ISSN: 1692-7257 - Volumen 1 - Número 19 - Año 2012

Revista Colombiana de Tecnologías de Avanzada

Recibido: 17 de enero de 2011
Aceptado: 12 de marzo de 2011

MORPHOLOGICAL NEURAL NETWORK IMPLEMENTATION OF RECOGNITION OF SINGLE WORDS

APLICACIÓN DE REDES NEURONALES MORFOLÓGICAS AL RECONOCIMIENTO DE VOCABLOS SIMPLES

Ing. Luis Fernando. Gélvez R, PhD. José Orlando Maldonado Bautista

Universidad de Pamplona, Facultad de Ingenierías y Arquitectura.
Grupo de Investigación en Ciencias Computacionales
Ciudadela Universitaria. Pamplona, Norte de Santander, Colombia.
E-mail: luisfgelvezr@unipamplona.edu.co, orlmaldonado@gmail.com

Abstract: In this Project the more used methods for the voice features extraction that give a good description of the phoneme regardless of the speaker are researched. Also traditional speech recognition techniques are examined, specially, those directed at simple words recognition in order to establish a guideline under which morphologic neural networks can be evaluated as voice recognition technique.

Keywords: Speech recognition, machine learning, morphological neural network.

Resumen: En este trabajo se realiza un estudio de los métodos más utilizados para la extracción de características de voz que permitan obtener una buena descripción de los fonemas independientemente del hablante. Se examinan también, las técnicas tradicionales en el reconocimiento de habla, en especial, las orientadas al reconocimiento de vocablos simples para crear un marco de referencia bajo el cual se pueda evaluar el rendimiento de las redes neuronales morfológicas como técnica de reconocimiento de voz.

Palabras clave: Reconocimiento de voz, aprendizaje automático, redes neuronales morfológicas.

1. INTRODUCCIÓN

El reconocimiento del habla es un área que por años ha sido objeto de investigación debido a sus aplicaciones, entre las cuales destacan el dictado automático, el control por comandos o el apoyo a personas discapacitadas. Numerosas técnicas han sido propuestas para mejorar la comunicación por voz entre el ser humano y el ordenador, todas utilizando enfoques variados y algunas de ellas con resultados significativos, como el uso de Máquinas de Vectores de Soporte o los conocidos Modelos Ocultos de Markov. En aras de generar un aporte al estudio de reconocimiento del habla, se ha propuesto en este trabajo el uso de las Redes Neuronales Morfológicas como una nueva técnica de reconocimiento de voz. Hasta el momento, estas redes sólo han sido empleadas en el análisis de imágenes, donde los resultados obtenidos han mostrado sus propiedades como memorias de reconstrucción perfecta.

En la primera parte de este estudio se ha hecho énfasis en la selección de técnicas de caracterización que permitan describir con claridad la voz desde un punto de vista fonético. Esto se ha hecho con el fin de que los algoritmos de clasificación tomados como referencia y la propuesta de las redes neuronales morfológicas puedan arrojar resultados considerables en el

reconocimiento de vocablos independientemente del locutor. Entre las técnicas de caracterización que se tuvieron en cuenta en el presente trabajo se encuentran las más comunes como la extracción del pitch o la codificación predictiva lineal, y otras mejor conocidas por su capacidad descriptora como los coeficientes cepstrales en la frecuencia de Mel y la descomposición Wavelet.

Una vez obtenida una buena técnica descriptora de la voz, se entra en la segunda fase de este trabajo, la cual hace referencia al estudio de las redes neuronales morfológicas frente a los métodos tradicionales de reconocimiento de voz. Esta propuesta aparece como una alternativa no lineal a las ya conocidas memorias asociativas, y en especial, a las memorias de Hopfield, cuyo sistema algebraico subyacente está basado en las operaciones de adición y de multiplicación. El modelo alternativo expuesto por las redes neuronales morfológicas sustituye este sistema por una estructura basada en el álgebra de Lattice, donde las operaciones se basan en máximos y mínimos de sumas. De esta forma, al ser un sistema no lineal, se obtiene gran robustez frente al ruido aditivo y sustractivo que puedan presentar las muestras.

Inicialmente las redes neuronales morfológicas propuestas en 1996 por Gerhard X. Ritter surgieron como una solución al problema de reconstrucción, y hasta a la fecha sólo han sido aplicadas en el análisis de imágenes. En este estudio se ha propuesto un clasificador combinando algunas características de los métodos tradicionales con las propiedades de reconstrucción perfecta de las Memorias Asociativas Morfológicas (MAM), que luego es aplicado al problema de reconocimiento de habla, y en particular, al reconocimiento de vocablos simples.

Cerrando este trabajo se presenta una comparación entre los resultados obtenidos con los métodos tradicionales de clasificación y la propuesta de clasificador híbrido. Se han planteado también algunas variantes del esquema general del clasificador y sus capacidades frente a distintas circunstancias de caracterización de las muestras de señales de voz.

2. CARACTERIZACIÓN DE LA SEÑAL DE VOZ

La voz es una onda sonora producida por las cuerdas vocales, y alterada por los demás elementos que conforman el aparato fonador humano; la variedad en la voz humana depende de la configuración del tracto vocal de cada persona. Tal complejidad en la producción de la voz requiere extraer características que describan de forma eficiente la señal para su posterior análisis. Las características que se tuvieron en cuenta para describir la señal incluyen la energía, cruces por cero, frecuencia fundamental, cepstrum, entre otras.

2.1 Base de datos

Se ha elaborado dos bases de datos para las pruebas, tomadas de personas de distintas regiones de Colombia con edades entre los 15 y 60 años; se tiene un total de 15 hablantes para la primera base de datos y 42 para la segunda. La primera base de datos tiene el objeto de medir la efectividad de los algoritmos implementados frente a fonemas vocálicos sostenidos (/a/, /e/, /i/, /o/, /u/); la duración de las vocales sostenidas tiene en promedio una duración de 700 milisegundos. La segunda base de datos consiste en palabras sencillas, en este caso los números del 1 al 10. Las muestras de ambas bases de datos han sido capturadas bajo una frecuencia de muestreo de 22050 Hz en un solo canal (mono) a 16 bits por muestra, suficiente para tener acceso al espectro de frecuencias de la voz humana (y sus patologías) y poder aislar ciertos ruidos que se introducen en el proceso de grabación. El nivel de volumen de la base de datos de vocablos fue normalizado debido a que unas muestras presentan niveles de intensidad muy bajos que pueden dificultar la caracterización. Los archivos fueron almacenados en formato *wav* debido a que éste es el formato soportado por la herramienta MATLAB. A las muestras capturadas no se les aplicó técnicas especiales de sustracción de ruido, esto con el fin de medir la robustez de los clasificadores utilizados frente a condiciones complejas.

2.2 Técnicas de caracterización

Se han aplicado técnicas de caracterización basadas en tiempo-frecuencia, energía, correlación de muestras.

Energía: En las señales discretas la energía para N muestras de la señal se define como:

$$E = \sum_{m=0}^{N-1} x(m)^2 \qquad (1)$$

Los cambios de energía en la señal de voz están directamente relacionados con la variación de la presión subglotal y la forma del tracto vocal. La energía resulta de gran utilidad a la hora de identificar sonidos sonoros y sordos en la señal, debido a que los primeros presentan valores más altos de energía que los segundos (Wikipedia, 2010-1).

- **Cruces por cero:** Un cruce por cero puede definirse como la ocurrencia en la cual dos muestras consecutivas tengan distinto signo (entre las dos muestras la señal tendrá que tomar el valor de cero) o cuando una muestra equivale a cero. El momento de cruce por cero puede expresarse como sigue:

$$z = \sum_{m=0}^{N-1} \left| \mathrm{sgn}[x(m)] - \mathrm{sgn}[x(m-1)] \right| \qquad (2)$$

- ***Pitch:*** Para la extracción del *pitch* o frecuencia fundamental de las señales de voz se ha utilizando una implementación basada en la razón de subarmónicos a armónicos (Sun, 2000). Partiendo del primer y segundo armónico (f_1 y f_2) la razón de subarmónicos a armónicos (SHR) se puede calcular como:

$$SHR = 0.5 \frac{DA(\log f_1) - DA(\log f_2)}{DA(\log f_1) + DA(\log f_2)} \qquad (3)$$

Donde *DA* es la función diferencia definida por:

$$DA(\log f) = SUMA(\log f)_{par} - SUMA(\log f)_{impar} \qquad (4)$$

$$SUMA(\log f)_{par} = \sum_{n=1}^{N} LOGA(\log f + \log(2n)) \qquad (5)$$

$$SUMA(\log f)_{impar} = \sum_{n=1}^{N} LOGA(\log f + \log(2n-1)) \qquad (6)$$

- **Coeficientes MFCC:** los coeficientes MFCC son una variante de los coeficientes obtenidos mediantes análisis cepstrum debido a que las bandas de frecuencia son mapeadas a la frecuencia de Mel (Wikipedia, 2010-2). Los coeficientes MFCC son calculados de la siguiente forma: *i*) se aplica una función ventana *ii*) se calcula la transformada de Fourier *iii*) se hace un mapeo de las frecuencias obtenidas a la escala de Mel *iv*) se obtiene el logaritmo de las frecuencias mapeadas *v*) Se obtiene la transformada discreta del coseno. El mapeo de las frecuencias se lleva a cabo mediante la siguiente expresión:

$$m = \frac{1000}{\ln(1+1000/700)} \ln(1 + f/700) \qquad (7)$$

- **Coeficientes LPC:** El análisis LPC se centra en la estrecha correlación que puede existir entre muestras consecutivas, preferiblemente entre tramos sonoros (Wikipedia, 2010-3). Entonces, la muestra actual se aproxima (es predicha) mediante la combinación lineal de las muestras anteriores:

$$\tilde{s}(n) = -\sum_{k=1}^{p} a_k s(n-k) + e(n) \qquad (8)$$

Donde a_k son los coeficientes de aproximación, *s(n)* la señal real, *s(n)* la señal predicha, *e(n)* el error de predicción lineal y *p* el orden de predicción del filtro polinomial.

- **Descomposición *wavelet*:** La transformada wavelet genera una representación de la señal en términos de una onda finita (*wavelet* madre) en sus versiones trasladas y escaladas (Sepúlveda, 2004). La transformada wavelet en su forma continua se define como:

$$Wf(u,s) = \left\langle f(t), \Psi_{u,s} \right\rangle = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \Psi\left(\frac{t-u}{s} \right) dt \qquad (9)$$

2.3 Selección de características

Al caracterizar el conjunto de muestras de voz, se genera una gran cantidad de variables, por lo que en algunos casos es necesario reducir la dimensionalidad para mejorar el tiempo de cómputo y a la vez condensar la variabilidad de los datos en un conjunto de datos más pequeño; un conjunto de muestras con demasiados parámetros hace que algunos algoritmos de clasificación no puedan identificar discrepancias. En este trabajo se ha aplicado análisis de componentes de principales en dos casos: el primero cuando una técnica de caracterización genera demasiadas variables y el segundo cuando se realiza combinación de varias características.

3. CLASIFICACIÓN DE PATRONES DE VOZ

3.1 Clasificadores clásicos

Para la fase de clasificación se ha dispuesto de algunos métodos tradicionales, con el objeto de tener un referente para medir luego la efectividad de la propuesta de las redes neuronales morfológicas.

Entre las técnicas utilizadas están: algoritmo de agrupamiento *k*-medias utilizando varias medidas de distancia para medir la similitud de las muestras con los prototipos de clase (distancia euclídea, distancia *city-block* (distancia de Manhattan), distancia de coseno y medida de correlación. También se han utilizado el análisis discriminante, una red neuronal tipo *feed-forward* con algoritmo *backpropagation*, y máquina de vectores de soporte (SVM).

3.2 Memorias asociativas morfológicas

En este estudio se ha propuesto el uso de memorias asociativas morfológicas (MAM) como clasificadores de señales de voz, pretendiendo aprovechar las propiedades de reconstrucción perfecta de dichas memorias. El objetivo de una memoria asociativa es recuperar un patrón de salida **y** a partir de un patrón de entrada **x**:

$$x \rightarrow [M] \rightarrow y \qquad (10)$$

Una asociación entre *k* patrones de entrada y salida puede ser expresada como:

$$\left\{ \left(x^k, y^k\right) \middle| \, k = 1, 2, \cdots, p \right\} \qquad (11)$$

En general, el problema de las memorias asociativas puede resumirse en dos fases: una fase de aprendizaje y una fase de recuperación. En la fase de aprendizaje la memoria es generada a partir de un conjunto de asociaciones de patrones de entrada y salida; en este caso, cuando se da $\mathbf{x}^k = \mathbf{y}^k$? k ? {1, 2,... ,p}, la memoria es heteroasociativa y en el caso contrario donde xk ? $\mathbf{y}^k$? k ? {1, 2,..., p}, la memoria es autoasociativa.

A diferencia de las redes neuronales tradicionales, cuyos cálculos utilizan sumas de productos (**R**, +, x), las memorias asociativas morfológicas propuestas por Ritter en 1996 se basan en un sistema algebraico que utiliza máximos y mínimos de sumas (**R**, ∨,∧, +, +'). Los símbolos ∨ y ∧ denotan las operaciones máximo y mínimo respectivamente con las siguientes condiciones:

$$a \vee (-\infty) = (-\infty) \vee a = a \quad \forall a \in \Re_{-\infty} \qquad (12)$$

$$a \wedge (-\infty) = (-\infty) \wedge a = a \quad \forall a \in \Re_{\infty} \qquad (13)$$

El modelo para la red neuronal morfológica utilizando el sistema ($\mathbf{R}_{-8}$, ∨, +) se rige por las siguientes expresiones:

$$t_i(t+1) = \bigvee_{j=1}^{n} a_j(t) + w_{ij} \qquad (14)$$

$$a_i = (t+1) = f\left(t_i(t+1) - q_i\right) \qquad (15)$$

Donde a_i (*t*+1) representa el valor de la *j*-ésima neurona en el tiempo *t*, *n* representa el número de neuronas en la red, w_{ij} es el valor de la conexión sináptica entre la *i*-ésima y *j*-ésima neurona. $?_i(t+1)$ es el siguiente efecto total de entrada en la *i*-ésima neurona, $?_i$ un umbral, y *f* el siguiente estado de la función. De forma análoga, para el sistema ($\mathbf{R}_8$,∧, +') el siguiente efecto total de entrada en la *j*-ésima neurona puede ser calculado como:

$$t_i(t+1) = \bigwedge_{j=1}^{n} a_j(t) + w_{ij} \qquad (16)$$

El cómputo total de la red morfológica puede ser expresado en forma de matriz, mediante el producto *máx* y el producto *mín* respectivamente como sigue:

$$T(t+1) = W \vee a(t) \qquad (17)$$

$$T(t+1) = W \wedge a(t) \qquad (18)$$

Partiendo de la definición básica de una memoria asociativa, y dados los vectores $\mathbf{x} = (x_1,\ldots,x_n)' \in \mathbf{R}^n$ y $\mathbf{y} = (y_1,\ldots,y_m)' \in \mathbf{R}^m$, una memoria asociativa morfológica que recuperará el vector **y** a partir de la llave **x** está dada por:

$$w = y \vee (-x)' = \begin{pmatrix} y_1 - x_1 & \cdots & y_1 - x_n \\ \vdots & \ddots & \vdots \\ y_m - x_1 & \cdots & y_m - x_n \end{pmatrix} \qquad (19)$$

Una expresión para obtener la memoria asociativa morfológica óptima que recupera el vector $\mathbf{y}^?$ cuando es presentado el vector $\mathbf{x}^?$ para ? = 1,…,*k* está dada como sigue:

$$W = \bigwedge_{x=1}^{k} \left(y^x \vee \left(-x^x\right)'\right) \qquad (20)$$

Las propiedades de reconstrucción perfecta se pueden evidenciar mediante la siguiente expresión:

$$W \vee x = \begin{pmatrix} \bigvee_{i=1}^{n} (y_1 - x_i + x_i) \\ \vdots \\ \bigvee_{i=1}^{n} (y_m - x_i + x_i) \end{pmatrix} = y \qquad (21)$$

3.3 Clasificador basado en MAM

Se han planteado varios esquemas de clasificación que utilizan algunas propiedades de clasificadores clásicos como el *k-means* o el *knn* (*k-nearest neighbors*). Se definió un primer modelo de clasificación y partiendo de los resultados obtenidos se diseñaron otras alternativas para solucionar el problema de clasificación de vocablos simples.

El modelo de clasificación se basa en la idea de que la memoria asociativa morfológica puede relacionar cualquier patrón de entrada con un una versión generalizada de alguna de las clases existentes. Esto se puede lograr calculando prototipos de clase a partir del promedio de las muestras de cada clase:

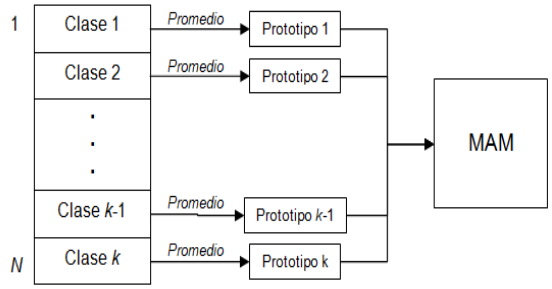

Fig. 1. Cálculo de prototipos de clase por promedios

Al presentar a la memoria una muestra de voz, se obtendrá un patrón de salida que luego será comparado con alguno de los prototipos de clase, utilizando para ello una medida de distancia. El prototipo con el cual guarde la menor distancia será aquel que determine la clase a la cual pertenece el patrón de entrada presentado a la memoria.

Partiendo de este modelo de clasificación se diseñaron varios experimentos donde se plantean alternativas para determinar las capacidades de las MAMs en el problema de clasificación de señales de voz:

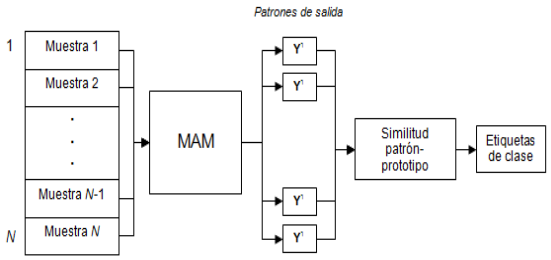

Fig. 2. Asignación de etiquetas de clase por medio de comparación patrón-prototipo

- ***Experimento 1:*** Calcular una matriz de características tal que contenga las variables de todas las técnicas de caracterización utilizadas y seguidamente aplicar un análisis de componentes principales. La idea es que al tener una gran cantidad de variables se puede disponer de un conjunto suficiente para recoger la máxima variabilidad de las muestras y por tanto, el clasificador pueda identificar más diferencias en las observaciones.

- ***Experimento 2:*** Aplicar el mismo procedimiento del experimento 1 pero esta vez incluyendo la variable del tiempo, esto es, incluir dentro de la matriz de características la duración de la señal. Este valor en una señal muestreada equivale al número de puntos del vector que la representa, pero para simplificar este valor se hace la conversión a milisegundos dividiendo el número de puntos entre la frecuencia de muestreo y multiplicando por mil.

- ***Experimento 3:*** Realizar un entrenamiento con pocas clases, es decir, partir del aprendizaje y reconocimiento de dos vocablos, e ir incrementando la cantidad de ellos hasta llegar al número total de clases existentes, esto con el fin de determinar hasta qué número de clases puede soportar el modelo propuesto.

- ***Experimento 4:*** Aplicar un entrenamiento basado en validación cruzada para 3 divisiones del conjunto de muestras; en este caso, cada clase se segmenta en 3 partes: el entrenamiento de la memoria se realiza con dos de esas 3 divisiones y la fase de recuperación o prueba de la memoria se lleva a cabo con la parte restante, es decir con 1/3 de las muestras. Lo anterior se ilustra en la Fig. 3:

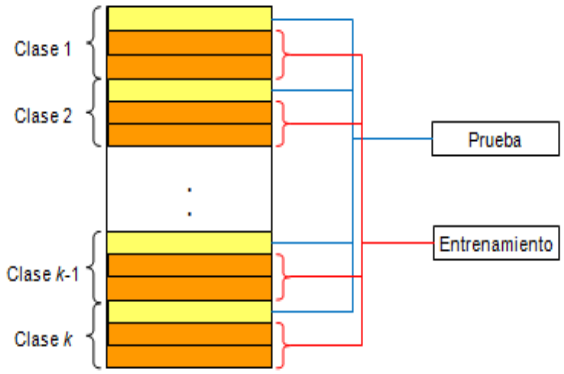

Fig. 3. Entrenamiento por validación cruzada para tres divisiones

- *Experimento 5:* Utilizar otras medidas de tendencia central para calcular los prototipos de clase que son utilizados para generar la MAM.

- *Experimento 6:* Implementar una de las propiedades del *k-medoids* que consiste en calcular el promedio de cada clase, y para cada uno de estos valores, buscar la muestra más cercana. En este sentido, no es un promedio de clase el que representa la clase, sino la muestra más parecida a este promedio.

- Experimento 7: En este experimento se ha rediseñado gran parte del esquema original del clasificador. El siguiente pseudocódigo ilustra el nuevo modelo propuesto basado en entrenamiento *leave-one-out*:

Para i=1 hasta numeroDeObservaciones

1. Sacar una copia de la matriz de características vs.
2. Suprimir la fila i (observación actual) en la copia de la matriz VS para realizar un entrenamiento sin la observación actual.
3. Crear la memoria W con la versión traspuesta de la matriz resultante.
4. Presentar la fila i de la matriz VS como un patrón de entrada a la memoria W para obtener el patrón de salida Y.
5. Calcular la distancia del patrón Y a todas las observaciones de la matriz VS para generar un vector de distancias de longitud igual al número de observaciones.
6. Calcular el valor mínimo del vector de distancias y su posición h respectiva dentro del arreglo. Esta posición indica la observación que más se parece al patrón de salida Y.

Determinar a qué clase corresponde la observación h como sigue:

```
para k = 1 hasta numeroDeClases
limiteFinal=k*numeroDeObservaciones/numeroDeClases;
                    si (h <= limiteFinal)
                            etiqueta(i) = k;
                            romperCiclo;
                    fin
            fin
    fin
```

El objetivo es crear la MAM con todas las muestras menos la muestra que se pretende recuperar; esto implica que la muestra que se dejó por fuera será un nuevo elemento para la MAM. El patrón de salida es utilizado para buscar la muestra más parecida al patrón de entrada (se ha utilizado la distancia euclidiana) y luego, a través de la organización por defecto que posee la matriz de características, determinar la clase *k* a la que pertenece la muestra que se ha dejado por fuera del conjunto de entrenamiento.

4. RESULTADOS

Se presenta una comparación entre los resultados obtenidos por los clasificadores tradicionales y la propuesta de clasificador híbrido basado en memorias asociativas morfológicas.

Se ha medido la efectividad de los clasificadores frente a la base de datos de fonemas vocálicos y a la base de datos de vocablos. Los resultados para fonemas vocálicos de muestran a continuación:

Tabla 1. Comparativa de clasificadores en reconocimiento de fonemas vocálicos

Clasificador	Porcentaje de acierto
k-means	82.4
Classify (análisis discriminante)	85
Red feed-forward	60
Máquina de vectores de soporte	80
Clasificador híbrido basado en MAM	91

Se ha conseguido que el clasificador híbrido tenga una efectividad tal que ha mejorado inclusive el resultado obtenido con uno de los mejores clasificadores tradicionales de que se dispone; por el contrario, las redes *feed-forward* han mostrado ser no tan efectivas en este problema de reconocimiento.

A continuación se muestran los resultados obtenidos por el clasificador híbrido en los experimentos mencionados en la sección anterior. La matriz de características utilizada incluye coeficientes MFCC y coeficientes *wavelet*.

En el experimento 4, para la rotación del conjunto de prueba el clasificador híbrido obtuvo 54%, 56% y 44% de efectividad. Los resultados del clasificador híbrido usando varias medidas de tendencia central se muestran en la Tabla 4:

Tabla 2. Efectividad del clasificador híbrido en algunos experimentos

Experimento	Porcentaje de acierto
1	50
2	47
6	40
7	100

Tabla 3. Acierto del clasificador hibrido para varios números de clases (Experimento 3)

Número de clases	Porcentaje de acierto
2	100
3	93
4	81
5	78
6	75
7	67
8	63
9	58
10	56

Tabla 4 Porcentaje de acierto utilizando algunas medidas de tendencia central

Medida de tendencia central	Porcentaje de acierto
Media aritmética	56
Media geométrica	51
Media armónica	23
Media cuadrática	10
Mediana	43

Con respecto al problema de reconocimiento de vocablos simples es necesario aclarar que la comparativa se ha realizado paras varias combinaciones de características con el fin de tener un marco de referencia más amplio bajo el cual se puedan determinar las capacidades tanto de los clasificadores tradicionales como la basada en MAMs. En la tabla 5 se muestran los resultados de clasificación obtenidos con la base de vocablos:

Tabla 5. Comparativa de clasificadores en reconocimiento de vocablos simples

	Clasificador		
	k-means	*Classify*	*Híbrido*
Combinación	**Porcentaje de acierto**		
Energía + cruces por cero	46	65	100
Energía + MFCC	25	87	99.8
Energía + wavelet	28	69	99.7
Cruces por cero + MFCC	44	90	100
Cruces por cero + wavelet	33	76	100
MFCC + wavelet	45	74	99.7
Energía + cruces pro cero + MFCC	48	91	100
Energía + MFCC + wavelet	27	86	99.7
Energía + cruces por cero + wavelet	46	80	100
Cruces por cero + MFCC + wavelet	39	89	100
Energía + cruces por cero + MFCC + wavelet	48	90	100

El clasificador *k-means* no consigue ser efectivo para reconocer al menos el 50% de los vocablos, debido a que la forma como se calculan los prototipos de clase es poco eficaz para el tipo de características que representan cada muestra de voz. Los mejores resultados conseguidos por el clasificador de análisis discriminante (*Classify*) se obtuvieron para la combinación de características energía + CPC + MFCC y ésta misma añadiendo los coeficientes wavelet; no obstante, se observa una pérdida de variabilidad al añadir los coeficientes wavelet. El clasificador híbrido diseñado en el experimento 7 ha mostrado tener la máxima efectividad, e inclusive obtener un reconocimiento del 100% en 7 de las 11 combinaciones de características.

5. CONCLUSIONES

Los clasificadores basados en promedios de muestras resultan ser poco efectivos cuando se trata con señales de voz pertenecientes a vocablos. En el caso de fonemas vocálicos las muestras caracterizadas tienen más variabilidad y una medida de tendencia central puede ser utilizada para generar prototipos de clase.

Los clasificadores como redes neuronales feed-forward o máquinas de vectores de soporte requieren mucho tiempo de entrenamiento para poder converger hacia una solución óptima, en especial cuando se tratan numerosas muestras con muchas variables caracterizadoras.

El clasificador basado en memorias asociativas morfológicas converge rápidamente a una solución óptima independientemente del esquema de clasificación que presente.

Un clasificador basado en memorias asociativas morfológicas está en capacidad de reconocer una muestra de voz de una clase al almacenar en la MAM un número significativo de muestras de dicha clase.

REFERENCIAS

Makhoul, J. (2006). *Speech Procesing at BBN*, BBN Technologies.

Molina, C., Becerra, N., Huenupán, F, Garretón C. Y Wuth J. (2010). "Maximum Entropy-Based Reinforcement Learning Using a Confidence Measure in Speech Recognition for Telephone Speech". IEEE Trans. On audio,

Speech, and Language Processing, **Vol. 18**, No. 5.

Bou-Ghazale, S. Y. Hansen, J. "A Comparative Study of Traditional and Newly Proposed Features for Recognition of Speech Under Stress". IEEE Trans. On Speech Audio Processing, **Vol. 8**, pp. 429-442.

Fandiño, D. (2005). *Estado del arte en el reconocimiento automático de voz*. Universidad Nacional de Colombia.

Sanz P. Y Vera De Payer, E. (S.F). *Reconocimiento de comandos de voz aplicado a sistema robótico médico*, Universidad Nacional de Córdoba.

Zañartu, M. (2003). *Aplicaciones del análisis acústico en los estudios de la voz humana.* Seminario Internacional de Acústica, Universidad Pérez Rosales.

Jones, R. (2009). *Inteligibilidad del habla*. Cetear.

Sepúlveda, F. Y Castellanos, G. (2004) "Estimación de la frecuencia fundamental de señales de voz usando Transformada Wavelet". Scientia Et Technica.

Díaz, J., Sapienza, C., Rothman, H., Y Natour, Y. (2003). *Algoritmo robusto para la detección de la frecuencia fundamental en la voz basado en el espectrograma.* Ingeniería UC, Universidad de Carabobo, pp. 7-16.

Cáceres, J. (2007). *Transformada corta de Fourier y ventanas*. Stanford University.

Cortés, J., Medina, F., Y Chávez, J. (2007). "Del análisis de Fourier a las Wavelets". Scientia Et Technica.

Ravelli, E., Richard, G., Y Daudet, L. (2010). "Audio Signal Representations for Indexing in the Transform Domain". IEEE Trans. On Speech Audio Processing, **Vol. 18**, pp. 434-446.

Álvarez, A. (2001). *Algoritmos de extracción de características.* Universidad Politécnica de Madrid.

San Martin, y C., Carrillo, R. (2004). "Implementación de un reconocedor de palabras aisladas dependiente del locutor". Revista Facultad de Ingeniería U.T.A Chile, **Vol. 12**.

Terrádez, M. (S.F). *Análisis de componentes principales*. Universidad Abierta de Cataluña.

IBM SPSS. (2002). Guía para el análisis de datos.

Díaz, E. (2003). *Análisis discriminante*.

Galbiati, J. (2009). *Análisis discriminante*.

Barrón, R. (2006). *Memorias asociativas y redes neuronales morfológicas para la recuperación de patrones*. Instituto Politécnico Nacional, México D.F.

Ritter, G., Y Sussner, P. (1996). *An Introduction to Morphological Neural Networks*. Proceedings of the 13th International Conference on Pattern Recognition.

Sun, X. (2000). *A pitch determination algorithm based on subharmonic-to-harmonic ratio*. 6th International Conference of Spoken Language Processing.

Sepúlveda, F. (2004). *Extracción de parámetros de voz usando técnicas de análisis en tiempo-frecuencia*. Universidad Nacional de Colombia.

SITIOS WEB

Wikipedia. (2010-1). *Señal de voz*. http://es.wikipedia.org/wiki/Se%C3%B1al_de_voz. (26 de julio 2010).

Wikipedia. (2010-2). *Mel-frequency cepstrum*. http://en.wikipedia.org/wiki/Mel-frequency_cepstrum. (10 de agosto 2010).

Wikipedia. (2010-3). *Linear Predictive Coding*. http://en.wikipedia.org/wiki/Linear_predictive_coding. (12 de agosto 2010)