CLASIFICACIÓN DE FONEMAS MEDIANTE PAQUETES DE ONDITAS ORIENTADAS PERCEPTUALMENTE

H. Torres, H. L. Rufiner

UNER, Laboratorio de Cibernética, Ruta 11 Km 10, Paraná,
(3100), Entre Ríos, Argentina, cyberlab@fi.uner.edu.ar

RESUMEN:

Los paquetes de onditas (WP) son una extensión de la transformada de onditas (WT) introducida recientemente. Si bien su aplicación ha sido amplia en problemas de compresión, filtrado y supresión de ruido, poco se ha hecho en el problema de clasificación de señales. Los fonemas del habla constituyen un ejemplo de señales de duración variable de difícil separación. En este trabajo se presentan los resultados de la aplicación de paquetes de onditas diseñados en base a criterios perceptuales para el análisis y la clasificación de fonemas. Debido a la naturaleza dinámica de los patrones el clasificador utilizado es una red neuronal con retardos temporales (TDNN).

1. 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 (FT), que examina una señal a una resolución fija, la transformada de onditas (WT) posee la característica de hacerlo a distintas escalas (ó resoluciones). 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 [2]. 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 un trabajo reciente [3], se estudió la aplicación de las onditas al procesamiento de señales de voz, para su posterior clasificación en fonemas. La familia de onditas utilizadas fue la basada en las denominadas Symmlets [4]. Los fonemas fueron extraídos de la base de datos TIMIT [5] y se utilizó una red neuronal con retardos para realizar la clasificación de los patrones generados mediante las onditas. En este nuevo trabajo se ha reemplazado la WT por la de paquetes de onditas (WPT), donde la base se ha diseñado especialmente para comportarse en forma similar al oído. Esto se logra distribuyendo el ancho de banda de las señales de la base (filtros) de acuerdo a una escala psicoacústica denominada escala de Mel.

Este artículo se organizará de la siguiente forma. A continuación se presentan los fundamentos de la WT y la WPT. En la sección siguiente se explica la motivación perceptual basada en la escala de Mel. Seguidamente se describen los experimentos realizados. Finalmente se presentan los resultados de los mismos y se analizan en la última sección.

2. ONDITAS Y PAQUETES DE ONDITAS

El área de onditas empezó a desarrollarse a mediados de los años 80´s con el trabajo de Meyer [6]. Desde entonces, la WT ha demostrado ser una herramienta importante para el procesamiento de señales debido a su manera natural de analizar señales con discontinuidades y picos (transitorios). Excelentes referencias son Daubechies [4], Wojtaszczyk [7], y Mallat [8].

Para el caso de señales muestreadas, existe la transformada discreta de onditas (DWT), que además posee una implementación rápida denominada transformada rápida de onditas (FWT). La FWT se implementa con un árbol de filtros diádicos submuestreos por 2. Sin embargo la DWT es realmente un subconjunto de la WPT [9]. La WPT generaliza el análisis tiempo-frecuencia realizado por la D

RESUMEN:

En el presente trabajo se utilizan un perceptrón multicapa junto con una estructura OCON (*One Class One Net*) en la modelización del riesgo de intoxicación por digoxina. Este fármaco, ampliamente usado en el tratamiento de la insuficiencia cardíaca y la fibrilación auricular, puede dar lugar a una intoxicación del paciente debido a su estrecho ámbito terapeútico. La presente comunicación plantea la utilización de técnicas no lineales como son las redes neuronales para la modelización del riesgo de intoxicación.

el árbol o base adecuada para una aplicación particular. Para aplicaciones de compresión se ha desarrollado el algoritmo de la mejor base (BB) minimizando la entropia de la señal analizada [10].

3. ESCALA DE MEL

El Mel es una unidad de medida de la frecuencia de un tono percibida por el oído [11]. No se corresponde linealmente con la frecuencia física del mismo, dado que el oído humano no percibe el tono de una manera lineal. En base a experimentos psicoacústicos fue posible determinar la forma aproximada de la respuesta del oído. Una aproximación a esta respuesta está dada por la Ecuación 1.

$$F_{Mel} = \frac{1000}{\log_2\left(1+\frac{F_{Hz}}{1000}\right)} \qquad (1)$$

En donde $F_{Mel}(F_{Hz})$ es la frecuencia percibida. Como se puede apreciar es una relación logarítmica que "comprime" el espectro a frecuencias altas. Basado en estas ideas se puede realizar el ajuste del ancho de banda de los filtros involucrados en la generación del paquete de onditas utilizado para el análisis de manera que se ajusten a esta escala perceptual. Como consecuencia esto también fija la resolución temporal en cada banda. El ajuste se realizó sobre la base de las señales muestreadas a 16 KHz, completando un total de 19 bandas. La disposición del árbol de filtros correspondiente se aprecia en la Figura 1 y la partición del plano tiempo- frecuencia generada en la Figura 2.

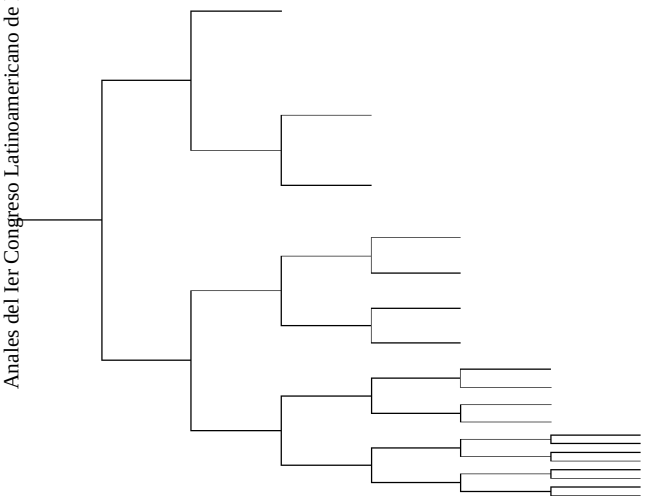

Figura 1: Diseño del árbol de filtros.

4. REDES CON RETARDOS

Existen muchas técnicas de clasificación que se han aplicado al caso de los fonemas [11]. En este trabajo se utiliza una red con retardos [12] con diferentes arquitecturas para realizar una clasificación de los fonemas escogidos luego de procesarlos con la WPT. Una razón para elegir este tipo de clasificador es que permite la utilización del clásico algoritmo de retropropagación casi sin modificaciones. En pruebas de comparación [2] funcionó mejor que otras arquitecturas de redes recurrentes simples como las de Jordan y Elman [13].

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.

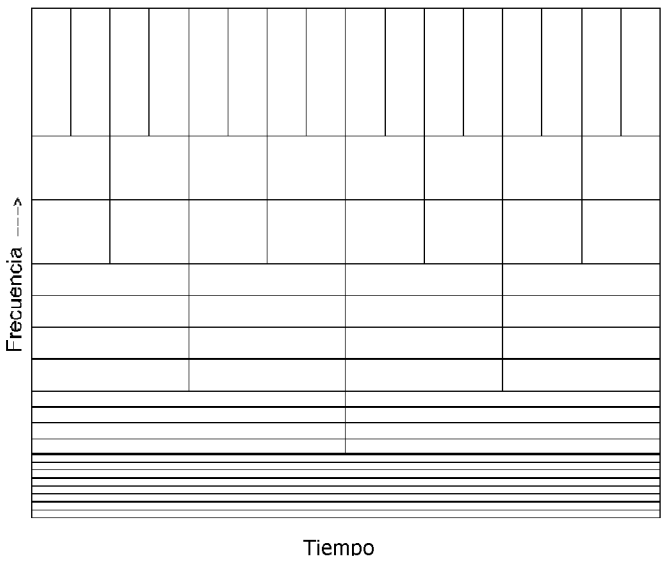

Figura 2: Partición del plano tiempo-frecuencia

5. EXPERIMENTOS

Los fonemas escogidos para los experimentos fueron tomados de la base de datos TIMIT [6]. Este corpus fue construido especialmente por Texas Instruments y el Massachusetts Institute of Technology para realizar experimentos con sistemas de reconocimiento automático del habla. 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 totalizando unos 650 Mbytes de información. Los datos fueron digitalizados a 16 KHz, 16 bits por muestra. Por el tamaño de TIMIT los experimentos se enfocaron sobre un subconjunto del total de fonemas (/b/, /d/, / jh/, /eh/, /ih/) [2].

Los experimentos consistieron en variar el ancho de las ventanas consisderadas, la familia de onditas inicial (Splines y Daubechies [4]) y varios pruebas con submuestreo (agrupamiento) de los coeficientes obtenidos. Esto último se justifica en que la resolución temporal puede ser excesiva para algunos anchos de ventana, principalmente en las altas frecuencias. Por otra parte ayuda a la clasificación porque reduce la dimensionalidad de los patrones. Los parámetros para cada familia de onditas madres fueron seleccionados por el método [14].

6. RESULTADOS

En las Tablas 1-4 se presentan los porcentajes de reconocimiento de las redes con retardos entrenadas con los patrones transformadaos para las diferentes familias y anchos de ventana utilizados. En la Tabla 1 se muestran los resultados de entrenar directamente con las señales transformadas. En la Tabla 2 se realiza previamente una integración en cada banda. En la Tabla 3 corresponde al agrupamiento de a 4 coeficientes y en la Tabla 4 de a 8 coeficientes.

	Splines		*Daubechis*	
Ancho	*TRN*	*TST*	*TRN*	*TST*
64	47.99	47.42	68.72	60.12
128	66.61	64.77	66.99	63.91
256	73.00	63.80	79.17	73.61

TABLA 1: RECONOCIMIENTO PARA LAS DIFERENTES FAMILIAS

	Splines		*Daubechis*	
Ancho	*TRN*	*TST*	*TRN*	*TST*
64	60.83	60.63	63.02	60.61
128	66.50	64.05	71.82	69.21
256	71.15	66.67	68.76	67.79

TABLA 2: RECONOCIMIENTO INTEGRADO POR BANDAS

	Splines		*Daubechis*	
Ancho	*TRN*	*TST*	*TRN*	*TST*
256	74.66	70.89	68.02	65.61
512	75.37	72.95	82.39	79.49

TABLA 3: RECONOCIMIENTO AGRUPAMIENTO DE A 4

	Splines		*Daubechis*	
Ancho	*TRN*	*TST*	*TRN*	*TST*
512	80.29	74.59	79.33	74.3

TABLA 4: RECONOCIMIENTO AGRUPAMIENTO DE A 8

En la Figura 3 se puede apreciar el escalograma resultante de aplicar la WPT con la familia daubechies, correspondiente a los fonemas /eh/ y /jh/ respectivamente.

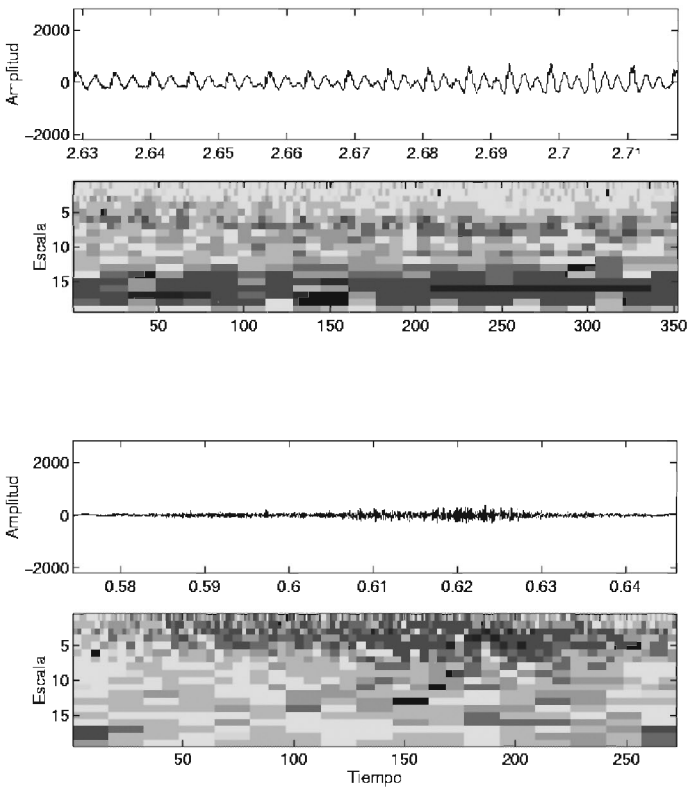

TABLA 3: ESCALORAMAS DE /EH/ Y /JH/

7. DISCUSIÓN Y CONCLUSIONES

En este trabajo se presentó una base de paquetes de onditas diseñada especialmente para el análisis y clasificación de fonemas.

Como se puede apreciar en la Tabla 2 los mejores resultados son los obtenidos para Daubechies con una ventana de 512 y submuestreo por 4. Estos son mejores que los reportados en [2] para la DWT con idénticas onditas madres. Esto se debe a la mejor discriminación de las frecuencias formantes que no se obtiene con la resolución de 1 coeficiente por octaba de la DWT. El

comportamiento del clasificador es inclusive mejor que el obtenido procesando las señales con la FT. Esto puede deberse a la mejor resolución temporal en las bandas de alta frecuencia que permite diferenciar mejor a los fonemas con componentes transitorias.

8. AGRADECIMIENTOS

Este trabajo se realizó con el apoyo de la Universidad Nacional de Entre Ríos y el CONICET (Argentina).

9. REFERENCIAS

[1] Rioul, O., Vetterli, M., "Wavelets and Signal Processing", IEEE Magazine on Signal Processing, pp. 14-38, October 1991

[2] 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.

[3] Rufiner H. L., Goddard J., "Procesamiento y Clasificación de Fonemas mediante Onditas y Redes con Retardos", Revista Argentina de Bioingeniería, Vol 4, N°1, Marzo 1998.

[4] Daubechies, I. "Ten Lectures on Wavelets", Rutgers University and AT&T Bell Laboratories, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.

[5] Garofolo, Lamel, Fisher, Fiscus, Pallett, Dahlgren, "DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus Documentation", National Institute of Standards and Technology, February 1993

[6] Meyer, y., "Principe d'incertitude, bases hilbertiennes et algebres d'operateur", Seminaire Bourbaki 38 no. 662 (1985-6).

[7] Wojtaszczyk, P., "A Mathematical Introduction to Wavelets", London Mathematical Society Student Texts No. 37, Cambridge University Press, 1997

[8] 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

[9] M.A. Cody, *"The Wavelet Packet Transform : Extending the Wavelet Transform"*, Dr. Dobb's Journal, April 1994.

[10] M.D. Wickerhauser, *"Acoustic Signal Compression with Wave Packets"*, Yale University, 1991.

[11] J. Deller, J. Proakis, J. Hansen, *"Discrete Time Processing of Speech Signals"*. Macmillan Publishing, NewYork, 1993.

[12] 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.

[13] J.L. Elman, "Finding structure in time", Cognitive Science 14 (1990) 179-211.

[14] Rufiner, H.L., Goddard, J., "A Method of Wavelet Selection in Phone Recognition", enviado al 40th Midwest Symposium on Circuits and Systems, California, 1997.