ANGELO MAIA CISTER
COPPE-Sistemas
Universidade Federal do Rio de Janeiro
CARLOS FELIPE JACCOUD
Empresa Brasileira de Telecomunicações - EMBRATEL
Rio de Janeiro
BRASIL

Reconocimiento del habla mediante una Red Neuronal con pre-procesamiento Wavelet

En este trabajo presentamos un estudio de reconocimiento del habla que utiliza como pre-procesamiento una Función Wavelet No Ortogonal, conocida como Función Morlet. El prototipo fue implementado en Matlab.
El reconocimiento del habla es realizado por una Red Neural Artificial (Backpropragation). Este estudio presenta, como característica relevante, la independencia del locutor.
Es elaborada una breve relación de los trabajos que han sido realizados sobre Reconocimiento del Habla, presentando sus principales investigadores. En el trabajo, también presentamos un estudio teórico sobre conocimiento para comprender textos más complejos.
El trabajo finaliza discutiendo las ventajas y desventajas del uso de una Función Wavelet en conjunto con una RNA. Recomendamos nuevos caminos a partir de nuestra experiencia en el desarrollo del presente trabajo.

1. El ser humano, el habla y el lenguaje natural.

El habla, la señal sonora producida por el hombre, resulta un fenómeno mucho más complicado que lo que podemos imaginar. Cada locutor presenta una serie de características que lo distinguen de otros locutores. Entre estas características podemos citar: la frecuencia, los vicios del lenguaje (muchas veces por razones culturales e/o intelectuales), entonación (por razones regionales) y hasta el tono con que cada locutor pronuncia las palabras (evidenciando su estado de espíritu y/o humor) [Johnston, 1994]. Podemos identificar el sexo del locutor a través de la frecuencia de voz. Si alguien dice que es fácil identificar si el locutor es hombre o mujer, sería suficiente realizar un pre-procesamiento de la voz y generar un espectrograma para confirmar que la voz femenina posee una frecuencia mayor que la masculina.
El análisis y observación de las características del habla resulta simple cuando utilizamos el cerebro, que es un poderoso computador en el análisis cualitativo, al contrario del computador que es excelente en lo relacionado al análisis cuantitativo.
En este trabajo, presentamos un estudio sobre el reconocimiento del habla utilizando como pre-procesamiento la Transformada Wavelet y el reconocimiento del habla es realizado por una Red Neural Artificial -RNA- de tipo Multicamadas con Retropropagación de Errores (Backpropagation).

2. Un breve estudio del estado del arte.

Los estudios sobre reconocimiento del habla se iniciaron, aproximadamente, hace cuatro décadas. A partir de 1950 se iniciaron los estudios de investigadores que buscaban explorar las ideas básicas de la fonética-acústica, sin resultados satisfactorios.
Entre los años 1950 hasta 1959, los estudios fueron lidereados básicamente por investigadores americanos e ingleses. Dentro de las investigaciones más relevantes durante esta década debemos resaltar los trabajos de Olson y Belar, de los Laboratorios RCA, y los trabajos de Fry y Denes, de College University en Inglaterra.
A partir de 1960, se inicia una verdadera explosión en las investigaciones en reconocimiento del habla, y en este momento los japoneses en escena. Sin dudas la técnica más expresiva desarrollada en esta época fue el Análisis del Cruzamiento por Cero (Zero-Crossing Analysis) que conseguía distinguir entre varias regiones del estímulo facilitando, en gran medida, el reconocimiento. Una otra técnica, no menos importante, desarrollada en la década del 60, es la Programación Dinámica para alineamiento temporal. Esta técnica fue desarrollada por el científico ruso Vintsync. A pesar que la base del concepto fue

considerada rudimentaria, las versiones más avanzadas de su algoritmo son muy conocidas en accidente, que utilizó ampliamente este algoritmo a inicios de la década de 80.
En 1970, ya se contaba con un gran número de investigadores en el área. Investigadores rusos estudiaron diversas maneras para reconocer un modelo del habla, los japoneses presentaron la idea de Codificación Predictiva Lineal (LPC-Linear Predictive Coding), que viene siendo una serie utilizada con satisfacción hasta nuestros días. En los laboratorios Bell se inició una serie de experimentos con el objetivo de crear un sistema de reconocimiento del habla con independencia de locución. [Rabiner et al, 1979].
Una "nueva" tecnología que fue utilizada a mediados de 1980, fue la idea de aplicar las Redes Neuronales Artificiales (RNA) [Morgan, 1995], [Jelinek, 1976] para el reconocimiento de patrones del habla. Las RNA, inicialmente, fueron introducidas durante la década del 50, sin embargo sus resultados no fueron considerados satisfactorios debido a un conjunto de problemas operacionales que fueron subsanados a medida que el reconocimiento avanzó.
La década del 80 es, reconocidamente, la década donde el avance en la pesquisa en reconocimiento del habla presenta sus mejores resultados. Como ejemplo podemos mencionar el sistema desarrollado por la Agencia de Proyectos de Investigaciones Avanzadas para Defensa (DARPA-USA) que reconoce con precisión mil palabras en comunicación continua. Se trata de palabras claves utilizadas para el control y comando de misiles. El programa continua en la década del 90 enfatizando técnicas de Inteligencia Artificial, como es el caso del área de estudio en procesamiento del lenguaje natural [Bahl et al, 1989].

3. La Transformada Wavelet

La Transformada Wavelet - TW es una "moderna" técnica de análisis que viene siendo desarrollada y difundida durante los últimos 10 años. Una característica que distingue TW es la transformación del dominio temporal para el dominio tiempo-frecuencial. Por esta razón muchos autores la denominan Análisis Tiempo-Frecuencial diferenciándola del secular Análisis de Fourier (Transformada de Fourier-TF) que transforma la señal para el dominio frecuencial.
Otra característica importante en la TW es el concepto de Soporte Compacto, que implica que cualquier punto fuera del intervalo especificado para la función, será nulo. El concepto de Soporte Compacto permite a la TW cambiar del dominio temporal para el dominio frecuencial [Cody, 1992].
Una ventaja de la TW en relación a la TF es que la primera tiene la capacidad de detectar señales transientes, es decir, señales que varían muy rápidamente en el tiempo. Esta ventaja se debe a la característica de la TW de ser una ventana de barredura deslizante con dimensión variable, que permite una resolución tanto a nivel de tiempo como de frecuencia.
Aunque iniciamos esta sección presentando la TW como una técnica moderna, su verdadera concepción data de inicios de siglo. Entre los autores existen discrepancias en relación al momento exacto del surgimiento y de los creadores que desarrollaron el concepto de acuerdo con Vetterli [Vetterli, 1995], la primera mención a este concepto se encuentra en la tesis de Haar en 1904. Sin embargo Y. Meyer [Meyer, 1992], manifiesta que el concepto fue desarrollado en los trabajos de Lusin y Calderón en 1921.
Lo que nos importa, en este momento, es que la función desarrollada por Haar es una base matemática *Wavelet* y que dentro de las investigaciones modernas con respecto a TW, podemos decir que la base de Haar es un límite inferior. Es decir, se localiza extremadamente en el tiempo, y como un límite superior tenemos la función sinc(x) que es extremadamente localizada en la frecuencia. De esta forma cabe, a la comunidad científica descubrir nuevas funciones con las características de la TW que nos auxilien en los experimentos, sabiendo que las familias de funciones que por ventura sean descubiertas estarían limitadas por estas dos funciones.

3.1 El Interés por la Transformada.

Es un intento de suplir las dificultades que están presentadas por la Transformada de Fourier - TF, como la discontinuidad mencionada anteriormente, surgió la necesidad de analizar la señal, tanto a nivel temporal como frecuencial, es decir, tener una buena resolución tiempo-frecuencia.
J. Morlet, geofísico francés, a partir de estudios anteriores, creó una función que multiplicaba una función armónica compleja con una función gaussiana, la cual llamó analizadora (función ventana). La función analizadora puede ser ajustada, dilatada o contraída. De esta forma, mientras mayor es la anchura de la ventana menor son los picos de frecuencia, en consecuencia la resolución temporal aumenta.

Posteriormente, el matemático francés Y. Meyer reconoció el trabajo de Morlet como parte del campo de estudios en análisis armónicos, presentando una serie de familias de funciones *Wavelet*, que es más eficiente para el modelaje de fenómenos complejos. Las investigaciones de Meyer fueron comprobadas por los científicos norteamericanos Stephane. Mallat de New York University e Ingrid Daubechies de los laboratorios Bell [Vetterli, 1995].

3.2 El análisis *Wavelet* propiamente dicho.

Cuando se desea la expansión de una señal, la idea principal dadas las funciones bases , es la localización en el tiempo y en la frecuencia. En la Transformada de Fourier, por ejemplo, las funciones utilizadas en el análisis son infinitamente precisas en lo que se refiere a la localización frecuencial, sin embargo, resultan pobres en el aspecto relativo al tiempo debido a su expansión infinita.
Cuando creamos una función o, mejor dicho, una familia de funciones que nos puedan ofrecer localizaciones precisas en el tiempo y en la frecuencia sería necesario, entre algunas , dos características principales.

- Traslación

Una traslación en el tiempo t_0 , resulta la multiplicación por un factor de fase en dominio frecuencial, esto es:

$$f(t - t_0) \qquad e^{-jwt_0}F(w) \tag{1}$$

Convencionalmente, una traslación en la frecuencia resulta un factor de fase, o modulación por un exponencial complejo, en el dominio temporal

$$e^{-jw_0t}f(t) \qquad F(w - w_0) \tag{2}$$

- Escalonamiento

En el tiempo resulta un escalonamiento inverso en la frecuencia, como es presentado en la siguiente transformada:

$$f(at) \Leftrightarrow \frac{1}{|a|}F\left(\frac{w}{a}\right); \ a \neq 0 \tag{3}$$

A partir de estas dos características fundamentales, podemos definir una familia de funciones del tipo:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right); \quad a,b \in \mathcal{R} \neq 0 \tag{4}$$

donde ψ es llamada *Wavelet* madre o *Wavelet* base, que tiene su localización tanto en el tiempo como en la frecuencia. $\psi(t)$ generalmente es un filtro pasa-banda. Cuando *a* es muy grande ($a \gg 1$), corresponde a una función que tiende a poseer muchos términos para ser analizados.
Para *a's* pequeños ($0 < a < 1$), las funciones tienden a poseer términos menores en el comportamiento de la señal.
Esta situación nos lleva a la siguiente conclusión: ***la escala es proporcional a la duración de las funciones bases utilizadas en la expansión de la señal.***
Matemáticamente tenemos:

$$f = \sum_i \alpha_i \psi_i \ ; \ i \in Z \ ; \{\psi_i\} \in \mathcal{H} \tag{5}$$

donde $\{\psi_i\}$ es completo en H (Espacio de Hilbert).
La experiencia aquí relatada sólo considera cuando $\{\psi_i\}$ es una base, es decir, los vectores ψ_i son linealmente independientes (LI).
Una función $f(t)$ definida en R es considerada una función en el espacio cuadrático - integrable, o sea, en el Espacio de Hilbert L^2 (R) [Halmos, 1970, 1982], si $|f(t)|^2$ es integrable, es decir, $\int_{t \in R} |f(t)|^2 \, dt < \infty$

donde su producto interno en L^2 (R) es definido por $\langle f, g \rangle = \int_{t \in \mathcal{R}} f(t)\overline{g(t)} \, dt$ y su norma (distancia) definida por $\|f\|^2 = \langle f, f \rangle = \int_{t \in \mathcal{R}} |f(t)|^2 \, dt$

En este contexto definimos la TW Continua de una función $f \in L^2$ (R) como:

$$Tf(a,b) = \langle f, \psi_{a,b} \rangle = \frac{1}{\sqrt{a}} \int f(t) \overline{\psi\left(\frac{t-b}{a}\right)} \, dt \tag{6}$$

Aquí, $\psi_{a,b}$ desempeña el mismo papel que e^{jwt} en la TF.
Observamos que si a es constante podemos interpretar (6) como una convolución de la función $f(t)$ con la analizadora.
De este modo, la representación final de la TW es una sucesión de convoluciones de la función f con una analizadora cuya anchura (tamaño de la ventana) es variable.
Y. Daubechies en su trabajo [Daubechies, 1978] presenta la forma discretizada de los parámetros a y b.
$a = a_0^m$ (anchura de la ventana)
$b = nb_0 a_0^m$ (traslación de la ventana)
Obsérvese la dependencia de la traslación en relación a la anchura de la ventana.

- Propiedades de la transformada de *Wavelet*

Dentro de las condiciones y propiedades necesarias para poder definir una TW de clase *m*, además del estacionamiento y la traslación, podemos mencionar:

1. Condición de la Regularidad
$$\frac{\partial^k}{\partial x^k} \psi(x) \in L^\infty(\Re); 0 \le k \le m$$
2. Decaimiento Asintótico
$$\lim_{k \to \infty} \frac{\partial^k}{\partial x^k} \psi(x) = 0$$
3. Energía Finira
$$\int_{-\infty}^{+\infty} x^k \psi(x) \partial x = 0; \ 0 \le k \le m$$

4. Multiresolución y banco de filtros.

La posibilidad de poder considerar una señal con diferentes niveles de resolución, es decir con multiresolución, es muy utilizada en procesamiento de imagen/vídeo y visión por computador, donde

normalmente las versiones de imágenes con baja resolución son utilizadas no una primera aproximación de los algoritmos computacionales. Un ejemplo práctico es la consulta de bases de datos de imágenes en WWW, donde inicialmente son transmitidas con menor retardo las versiones de baja resolución de las imágenes. Los detalles solo son transmitidos si la imagen es de interés.
El concepto de Multiresolución puede ser implementado utilizando técnicas de Procesamiento Digital de Señales-PDS a través de bancos de filtros digitales Pasa-altos-FPA y pasa bajos-FPB, como representamos en la Figura 1. Consideremos, por ejemplo, una señal s(n) con rango útil de frecuencia igual a Bw_s y frecuencia y frecuencia de muestreo igual a f_a. Como primera aproximación de esa señal podrían ser consideradas las señales $S_{HI}(n)$ y $S_{LI}(n)$, con anchura de rangos iguales a BW_{SHI} y BW_{SLI}, tales que BW_{SHI} + BW_{SLI}, y con frecuencia de muestreo iguales a $f_a/2$. Estas señales componen una representación de baja resolución de s(n).

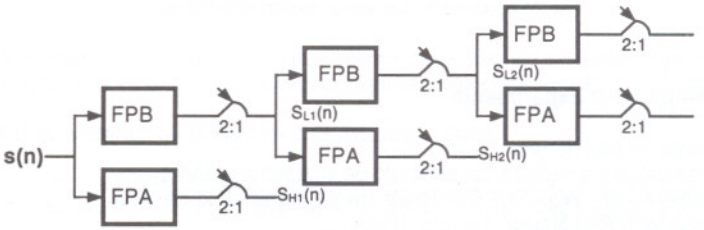

Fig. 1 Banco de Filtros Multiresolución

La subdivisión del rango de frecuencias de BW_{SLI} en dos nuevos rangos de la misma anchura, BW_{SL2} y BW_{SH2}, garantizará, a partir de las señales de $S_{HI}(n)$, $S_{L2}(n)$ una mejor representación de s(n). A cada subdivisión de los rangos de frecuencia, será obtenida una mayor resolución espectral, para una dad región del espectro. Este proceso puede ser repetido indefinidamente, considerando que la complejidad de implementación del método es proporcional al número de sub-bandas generadas. Genéricamente, una banco de filtros multiresolución tendría una respuesta de frecuencia como mostramos en la Figura 2.
Una de las soluciones más eficientes para la implementación de bancos de filtros consiste en el uso de Filtros de Espejos en Cuadratura - QMF. Estos filtros, considerados junto con los pares (Filtro Pasa-bajos + Filtro Pasa-altos), presentan una respuesta en frecuencia simétrica en relación a la frecuencia central del rango de frecuencias consideradas. Garantizan para la estructura una ganancia unitaria útil de la señal analizada y sin ocurrencia de distorsiones relativas a la superposición espectral, que es permitida entre los rangos de transición de aquellos filtros. Una de las formas posibles de utilización de estos filtros es una estructura en árbol como presentamos en la Figura 1.

¿Cuál es la relación entre las estructuras y la TW? A partir de la descripción elaborada por Mallat [Chui, 1992] y Meller [Meller, 1992] el concepto de multiresolución constituye la base matemática de la teoría de *Wavelet*, ya que su descomposición es un método de aproximaciones sucesivas que adiciona más proyecciones en el "espacio de detalles" o "espacios barridos" por las *Wavelets* y sus traslaciones con escalas distintas. De este modo, a partir de un Banco de Filtros *Wavelet* puede obtenerse un análisis espectral con diferentes niveles de resolución.

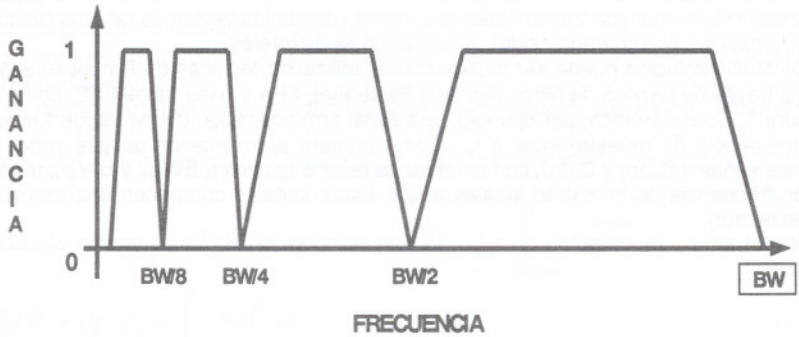

Fig. 2 Respuesta en frecuencia del Banco de Filtros

5. Resultados Obtenidos en el Experimento

Dividimos el experimento en tres partes:
FASE 1: Adquisición de datos y/o captura del estímulo de voz(señal tratado).
FASE 2: Pre-procesamiento (BF *Wavelet* - Extracción de los parámetros de voz).
FASE 3: Entrenamiento de la Red Neural Artificial - RNA

5.1 FASE 1: Adquisición de datos y/o captura del estímulo de voz (señal tratado).

Para la captura de la señal de voz utilizamos un micrófono unidireccional. Las señales fueron recogidas a partir de tres hombres adultos y un joven (adolescente), sin ningún tratamiento acústico ambiental. La única preocupación fue no tener ruidos que pudiesen interferir, comprometiendo la calidad de la grabación.
Tres de los cuatro locutores repitieron cinco veces los estímulos referentes a los números de cero (0) a nueve (9), de manera aleatoria buscando eliminar cualquier vicio de habla relativo a un determinado estímulo.
Las grabaciones se realizaron directamente en el computador. Para esto utilizamos un programa cedido por el fabricante de la placa de sonido (*Wave Studio*). Las grabaciones mantuvieron como característica los siguientes parámetros:

- Frecuencia de muestreo: 8 KHz
- Volumen de grabación: Máximo
- Modo de grabación: Mono
  Resolución: 16 bits

Como la duración máxima de cada uno de los estímulos es del orden de 0,5 *s*, optamos por un tamaño igual a 4096 muestras con duración igual 0,512 *s* = número de muestras/frecuencia de muestreo - f_a para cada uno de los estímulos, correspondiendo a la base de datos que contiene todos los estímulos alcanzó un tamaño igual a 25x8x10=2000 Kb [número de estímulos por clase x tamaño de cada estímulo de clase x cantidad de clases de estímulo (0=9)].

5.2 FASE 2: Pre - procesamiento (BF Wavelet - Extracción de los parámetros de la voz)

Esta fase considera la adquisición de parámetros necesaria para el reconocimiento de la señal.
La fase 2 consiste de:
1. Filtrado Digital (partes real e imaginaria)
2. Reducción de muestras (Desampled)

3. Cálculo de las energías medias por bloques de muestras (entre 10 y 30 ms de duración).

5.2.1 Filtrado digital.

El filtrado digital es la aplicación directa del concepto de los Filtros QMF. Como función prototipo fue utilizada la Función Morlet con alteraciones en los parámetros de la función gaussiana.
La función Morlet original no resultó adecuada para analizar las señales con frecuencia de muestreo de 8 KHz, por esta razón fuimos obligados a introducir una constante β de valor igual a $2x10^6$ para que pudiese anular el orden de grandeza de la frecuencia de muestreo. Por la función de Morlet teníamos:

$$e^{-\frac{\left(\frac{t}{a}\right)^2}{a}} \text{ donde } t = nT_a = \frac{n}{f_a} \Rightarrow e^{\left(\frac{n^2}{2a^2}\right)\frac{1}{f_a^2}}$$

$$f_a^2 = \frac{1}{64.10^6} \text{ tenemos: } e^{-\left(\frac{\beta n^2}{2a^2 f^2_a}\right)} \Rightarrow e^{-\left(\frac{n^2}{64.a^2}\right)}$$

$$\text{con } \beta = 2.10^6$$

Con estas alteraciones conseguimos el decaimiento necesario de la función gaussiana. También más oscilaciones dentro de la ventana, lo que facilitó capturar más características de la señal a ser analizada.

5.3 FASE 3: Entrenamiento de la Red Neural Artificial - RNA.

Para la realización del experimento, optamos por utilizar una red de tipo multicamadas con retropropagación de errores y con control del Momento y entrada de los datos en lote. El concepto de Momento fue introducido por Rumelhart en 1986 [Wasserman, 1989] [Haykin, 1994]. Básicamente este concepto auxilia en la convergencia y estabilidad del proceso.
Utilizando el método del Momento, la red tiende a "pasear" por la superficie de error más suavemente (a pasos más cortos). Actuando como un filtro pasa-bajos, permite que la red ignore pequeñas características en la superficie de error evitándose que caiga en un mínimo local.
Los parámetros que hicieron a nuestra red converger fueron los siguientes: *momentum* = 0.2 e *lr* = 0.01.
El sistema se mantuvo estable para los valores presentados. A partir de este momento nos preocupamos solamente por monitorear el nivel de error aceptable (confiabilidad).
La característica básica para el procesamiento en lote es que todos los vectores son presentados de una sola vez, actualizando los pesos sinápticos solamente al final de cada época. Denomínase época a la presentación de todo el conjunto de datos de entrenamiento de la red.
Como conjunto de entrenamiento utilizamos una base de datos de 200 elementos distribuidos uniformemente entre las clases de estímulos y locutores, y como Conjunto de prueba 460 elementos distribuidos aleatoriamente entre las clases de estímulos, incluyendo, los estímulos contaminados. En consecuencia, un determinado locutor podría comparecer con más estímulos dentro de una clase que otro locutor.
A continuación pretendemos demostrar cuantitativamente los resultados del experimento.
Inicialmente realizaremos una comparación entre las tres RNA (RNA2, RNA3, RNA5), entrenadas por nosotros y diferenciadas por el número de familias Wavelet.
Para el Conjunto de Entrenamiento, todos obtuvieron 100% de aciertos en la base de datos de prueba. Los datos que presentamos a seguir están basados en la base de entrenamiento.

Principales características entre las Redes Neurales.

	RNA2	RNA3	RNA5
Número de Familias	2	3	5
Número de filtros por familia	5	5	5
Total de filtros	10	15	25
Valores de los Coeficientes (de la madre para las niñas)	15, 15, 31, 63, 127	15, 15, 31, 63, 127	15, 15, 31, 63, 127
Tiempo de entrenamiento	35 minutos	65 minutos	80 minutos

Tabla 1 - Parámetros de los filtros y tiempo de entrenamiento por las diferentes RNA.

Resultados generales de los aciertos.

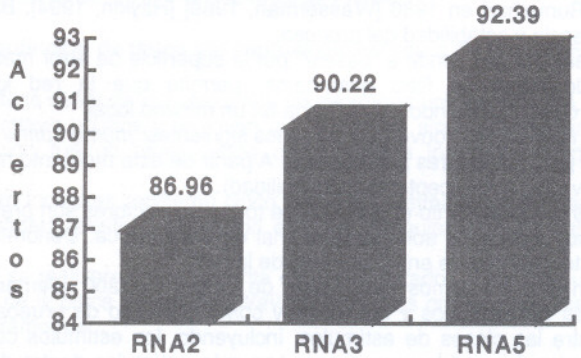

Grafico 1. Comparacion entre las Redes Neuronales (Datos contaminados+Datos no contaminados)

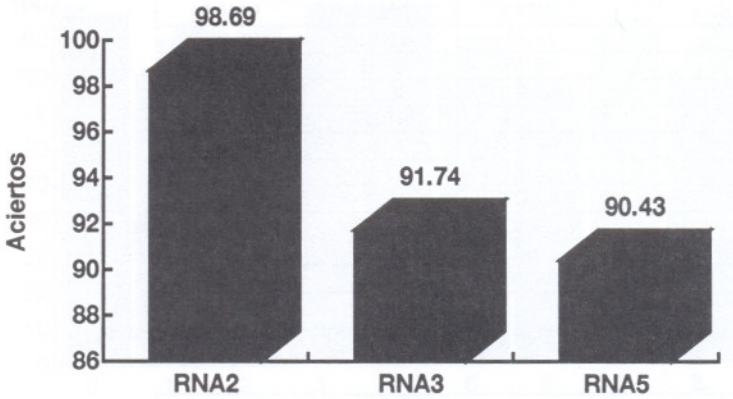

Grafico 2. Comparación entre las Redes Neuronales
(Datos no contaminados)

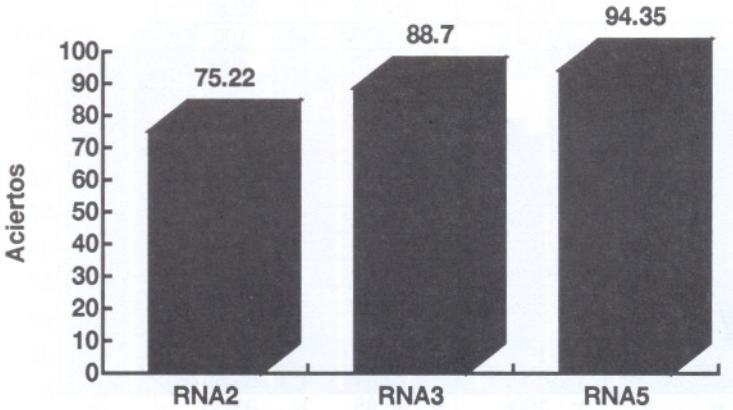

Grafico 3. Comparación entre las Redes Neuronales
(Datos contaminados)

Resultados en función de los estímulos de voz (Valor porcentual)

Estímulos	RNA2	RNA3	RNA5
0	97.83	95.65	95.65
1	80.43	89.13	93.48
2	82.91	89.13	91.30
3	89.13	97.83	97.83
4	95.65	97.83	97.83
5	84.78	89.13	89.13
6	82.61	73.91	78.26
7	89.13	93.48	89.13
8	80.43	86.96	93.48
9	86.96	89.13	97.83

Tabla 2- Datos generales de aciertos (Datos contaminados + no contaminados)

Estímulos	RNA2	RNA3	RNA5
0	95.65	95.65	95.65
1	60.87	82.61	86.97
2	69.56	82.61	95.65
3	78.26	95.65	100
4	86.97	95.65	100
5	69.56	86.97	95.65
6	69.56	73.91	86.97
7	78.26	91.30	82.61
8	65.21	91.30	100
9	73.91	86.97	100

Tabla 3- Datos contaminados.

Estímulos	RNA2	RNA3	RNA5
0	100	95.65	95.65
1	100	95.65	100
2	95.65	95.65	86.97
3	100	100	95.65
4	100	100	95.65
5	100	86.97	82.61
6	95.65	73.91	69.56
7	100	95.65	95.65
8	95.65	82.61	86.96
9	100	91.30	95.65

Tabla 4- Datos no contaminados.

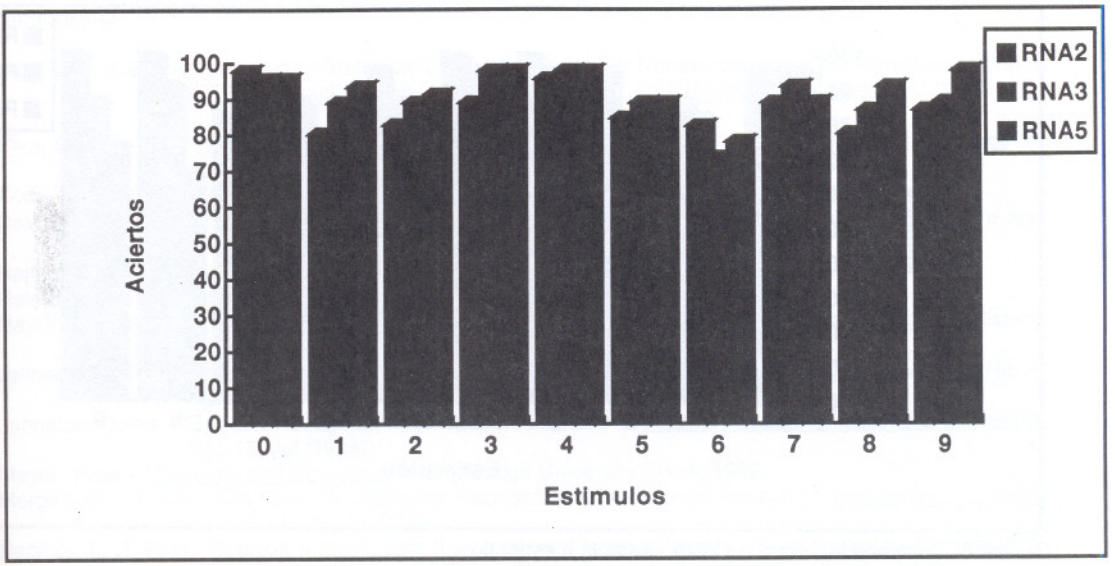

Gráfico 4- Comparación entre las RNA estímulo a estímulo.
(Datos contaminados + no contaminados)

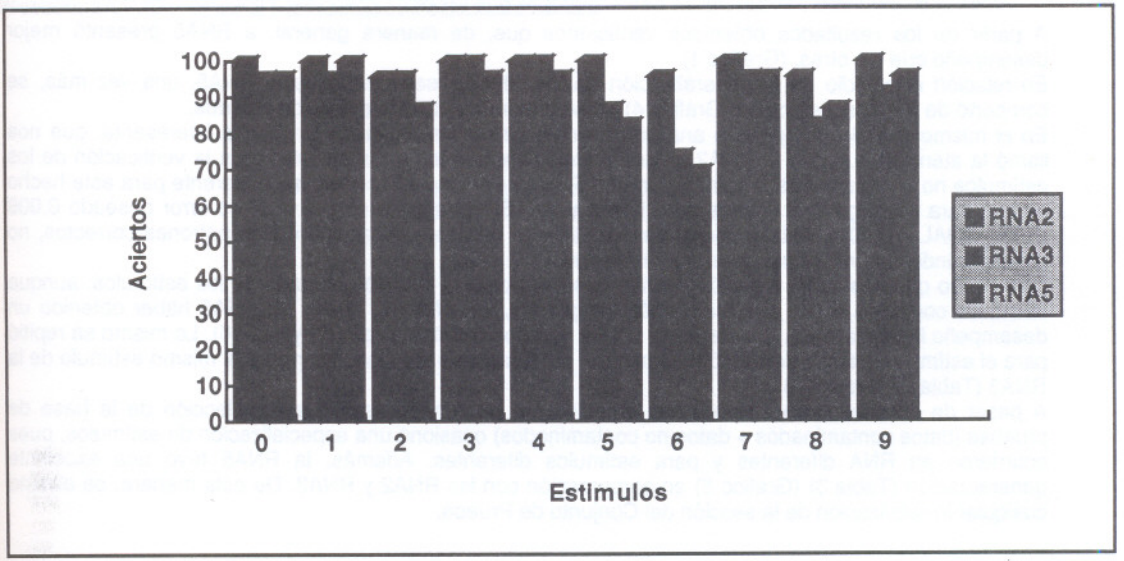

Gráfico 5- Comparación entre las RNA estímulo a estímulo.
(Datos no contaminados).

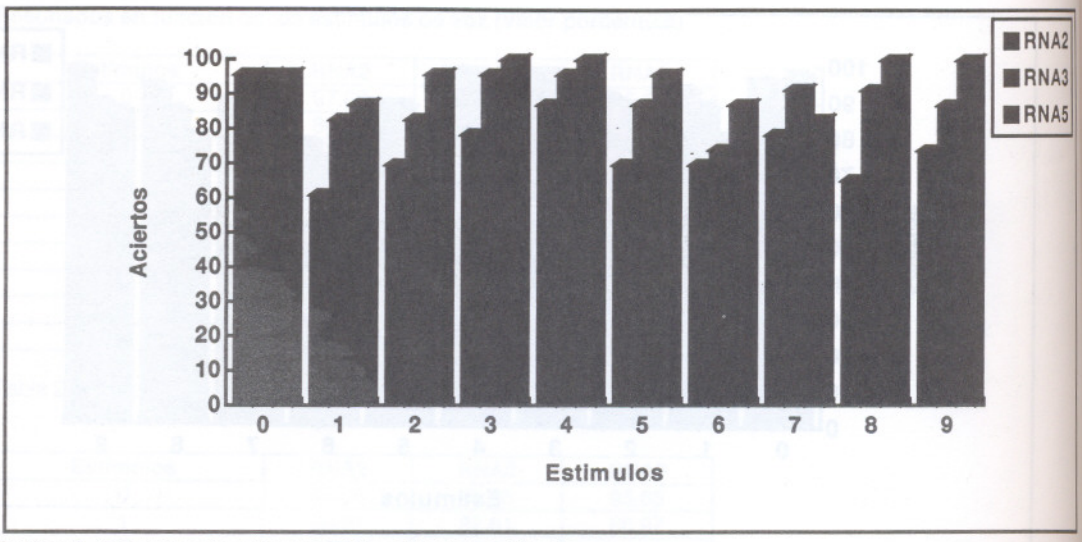

Gráfico 6- Comparación entre las RNA estímulo a estímulo.
(Datos contaminados)

6. Análisis general de los resultados obtenidos y conclusiones.

A partir de los resultados obtenidos verificamos que, de manera general, a RNA5 presentó mejor desempeño que las otras. (Gráfico 1)
En relación al estudio de la generalización de las RNA, observamos que la RNA5, una vez más, se comportó de forma satisfactoria (Gráfico 4) presentando una razón de 90% de aciertos.
En el mismo estudio, fue posible analizar diversas clases de estímulo. Un hecho interesante, que nos llamó la atención, fue que la RNA2 se comportó mejor que las otras en relación a la verificación de los estímulos no contaminados (Tabla 4) (Gráfico 5). Una explicación aceptable y coherente para este hecho es que para conseguir los valores para RNA3 y RNA5, trabajamos con el valor de error deseado 0.008 (ERR_GOAL = 0.008). Esta situación llevó a RNA2 a una especialización en los patrones correctos, no generalizándose para los patrones contaminados.
Otro hecho que nos llamó atención cuando pasamos para el estudio separado de los estímulos; aunque RNA5 se comportó mejor que las demás en general, se refiere al hecho de RNA5 haber obtenido un desempeño insatisfactorio, en relación a RNA2, para los estímulos cero (0) y seis (6). Lo mismo se repitió para el estímulo siete (7) de RNA5, que tuvo un porcentaje de acierto menor para el mismo estímulo de la RNA3 (Tabla 2) (Gráfico 4).
A pesar de observar estas situaciones notables, no podemos inferir que la selección de la base de pruebas (datos contaminados + datos no contaminados) ocasionó una especialización de estímulos, pues ocurrieron en RNA diferentes y para estímulos diferentes. Además, la RNA5 tuvo una excelente generalización (Tabla 3) (Gráfico 6) en comparación con las RNA2 y RNA3. De esta manera, se elimina cualquier inviabilización de la sección del Conjunto de Prueba.

Bibliografía

Bahl, L.R. et al - "A Tree- Based Statistical Language Model For Natural Language Speech Recognition" - IEEE Transactions on Acoustics, Speech, And Signal Processing, Vol. 37, No. 7, July/1989.

Chui, Charles K. - "WAVELETS ANALYSIS AND ITS APPLICATIONS - An Introduction to Wavelets " - Academic Press, Vol. 1, 1992.

Cody, Mac. - "The Fast *Wavelet* Transform " - Dr. Dobbs journal, AP/1992.

Daubechies, I. - "Orthonormal Bases of Compactly Supported Wavelets" -Communications on Pure and Applied Mathematics, Vol. XLI: 909-996, 1988.

Halmos, P. R. - "Finite-Dimensional Vector Space"- Springer-Verlag New York, 1970.

Halmos, P. R. - "A Hilbert Space Problem Book " - Springer-Verlag New York, 1992.

Haykin, Simon - "Neural Networks - A Comprehensive Foundation" - Macmillan College Publishing Company, Inc., 1994.

Jelinek, F. - "Continuos Speech Recognition by Statistical Methods" - Proceedings of IEEE, Vol. 64, No. 4, April/1976.

Johnston, Denis - "Speech Recognition For Speech Services" - Britsh Telecommunications Engineering, Vol. 13, Jul./1994.

Meyer, Yves - "Wavelets and Operators " - Cambridge University Press, 1992.

Morgan, N. - "Neural Networks For Statistical Recognition Of Continuos Speech " - Proceedings of IEEE, Vol. 83, No. 5, May/1995.

Rabiner, L. R. Et al, "Speaker Independent Recognition of Isolated Words Using Clustering Techniques ", IEEE Trans. Acoustics, Speech, Signal Proc., ASSP -27:336-349, Aug./1979.

Vetterli, M.; Kovacevic, J. - "Wavelets and Subband Coding"- Prentice Hall PTR, Englewood Cliffs, New Jersey,1995.

Wasserman,P. D. - "Neural Computing - Theory and Practice" - Van NostrandND , Reinhold, NY, 1989.