas escogidos. Una razón para elegir este tipo de clasificador es que permite la utilización del clásico algoritmo de retropropagación casi sin modificaciones. Por otro lado su potencia es adecuada para la clasificación de una secuencia en el tiempo de patrones, como se presentan con los *frames* asociados con cada fonema. En pruebas de comparación [3] funcionó mejor que otras arquitecturas de redes recurrentes simples como las de Jordan y Elman [15].

Las redes neuronales con retardos consisten en unidades elementales similares a las de un perceptron, pero modificadas a fin de que puedan procesar información generada en distintos instantes. A las entradas sin retardos de cada neurona, se les agregan las entradas correspondientes a instantes distintos. De este modo, una unidad neuronal de estas características es capaz de relacionar y procesar conjuntamente la entrada actual con eventos anteriores.

Esta arquitectura es equivalente a una red neuronal anteroalimentada con una capa oculta que recibe un patrón espacial con (m+1) dimensiones x generado por un línea de retardo a partir de la secuencia temporal. Esto es, si los valores deseados para la salida son especificados para distintos tiempos t, entonces se puede utilizar el algoritmo de retropropagación para entrenar esta red como un reconocedor de secuencias.

Este tipo de redes ha sido aplicado exitosamente al reconocimiento de voz [16, 17, 18, 19]. En las TDNN de Waibel se incluye generalmente una capa oculta extra (también con retardos) y se entrenan por el algoritmo de retropropagación a través del tiempo. De esta forma cada unidad neuronal con retardos posee la capacidad de extraer relaciones temporales entre distintos instantes de la entrada. Las transiciones locales de corta duración

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. L. Rufiner & J. Goddard; "Procesamiento y Clasificación de Fonemas mediante Onditas y Redes con Retardos"
Revista Argentina de Bioingeniería. Vol. 4, No. 1, pp. 11-20, Mar, 1998.

son tratadas por las capas más bajas, mientras que las capas más altas relacionan información que involucra a períodos de tiempo más largos.

En la *Figura 5* se aprecia un esquema de una red con retardos. En este trabajo se utilizó una versión con un retardo que a su vez posee una memoria de decaimiento exponencial.

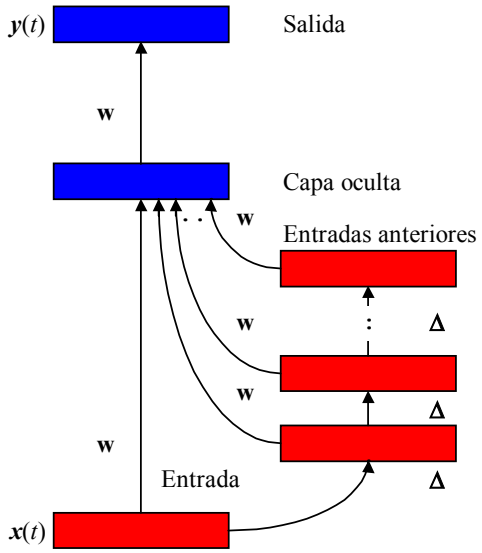


Figura 5: TDNN General

Detalles del procesamiento y el entrenamiento

En este tipo de problemas de clasificación se requiere seguir la evolución temporal de los patrones, para ello se utiliza una ventana temporal que se desplaza sobre la señal produciendo una secuencia de "frames". El tamaño del frame varió en este caso entre 64, 128 y 256 muestras, lo que equivale a 4, 8 y 16 mseg. Como ya se mencionó la familia de Wavelets empleada fue Symmlets con 10 momentos desvanecientes, este parámetro se eligió con el método presentado en [20]. Cada frame se etiquetó de acuerdo al fonema al cual pertenecía. Se varió la forma de activación deseada según se tratara de consonantes (sigmoidea) o vocales (escalón). El tamaño de los archivos de entrenamiento así generados varió entre los 13 y los 16 Mbytes. Se generó de la misma forma un archivo de prueba para determinar la capacidad de generalización de la red a datos no vistos durante el entrenamiento. En todos los casos el entrenamiento se detuvo en el pico de generalización de la red, sin embargo un buen criterio podría ser descartar el primer pico que podría llegar a ser prematuro y esperar al segundo. En ese caso la diferencia de resultados entre entrenamiento y prueba podría ser mayor. El tiempo promedio de entrenamiento de cada red fue de unas 40 horas (en una computadora tipo Pentium a 100 MHz). De cada red se obtuvieron porcentajes de reconocimiento para entrenamiento y prueba y las correspondientes matrices de confusión. En la figura 6 se muestra un esquema de como se realizaron los experimentos.

Existe una variación muy grande en la duración total de los distintos fonemas lo que dificulta su clasificación. El número de nodos por capa se determinó de manera empírica, teniendo en cuenta factores como dimensiones de la entrada y la salida, y cantidad total de patrones.

En la *Figura 7* se observa el tipo de patrones generados para el entrenamiento con onditas comparado con el análisis clásico basado en Fourier.

En la Tabla 1 se pueden observar las distribuciones de los patrones en los archivos de entrenamiento y prueba. Esto es importante a la hora de analizar los resultados finales ya que, como puede observarse, el desempeño global de las redes está fuertemente sesgado por su comportamiento frente a las vocales, que representan del orden del 85% de los patrones. Por otra parte se puede observar que la distribución entre el archivo de prueba y el de entrenamiento es similar.

Resultados

En esta sección se presentan los resultados de los experimentos. En la Tabla 2 se muestran los resultados de precisión de clasificación de las redes neuronales para los distintos fonemas. En la Tabla 3 se exhiben las matrices de confusión correspondientes.

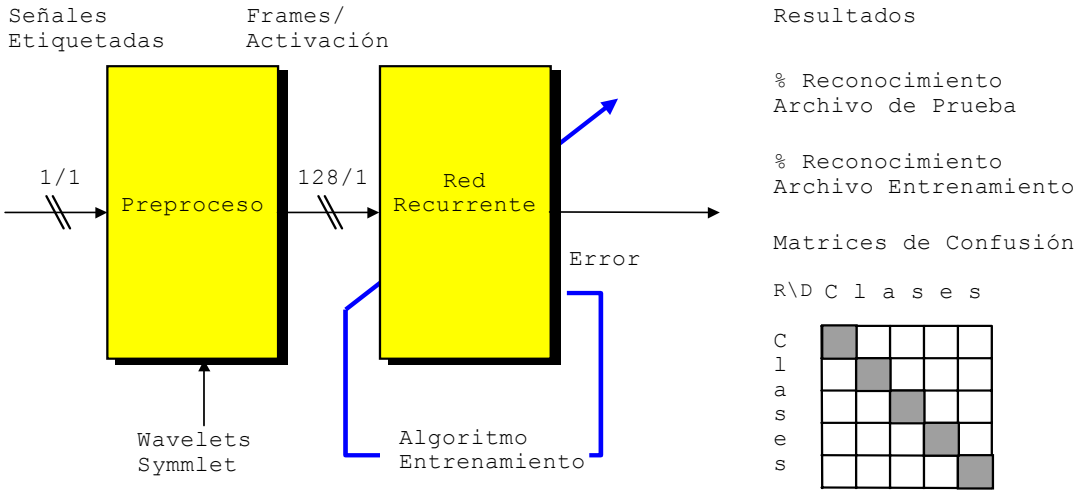


Figura 6: Entrenamiento de las Redes Recurrentes

Discusión y Conclusiones

En el presente trabajo se han presentado los resultados de la clasificación de un conjunto de fonemas de TIMIT procesados mediante las onditas denominadas Symmlets con el fin de evaluar esta alternativa como parte de un sistema de RAH. Hay que recalcar que esta es solo una de las numerosas posibilidades para escoger análisis relacionados con la Transformada de Onditas. Por lo tanto no se puede dar una respuesta definitiva acerca de la conveniencia de este tipo de análisis frente a otros más clásicos.

Un problema importante a resolver en este sentido es encontrar métodos adecuados para escoger una familia determinada de Onditas para el problema de clasificación. En el presente trabajo se utilizó el método presentado en [20] para elegir la cantidad de momentos de desvanecimiento de la Ondita utilizada. Este mismo método puede utilizarse para seleccionar también la familia. Para una comparación de los resultados experimentales con varias familias y el método de selección ver [20, 3].

Evidentemente el tamaño de la ventana de análisis ha influido fuertemente en los resultados (ver Tabla 2). Si bien, en general, la precisión mejora a medida que aumenta el ancho de la ventana, las mejoras más importantes aparecen en los fonemas con componentes transitorias (/b/, /d/, ver Tabla 3). Esto puede deberse a que el clasificador cuenta con mayor cantidad de información temporal acerca de los patrones a medida que se aumenta el ancho de la ventana. El ancho de la ventana, sin embargo, afecta la cantidad de conexiones de la red. En este caso el tamaño va desde unas 16000 conexiones en la red más pequeña, hasta casi 100.000 en la más grande. El hecho de que las redes neuronales con más conexiones convergieron en menos épocas refuerza la suposición de que los patrones poseían información más relevante acerca de los fonemas (a costa de aumentar sus dimensiones).

En resultados obtenidos con la Transformada de Fourier de Tiempo Corto[3] se puede apreciar un comportamiento que se podría denominar inverso, es decir que los mejores resultados se observan sobre los fonemas más estables (/eh/, /ih/, /jh/). Esto sería una consecuencia de la naturaleza de cada uno de los análisis implicados y podría sugerir la implementación de una estrategia híbrida que permitiera utilizar ambos tipos de análisis simultáneamente para mejorar el desempeño global del clasificador (a expensas de incrementar el costo computacional del cálculo)

El punto anterior nos regresa a la pregunta acerca de cuál es el preprocesamiento óptimo para el caso de RAH. En un trabajo reciente [21] se compararon una serie de alternativas (FFT, bancos de filtros, modelos de oído y LPC) para un sistema de RAH sobre TIMIT con redes recurrentes. En este trabajo no se encontró una

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. L. Rufiner & J. Goddard; "Procesamiento y Clasificación de Fonemas mediante Onditas y Redes con Retardos"
Revista Argentina de Bioingeniería. Vol. 4, No. 1, pp. 11-20, Mar, 1998.

diferencia significativa entre los distintos análisis, aunque no se experimento con onditas, y se concluyó que eran mucho más significativos los cambios de arquitectura o estructura de las redes. Sin embargo en [22] se comparó a Fourier contra Onditas aplicados a RAH, y los resultados fueron favorables para Onditas. En este último se trabaja con la Transformada Wavelet Continua Muestreada (TWCM) y se muestrean los coeficientes en escala de Mel, mejorando la resolución en frecuencia con respecto al caso diádico, aunque este enfoque requiere más tiempo de cálculo que el algoritmo rápido.

Actualmente se está trabajando en un esquema basado en Paquetes de Onditas para resolver el problema de la resolución en frecuencia. Otras técnicas emparentadas con Onditas, como Matching Pursuit, podrían constituir otra alternativa.

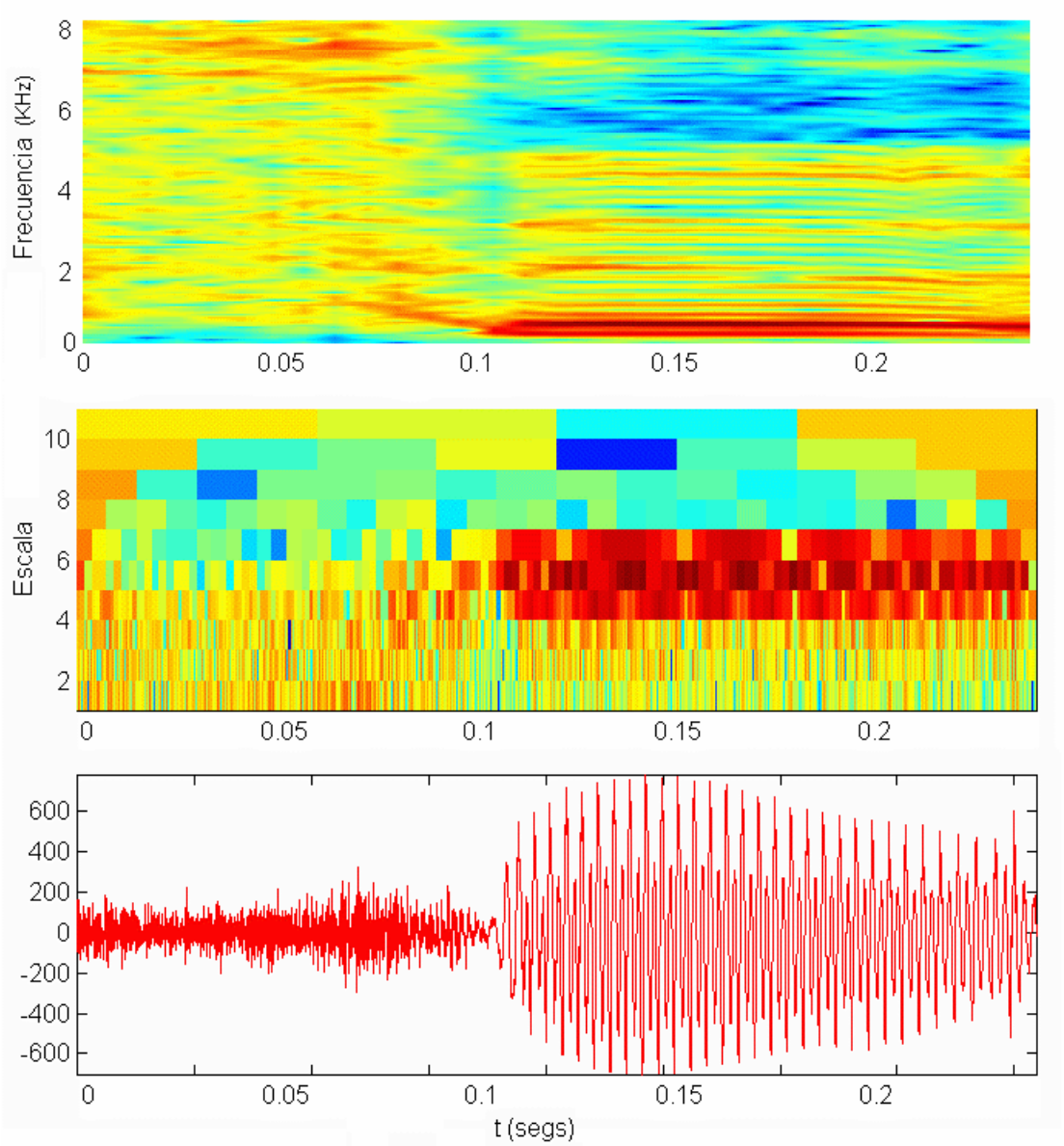


Figura 7: Espectrograma y Escalograma de la sílaba 'su' (inglés)

Tabla 1: Distribución de los patrones

Coeficientes		TRN		TST	
Onditas	**Fonema**	**Frames**	**Distribución (%)**	**Frames**	**Distribución (%)**
256	b	158	4.80	52	5.23
	d	287	8.70	74	7.44
	jh	167	5.10	42	4.22
	eh	1341	40.70	406	40.80
	ih	1345	40.80	421	42.31

Tabla 2: Precisión de las redes

Symmlet (10)	Estructura	Entrenamiento	Prueba	ECM	Épocas
64	64+64/125/5	61.41	56.14	0.2575 / 0.2756	100
128	128+128/154/5	66.43	64.30	0.2457 / 0.2649	60
256	256+256/188/5	73.74	69.85	0.2050 / 0.2267	50

Tabla 3: Matrices de Confusión

	R/D	b	d	jh	Eh	ih
	b	25.25	0.54	0.19	0.03	0.03
	d	22.62	68.71	26.60	0.40	0.68
64	jh	8.85	5.44	65.85	0.07	0.17
TRN	eh	19.02	17.82	4.34	86.63	57.95
	ih	24.26	7.48	3.02	12.87	41.18

	R/D	b	d	jh	eh	ih
	b	16.49	4.23	0.34	0.05	0
	d	21.65	46.76	31.54	0.44	1.12
64	jh	8.25	4.51	61.74	0.27	0.05
TST	eh	27.84	25.63	2.01	82.46	62.22
	ih	25.77	18.87	4.36	16.77	36.61

	R/D	b	d	jh	eh	ih
	b	63.33	0.88	0.0	0.22	0.33
128	D	23.33	84.46	2.01	0.58	0.52
TRN	jh	2.00	8.80	97.71	0.18	0.33
	eh	4.67	1.76	0.0	41.19	12.74
	ih	6.67	4.11	0.29	57.83	86.07

	R/D	b	d	jh	eh	ih
	b	24.32	4.00	0.0	0.47	0.46
128	d	62.16	73.33	5.68	0.12	0.93
TST	jh	0.0	9.33	93.18	0.24	0.12
	eh	0.o	1.33	0.0	40.45	12.88
	ih	13.51	12.00	1.14	58.72	85.61

	R/D	b	d	jh	eh	ih
	b	82.28	1.39	0	0.52	1.56
	d	15.82	95.47	15.57	1.27	1.71
256	jh	0.0	2.44	83.83	0.0	0.07
TRN	eh	0.0	0.35	0.0	75.69	31.75
	ih	1.90	0.35	0.60	22.52	64.91

	R/D	b	d	jh	eh	ih
	b	63.46	6.76	0.0	0.0	0.48
	d	32.69	86.49	23.81	0.99	3.56
256	jh	0.0	0.0	76.19	0.0	0.24
TST	eh	3.85	2.70	0.0	75.62	34.20
	ih	0.0	4.05	0.0	23.40	61.52

Referencias

1. Rioul, O., Vetterli, M., "Wavelets and Signal Processing", IEEE Magazine on Signal Processing, pp. 14-38, October 1991
2. Daubechies, I. "Ten Lectures on Wavelets", Rutgers University and AT&T Bell Laboratories, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992
3. Rufiner, H.L. "Comparación entre Análisis Wavelets y Fourier aplicados al Reconocimiento Automático del Hablar", Tesis de Maestría en Ingeniería Biomédica, U.A.M.-I, Diciembre 1996.
4. Garofolo, Lamel, Fisher, Fiscus, Pallett, Dahlgren, "DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus Documentation", National Institute of Standards and Technology, February 1993.
5. Meyer, y., "Principe d'incertitude, bases hilbertiennes et algebres d'operateur", Seminaire Bourbaki 38 no. 662 (1985-6) .
6. Wojtaszczyk, P.,"A Mathematical Introduction to Wavelets", London Mathematical Society Student Texts No. 37, Cambridge University Press, 1997
7. Mallat, S.G., "A Theory of Multiresolution Signal Decomposition: the Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, 1989

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. L. Rufiner & J. Goddard; "Procesamiento y Clasificación de Fonemas mediante Onditas y Redes con Retardos"
Revista Argentina de Bioingeniería. Vol. 4, No. 1, pp. 11-20, Mar, 1998.

8. Mallat, S.G., "Multiresolution approximation and wavelet orthonormal bases of $L^2(R)$",Trans. Amer. Math. Soc. 315, 1989, pp. 69-88
9. Qian, S., Chen, D., Joint Time-Frequency Analysis, Prentice-Hall, 1996.
10. Woodland, P., Young, S., The HTK tiered-state continuous speech recognizer, Eurospeech '93, pp.2207-2210, 1993
11. Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., Lang, K.,"Phoneme Recognition using Time-Delay Neural Networks", IEEE Trans. ASSP Vol. 37, No. 3, 1989
12. Nádas, A., Some relationships between ANNs and HMMs, en Artificial Neural Networks for Speech and Vision, Mammone (editor), Chapman & Hall, 1994
13. Rahim,M.G., A Self-learning neural tree network for phone recognition, en Artificial Neural Networks for Speech and Vision, Mammone (editor), Chapman & Hall, 1994
14. Morgan, N., Bourland, H., An introduction to the hybrid HMM/connectionist approach, IEEE Signal Processing Magazine, May 1995
15. J.L. Elman, "Finding structure in time", Cognitive Science 14 (1990) 179-211.
16. D.W. Tank, J.J. Hopfield, "Concentrating information in time : Analog Neural Networks with Applications to Speech Recognition Problems", IEEE First International Conference on Neural Networks, 1987.
17. J.L. Elman, D. Zipser, "Learning the Hidden Structure of Speech", JASA 83, pp.1615-1626, 1988.
18. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, K. Lang; "Phoneme Recognition Using Time-Delay Neural Networks". IEEE Trans. ASSP Vol. 37, No 3 (1989).
19. R. Lippman, "Review of Neural Networks for Speech Recognition", Neural Computation, 1 (1), 1-38.
20. Rufiner, H.L., Goddard, J.,"A Method of Wavelet Selection in Phone Recognition", enviado al 40th Midwest Symposium on Circuits and Systems, California, 1997
21. Robinson, T., Holdsworth, J., Patterson, R., Fallside, F.,"A comparison of preprocessors for the cambridge recurrent error propagation network speech recognition system",
22. Richard F. Favero; "Comparison of Perceptual Scaling of Wavelet for Speech Recognition", SST-94.
23. T. Robinson, J. Holdsworth, R. Patterson, F.Fallside, "A comparison of preprocessors for the cambridge recurrent error propagation network speech recognition system", Cambridge University Engineering Department.