ENTRENAMIENTO DISCRIMINATIVO POR DISTANCIA DE MAHALANOBIS PARA DETECCIÓN DE PATOLOGÍAS DE VOZ

DISCRIMINATIVE TRAINING BASED ON MAHALANOBIS DISTANCE FOR PATHOLOGIC VOICE DETECTION

M. SARRIA-PAJA
Universidad Nacional de Colombia – Sede Manizales,mosarriap@unal.edu.co

G. CASTELLANOS-DOMÍNGUEZ
Universidad Nacional de Colombia – Sede Manizales,cgcastellanosd@unal.edu.co

Recibido para revisar Marzo 17 de 2009, aceptado Septiembre 8 de 2009, versión final Octubre 13 de 2009

RESUMEN: Este artículo presenta una técnica de entrenamiento discriminativo para Modelos Ocultos de Markov, orientado a identificación de patologías de voz. Esta técnica busca maximizar el área que encierra la Curva ROC (*Receiver Operating Characteristic*) ajustando los parámetros del modelo, empleando como función objetivo la distancia de Mahalanobis. Los resultados muestran que la técnica propuesta mejora significativamente la precisión en un sistema de clasificación comparado con otros criterios de entrenamiento. Los resultados son obtenidos empleando la base de datos de patologías de voz MEEIVL.

PALABRAS CLAVE: HMM, MLE, Entrenamiento discriminativo, patologías de voz, curva ROC.

ABSTRACT: This paper presents an approach that improves discriminative training criterion for Hidden Markov Models, and oriented to voice pathological identification. This technique aims at maximizing the Area under Curve of a Receiver Operating Characteristic curve by adjusting the model parameters using as objective function the Mahalanobis distance. The results show that the proposed technique outperforms significantly the accuracy in a classification system comparing with other training criteria. Results are provided using the MEEIVL voice disorders database.

KEYWORDS: HMM, MLE, Discriminative training, voice pathology, ROC curve.

1. INTRODUCCIÓN

Los Modelos Ocultos de Markov (*Hidden Markov Models – HMM*) han sido ampliamente utilizados en sistemas de reconocimiento de voz, especialmente en la solución de problemas tales como identificación o verificación de hablante, ubicándose como una herramienta estándar para modelar las variaciones estocásticas presentes en este tipo de señales [1]. Un problema de especial interés en aplicaciones biomédicas es la detección de patologías en señales de voz, donde el principal objetivo es generar herramientas de diagnóstico asistido mediante técnicas no invasivas [2]. El proceso automático de voz para detección de patologías tiene sus ventajas: el análisis es cuantitativo y no invasivo, permitiendo identificar y monitorear enfermedades del tracto vocal, y adicionalmente reducir costos.

Durante la fonación sostenida de vocales, la voz normal es una señal regular y cuasi-periódica, y cambios abruptos en su forma de onda se pueden percibir como posibles trastornos. Emplear medidas de distorsión clásicas complementadas con otro tipo de características dinámicas, como se ha señalado en algunos estudios [3], es una de las formas más eficientes de capturar la mayor

cantidad de información disponible en las señales acústicas, considerando también los cambios en su estructura temporal, permitiendo modelar de forma adecuada estos fenómenos. Este tipo de características combinadas con clasificadores dinámicos, (por ejemplo HMM), se han empleado en la detección de patologías de voz de forma satisfactoria [4].

El entrenamiento de los HMM implica el ajuste de los parámetros de un modelo, tal que se extraiga la máxima información de las secuencias de observación. Entre los métodos conocidos están el criterio basado en la estimación de máxima verosimilitud (*Maximum Likelihood Estimation* - MLE) [5], donde se optimiza la descripción del respectivo modelo para un conjunto dado de observaciones (*función de verosimilitud*), sin tener relación explícita con el rendimiento del clasificador, por lo cual este es un criterio de entrenamiento generativo. Por otro lado están los métodos de entrenamiento discriminativo, por ejemplo, la técnica de Máxima Información Mutua (*Maximum Mutual Information* – MMI) [6], donde se busca optimizar la probabilidad a posteriori de los datos de entrenamiento y, por lo tanto la separabilidad entre clases, o el criterio de Mínimo Error de Clasificación (*Minimum Classification Error* - MCE) [7] donde se minimiza el error de clasificación mediante la formulación de una función de error empírica.

En cuanto a las medidas de desempeño, en el caso de los sistemas de diagnóstico asistido, se ha sugerido el empleo de medidas de desempeño mucho más robustas que el error de clasificación o la precisión, por ejemplo, el empleo de la curva ROC *(Receiver Operating Characteristic curve)* [8], que tiene la capacidad de representar el desempeño global del sistema en diferentes puntos de operación, un indicador muy importante es el del área bajo la curva ROC ó ABC, lo cual ha llevado a la formulación de nuevos criterios de entrenamiento discriminativos que emplean la maximización del ABC como función objetivo.

Un ejemplo concreto de entrenamiento discriminativo, que optimiza el ABC ajustando los parámetros del modelo, se presenta en [9], conocido como *FOM training* y propuesto para ajustar los parámetros de modelos de mezclas de Gaussianas, mostrando que la capacidad discriminativa del sistema mejora significativamente. Sin embargo, no se presenta una función analítica directamente asociada a la misma ABC.

Otro enfoque propone integrar a la etapa de entrenamiento algunas métricas de interés además del error de clasificación (especificidad, sensibilidad)[10]. Criterio conocido como MfoM (*Maximal Figure of Merit*), y aunque trabaja sobre medidas muy relacionadas a la curva ROC, no optimiza una función relacionada directamente con la curva o construcción de la misma.

En estos dos trabajos el principal inconveniente es la ausencia de una función que esté relacionada directamente con la curva ROC o el ABC. Esta dificultad es la que no ha permitido que el enfoque haya sido formalizado, restringiendo su uso.

En este sentido y para superar este inconveniente, se propone emplear como criterio de entrenamiento discriminativo la optimización de una medida de distancia, que este mas relacionada al área que encierra la curva ROC, esto debido a que el ABC, es directamente proporcional a la separación que tienen las distribuciones de probabilidad para cada una de las clases, generadas a partir de los HMM. Por lo que se emplea como criterio de entrenamiento la optimización de la distancia de Mahalanobis, mediante una técnica basada en el cálculo de gradientes para ajustar los parámetros del modelo.

La comparación de los métodos de entrenamiento empleados (MLE, MMI, MCE y MFoM), con el método propuesto, se realiza sobre la base de datos de patologías de voz desarrollada por *The Massachusetts Eye and Ear Infirmary Voice Laboratory (MEEIVL).* Y los resultados obtenidos muestran que la técnica de entrenamiento propuesta mejora sustancialmente las técnicas de entrenamiento conocidas tanto generativas como discriminativas.

Este manuscrito esta estructurado de la siguiente manera: En la sección 2 se hace una revisión del estado del arte sobre las técnicas de entrenamien-

to generativo y discriminativo, aplicables a HMM. En la sección 3 se describe el ajuste experimental, como la parametrización de la base de datos, la metodología de validación y la arquitectura del modelo. Las dos ultimas secciones presentan los resultados y conclusiones del trabajo.

2. MATERIALES Y MÉTODOS

Sea un conjunto de R observaciones de entrenamiento $\mathbf{Y}=\left\{\boldsymbol{\varphi}_r^{n\varphi_r}: r=1,\ldots,R\right\}$, con sus correspondientes categorías o etiquetas, $\mathbf{C}=\left\{\mathbf{c}^r: r=1,\ldots,R\right\}$, donde $\mathbf{c}^r\in\{c_m: m=1,\ldots,M\}$, siendo M el número total de clases. Cada registro $\boldsymbol{\varphi}_r^{n\varphi_r}$ se representa por una secuencia de longitud $n\varphi_r$ de vectores de características $\boldsymbol{\varphi}_r^{n\varphi_r}=\left\{\varphi_{r,t}: t=1,\ldots,n\varphi_r\right\}$.

Los modelos ocultos de Markov describen procesos estocásticos doblemente anidados, compuestos de una capa oculta que controla la evolución temporal de las características espectrales de una capa observable.

El conjunto total de parámetros de los HMM se denota por Θ y se compone por M modelos, es decir, $\Theta=\{\lambda_m\}$, donde λ_m denota los parámetros del HMM que representa la categoría o clase c_m. Un Modelo Oculto de Markov para una clase en particular esta definido por el conjunto de parámetros $\lambda_m=\left\{\boldsymbol{A}^{(m)},\boldsymbol{B}^{(m)},\boldsymbol{\pi}^{(m)}\right\}$, donde $\boldsymbol{A}^{(m)}$ es la matriz de transición de estados, y esta compuesta por las probabilidades discretas $a_{ij}^{(m)}$ que representa la probabilidad de pasar del estado s_i al estado s_j, $\boldsymbol{B}^{(m)}$ corresponde a la función densidad de probabilidad de observación, que en este caso corresponde a un modelo de mezclas de Gaussianas por estado, definido como:

$$b_j^{(m)}(\varphi_{r,t})=\sum_{k=1}^{K}\boldsymbol{c}^{(m)}{}_{jk}\boldsymbol{N}\left[\varphi_{r,t},\boldsymbol{\mu}^{(m)}{}_{jk},\boldsymbol{\Sigma}^{(m)}{}_{jk}\right] \quad (1)$$

Donde $\boldsymbol{\mu}^{(m)}{}_{jk}$ es el vector de medias y $\boldsymbol{\Sigma}^{(m)}{}_{jk}$ la matriz de covarianzas de la *k*-ésima mezcla del estado s_j, que por simplicidad se asume diagonal, es decir, $\boldsymbol{\Sigma}^{(m)}{}_{jk}=[\sigma^{2(m)}{}_{jkl}]_{l=1}^{p}$, y p la dimensión del vector de observación $\varphi_{r,t}$, además, $\boldsymbol{\pi}^{(m)}$ corresponde al vector de probabilidad de estado inicial [1-5].

A continuación, se explica cada uno de los criterios de entrenamiento y las funciones que se emplean para ajustar los parámetros de los HMM, la métrica de desempeño, y el método propuesto.

2.1 Criterios de entrenamiento

Criterio MLE. Se asume que la forma funcional de $P\left(\boldsymbol{\varphi}_r^{n\varphi_r}\middle|\mathbf{c}^r\right)$ es conocida, y puede estimarse al ajustar el conjunto de parámetros del modelo para de esta forma optimizar la descripción del respectivo modelo para un conjunto dado de observaciones. La función objetivo ML se define como:

$$f_{ML}(\Theta)=\sum_{r=1}^{R}\log\left(P\left(\boldsymbol{\varphi}_r^{n\varphi_r}\middle|\mathbf{c}^r\right)\right) \quad (2)$$

Cuya optimización se alcanza ajustando los parámetros de cada modelo, por separado, con los datos de entrenamiento de cada clase, de tal forma, que el valor de (2) alcance un máximo [5].

Criterio MMI. Dada una secuencia de observación, se debe escoger la clase c_m que tenga el mínimo de incertidumbre. Ésta condición puede alcanzarse minimizando la entropía condicional, $H(\mathbf{C}|\mathbf{Y})=H(\mathbf{C})-I(\mathbf{C};\mathbf{Y})$, cuya optimización implica minimizar la entropía $H(\mathbf{C})$, o bien maximizar la información mutua $I(\mathbf{C};\mathbf{Y})$. La primera tarea corresponde a hallar el modelo con el mínimo de entropía, que analíticamente es complejo e intratable. En la segunda aproximación, se maximiza la información mutua [6]:

$$f_{MMI}(\Theta)=\frac{1}{R}\sum_{r=1}^{R}\left(\log P\left(\boldsymbol{\varphi}_r^{n\varphi_r}\middle|\mathbf{c}^r\right)-\log\sum_{i=1}^{M}P\left(\boldsymbol{\varphi}_r^{n\varphi_r}\middle|c_i\right)P(c_i)\right) \quad (3)$$

Criterio MCE. Incluye una función de pérdida, proporcional al error de clasificación, $f_{MCE}(\Theta)=l_i\left(\boldsymbol{\varphi}_r^{n\varphi_r};\Theta\right)$, y que se asocia al costo de asignar la secuencia $\boldsymbol{\varphi}_r^{n\varphi_r}$ a la clase c_i, se

Define como:

$$l_i\left(\boldsymbol{\varphi}_r^{n\varphi_r};\Theta\right)=\begin{cases}0, & \boldsymbol{\varphi}_r^{n\varphi_r}\text{ asignado correctamente a } c_i\\ 1, & \boldsymbol{\varphi}_r^{n\varphi_r}\text{ asignado incorrectamente a } c_i\end{cases}$$

Debido a que ésta no es una función derivable, se ha propuesto en cambio la siguiente función:

$$l_i\left(d_i(\boldsymbol{\varphi})\right)=\frac{1}{1+\exp\left(-\gamma\, d(\boldsymbol{\varphi})+\alpha\right)} \quad (4)$$

Donde $d_i(\boldsymbol{\varphi})$ es de la forma:

$$d_i(\boldsymbol{\varphi})=-\mathrm{g}_i(\boldsymbol{\varphi};\lambda_i)+\log\left[\frac{1}{M-1}\sum_{j,\,j\neq i}\exp\left[\mathrm{g}_j(\boldsymbol{\varphi};\lambda_j)\eta\right]\right]^{\frac{1}{\eta}} \quad (5)$$

con $\mathrm{g}_i(\boldsymbol{\varphi};\lambda_i)$ definido como la función de verosimilitud condicional para la clase c_i y η es una constante positiva [7].

2.2 Curva ROC

La toma de decisiones clínicas exige la valoración de la utilidad de cualquier prueba diagnóstica, es decir, su capacidad para clasificar correctamente a los pacientes en categorías o estados en relación con la enfermedad (típicamente dos: estar o no estar enfermo, respuesta positiva o negativa). La curva más utilizada en la literatura médica para la toma de decisiones es la ROC, que representa la tasa de falso acierto o falsa aceptación (FP) en función de la tasa de acierto o aceptación verdadera (VP), para diferentes valores del umbral de decisión. La disposición de la ROC (figura 1) depende de la forma y del solapamiento de las distribuciones subyacentes de las clases (patológica, normal – positiva, negativa) [8].

En el caso de HMM, el cálculo de la curva ROC se hace mediante los cocientes o *scores* de verosimilitud estimados de cada registro con los modelos para cada clase. Con los *scores* obtenidos se crea un histograma, que para los registros que pertenecen a la clase positiva (clase 0) debería estar situado en su mayor parte a la derecha y para los que pertenecen a la clase negativa (clase 1) en su mayor parte a la izquierda. Así, la puntuación para la secuencia $\boldsymbol{\varphi}_r^{n\varphi_r}$ está dada por:

$$s_r=\log\left(P\left(\boldsymbol{\varphi}_r^{n\varphi_r}\middle|\lambda_0\right)\right)-\log\left(P\left(\boldsymbol{\varphi}_r^{n\varphi_r}\middle|\lambda_1\right)\right) \quad (6)$$

Donde λ_i está asociado a la clase $c_i, i=0,1$. El histograma normalizado se puede interpretar como una versión discreta de las funciones densidad de probabilidad de las clases.

Una mayor precisión diagnóstica de la prueba se traduce en el desplazamiento hacia arriba y a la izquierda de la curva ROC (figura 1), lo que sugiere que el ABC se puede emplear como un índice conveniente de la exactitud global de la prueba; el mejor indicador correspondería a un valor de 1 y el mínimo a uno de 0.5 (si fuera menor de 0.5 debería invertirse el criterio de decisión de la prueba). En este sentido, se ha propuesto emplear como criterio de entrenamiento la optimización del área que encierra la ROC, teniendo como restricción la no existencia de una función analítica que represente el ABC.

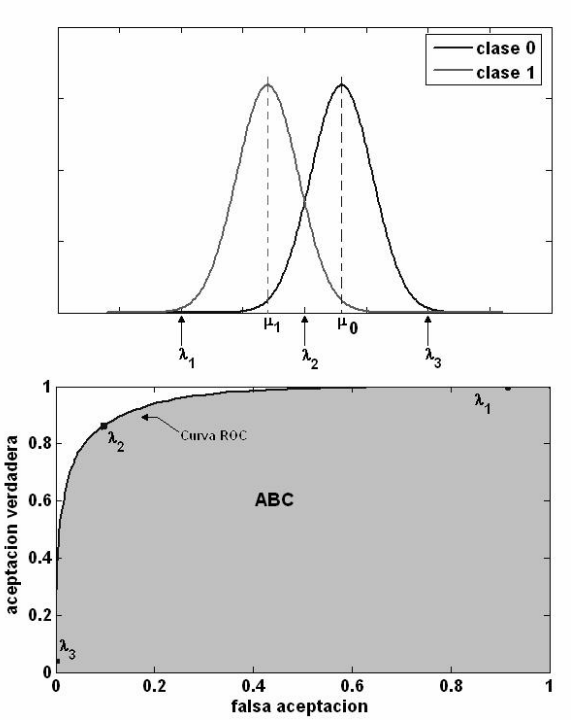

Figura. 1. Curva ROC
Figure. 1. ROC Curve

Criterio MFoM

Una primera aproximación de solución propone utilizar medidas indirectas, de tal forma que al optimizarlas sea posible maximizar el ABC de la curva ROC [10].
Debido a que el ABC de la curva ROC está directamente relacionado con el rendimiento del sistema de clasificación, es posible emplear medidas de desempeño de forma similar a como se

plantea el criterio MCE, pero con medidas más complejas y globales.

Dada una clase c_j, teniendo en cuenta que VP_j son las aceptaciones correctas, FP_j son las falsas aceptaciones y FN_j son los falsos rechazos, se pueden definir las siguientes medidas de precisión:

$$P_j = VP_j/(VP_j + FP_j) \quad (a)$$
$$R_j = VP_j/(VP_j + FN_j) \quad (b) \qquad (7)$$
$$F_j = 2VP_j/(FP_j + FN_j + 2VP_j) \quad (c)$$

Generalmente, la exactitud diagnóstica se expresa como sensibilidad y especificidad diagnósticas. Cuando se utiliza una prueba dicotómica (una cuyos resultados se puedan interpretar directamente como positivos o negativos), la **sensibilidad** (7b) es la probabilidad de clasificar correctamente a un individuo, cuyo estado real sea el definido como positivo respecto a la condición que estudia la prueba. El **valor predictivo positivo** (7a) Es la probabilidad de padecer la enfermedad si se obtiene un resultado positivo en el test. El valor predictivo positivo puede estimarse, por tanto, a partir de la proporción de pacientes con un resultado positivo en la prueba que finalmente resultaron estar enfermos.

Debido a que las medidas de (7) no son funciones derivables, es necesario aproximar la clasificación correcta o incorrecta de un registro mediante una función sigmoidal igual a la definida anteriormente para el criterio MCE:

$$TP_j \approx \sum_{\boldsymbol{\varphi}\in\mathbf{Y}} \left(1 - l_j\left(d_i(\boldsymbol{\varphi})\right)\right) 1\left(\boldsymbol{\varphi} \in c_j\right) \quad (a)$$
$$FP_j \approx \sum_{\boldsymbol{\varphi}\in\mathbf{Y}} \left(1 - l_j\left(d_i(\boldsymbol{\varphi})\right)\right) 1\left(\boldsymbol{\varphi} \notin c_j\right) \quad (b) \qquad (8)$$
$$FN_j \approx \sum_{\boldsymbol{\varphi}\in\mathbf{Y}} l_j\left(d_i(\boldsymbol{\varphi})\right) 1\left(\boldsymbol{\varphi} \in c_j\right) \quad (c)$$

Donde $\boldsymbol{\varphi}$ corresponde a una secuencia de observaciones en particular, $1(\cdot)$ es una función de indicación y es 1 si $(\cdot)$ es verdadero y 0 de otra forma. El criterio MFoM emplea la aproximación de (8) en (7c).

Por medio de estas medidas también se puede llegar a una forma del criterio MCE, tomando la expresión para los falsos negativos (falsos rechazos), y recordando la función de pérdida (4), se tiene la siguiente expresión:

$$f_{MCE}'(\Theta) = \sum_{\boldsymbol{\varphi}\in\mathbf{Y}} l_i\left(\boldsymbol{\varphi}_r^{n\varphi_r}; \Theta\right) / R \qquad (9)$$

Para este caso la optimización no se realiza en línea (*online*) sino por lotes (*batch*), y la actualización de los parámetros se realiza en dirección contraria al gradiente acumulado [11].

2.3 Método Propuesto

Cuando las distribuciones de probabilidad están separadas, tanto como es posible (figura 1) se puede asumir que el ABC alcanzara un valor máximo, por lo tanto, se propone utilizar una medida de distancia entre las distribuciones, cuya optimización indirectamente debe mejorar el ABC.

La distancia de Mahalanobis es la opción más clara, y la distancia entre distribuciones que mejor se ajusta a los requerimientos:

$$D^2 = \left(\mu_0 - \mu_1\right)^T S^{-1} \left(\mu_0 - \mu_1\right) \qquad (10)$$

Donde μ_i y S_i son las medias y varianzas respectivamente, de las distribuciones de cada clase ($i = 0,1$), S se calcula de la siguiente forma:

$$S = \left(\left(n_0 - 1\right) S_0 + \left(n_1 - 1\right) S_1\right) / N \qquad (11)$$

Donde n_0 es el número de registros de la clase 0, n_1 los registros de la clase 1, y $N = n_0 + n_1 - 2$. Analizando (10) es claro que existen al menos tres formas de hacer que la distancia D^2 sea máxima, bien maximizando la distancia entre medias (μ_i) de cada distribución, minimizando S de (11); por último, maximizando directamente la distancia D^2, tal como está definida (10).

En este trabajo se implementan dos formas: minimizar (11), FOM1 y maximizar (10), FOM2, en las cuales la media y varianza para la clase i-esima están definidas de la siguiente forma:

$$\mu_i = \frac{1}{n_i} \sum_{r=1}^{n_i} s_r 1\left(\boldsymbol{\varphi}_r^{n\varphi_r} \in c_i\right) \quad (a)$$
$$S_i = \frac{1}{n_i - 1} \sum_{r=1}^{n_i} \left(s_r - \mu_i\right)^2 \cdot 1\left(\boldsymbol{\varphi}_r^{n\varphi_r} \in c_i\right) \quad (b) \qquad (12)$$

Optimización:
Para actualizar los parámetros de cada uno de los modelos se emplea el algoritmo GPD (*Generalized Probabilistic Descend*) [12], es una técnica de optimización basada en el cálculo de gradientes, donde se definen las siguientes transformaciones sobre los parámetros a actualizar, que permiten mantener las restricciones probabilísticas de los HMM durante la adaptación:

$$\pi_j \to \hat{\pi}_j \text{ donde } \pi_j = \frac{e^{\hat{\pi}_j}}{\sum_k e^{\hat{\pi}_k}} \quad a)$$
$$a_{ij} \to \hat{a}_{ij} \text{ donde } a_{ij} = \frac{e^{\hat{a}_{ij}}}{\sum_k e^{\hat{a}_{kj}}} \quad b) \qquad (13)$$

Las transformaciones que se hacen sobre las componentes Gaussianas del modelo, se definen como:

$$c_{jk} \to \hat{c}_{jk} \text{ donde } c_{jk} = \frac{e^{\hat{c}_{ij}}}{\sum_k e^{\hat{c}_{kj}}} \quad a)$$
$$\mu_{jkl} \to \hat{\mu}_{jkl} = \frac{\mu_{jkl}}{\sigma_{jkl}} \quad b) \qquad (14)$$
$$\sigma_{jkl} \to \hat{\sigma}_{jkl} = \log \sigma_{jkl} \quad c)$$

La actualización de un parámetro θ en particular, se realiza de la siguiente forma:

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \varepsilon \frac{\partial f(\Theta)}{\partial \hat{\theta}} \qquad (15)$$

Donde ε es la tasa de aprendizaje, n indica la iteración actual y $\partial f(\Theta)/\partial \hat{\theta}$ es la derivada parcial de la función objetivo con respecto al parámetro $\hat{\theta}$. Finalmente para calcular el parámetro θ se emplean en (13) y (14).

3. MARCO EXPERIMENTAL

Los experimentos son llevados a cabo sobre la base de datos *MEEIVL*. Debido a la heterogeneidad de la base de datos (diferente frecuencia de muestreo en la adquisición de los registros), los registros utilizados fueron re-muestreados a una frecuencia de muestreo de 25 kHz y con una resolución de 16 bits. Corresponden a pronunciaciones de la vocal sostenida */ah/*. Se utilizaron 173 muestras de pacientes patológicos (con una amplia gama de patologías vocales orgánicas, neurológicas, traumáticas y psíquicas) y 53 muestras de pacientes normales, de acuerdo con los registros enumerados en [13] y como se sugiere en [14]. Los registros de pacientes patológicos tienen una duración aproximada de 1 s, mientras que en los registros de pacientes normales la duración es alrededor de 3 s.

Cada registro fue ventaneado uniformemente con una ventana Hanning de 40 ms, con un traslape del 50%. A cada ventana se le extrae un vector de *p=16* características, 12 MFCC (*Mel-Frequency Cepstrum Coefficients*), la energía de la ventana (*En*), la relación armónico ruido *(Harmonic-to-Noise Ratio - HNR)* [15], la energía de ruido normalizada (*Normalized Noise Energy* - NNE) [16] y la relación excitación glottal ruido (*Glottal to Noise Excitation Ratio* - GNE) [17].

Los MFCC son derivados del cálculo de la FFT (*Fast Fourier Transform*) [18]. Esta aproximación no paramétrica permite modelar los efectos de las patologías en la excitación (pliegues vocales) y en el sistema (tracto vocal), mientras que un enfoque paramétrico como *Linear Predictive Coefficients* (LPC) presenta problemas debido a que las patologías introducen no linealidades en el modelo [19].

Los parámetros relacionados con mediciones de ruido (HNR, NNE, GNE), están diseñados para medir la componente de ruido relativo en las señales de voz. Debido a que estas medidas dan una idea de la calidad y grado de normalidad de la voz [20].

Para determinar la capacidad de generalización de los sistemas se emplea un esquema de validación cruzada, con diferentes conjuntos de entrenamiento-validación (*k-fold*), escogidos de forma aleatoria del conjunto completo de datos. En este trabajo se emplean 9 conjuntos, utilizando para el entrenamiento el 70% de los ficheros y para la validación el 30% restante.

Figura 2. Vector de características extraídas de cada ventana
Figure 2. Feature vector extracted from each window

Para el entrenamiento de los HMM se emplea la arquitectura de un modelo ergódico (*full connected*) con 3 mezclas Gaussianas y 2 estados, puesto que fue la arquitectura que mostró los mejores resultados al entrenar el sistema mediante MLE. El algoritmo de optimización empleado es el GPD [7] para todos los criterios de entrenamiento analizados, con excepción del criterio MLE, que usa el algoritmo de Baum-Welch [5]. Adicionalmente para el criterio MCE se emplea la configuración *batch*, que representa un menor coste computacional y rendimiento similar a la configuración *online*[11].

4. RESULTADOS Y DISCUSION

Las pruebas iniciales se realizan con la técnica de entrenamiento estándar (MLE) que será la base de comparación para los demás criterios de entrenamiento. En la Tabla 1, se muestran los resultados obtenidos con el conjunto de entrenamiento. En este caso se observa que no hay una diferencia sustancial entre los diferentes métodos de entrenamiento. En la Tabla 2, se muestran los resultados obtenidos con el conjunto de validación. Donde se observa que en general todas las técnicas de entrenamiento discriminativo superan la técnica de entrenamiento generativo y adicionalmente que la técnica de entrenamiento propuesta, basada en optimizar directamente la distancia de Mahalanobis (FOM2) es superior a todas las demás técnicas, mostrando que el ABC y la precisión son los mas grandes, mostrando una clara superioridad en cuanto a capacidad de generalización.

Tabla 1. ABC y precisión (Conjunto de entrenamiento)
Table 1. AUC and accuracy (Training set)

Criterio de entrenamiento	ABC	Precisión
MLE	0.999±0.0007	99.6±0.44
MMI	0.999±0.0016	98.9±0.63
MCE	0.999±0.0003	99.5±0.69
FOM1(11)	0.998±0.0020	99.0±0.84
FOM2(10)	0.999±0.0009	99.3±0.66
MFoM	0.999±0.0002	99.6±0.64

Tabla 2. ABC y precisión (Conjunto de validación)
Table 2. AUC and accuracy (validation set)

Criterio de entrenamiento	ABC	Precisión
MLE	0.946±0.020	93.3±2.3
MMI	0.966±0.002	95.7±2.1
MCE	0.960±0.020	94.6±2.5
FOM1(11)	0.967±0.020	95.6±2.9
FOM2(10)	0.978±0.020	97.0±1.6
MFoM	0.951±0.020	94.0±3.0

Debido a que estos criterios de entrenamiento se basan en algoritmos de optimización iterativos, el coste computacional es ligeramente mayor que el del criterio MLE, a excepción del criterio basado en las métricas de desempeño (MFoM) el cual logra el desempeño mostrado en sólo una iteración, no obstante el costo computacional es justificable al obtenerse una ganancia en el desempeño del sistema de clasificación.

De igual forma en las Tablas 1 y 2 se muestran los porcentajes de acierto (precisión), que se calculan empleando la regla MAP (*Maximum A Posteriori*). Sin embargo esta medida no es suficiente para establecer claras diferencias entre los criterios de entrenamiento, y tampoco para estimar de forma adecuada el rendimiento de un sistema de clasificación, esto debido a que es posible obtener una tasa de acierto errónea al emplear un umbral de decisión mal seleccionado, por esta razón los resultados se complementan con la curva ROC (figura 3) y el ABC (tabla 2).

Otro aspecto a resaltar es que la otra estrategia de aprendizaje propuesta (FOM1), que consiste en minimizar (11), presenta buenos resultados, incluso superando la técnica de entrenamiento estándar y por un margen muy mínimo las técnicas de entrenamiento discriminativo. De esta forma se sustentan las suposiciones hechas con respecto a asociar una medida de distancia al área de la curva ROC.

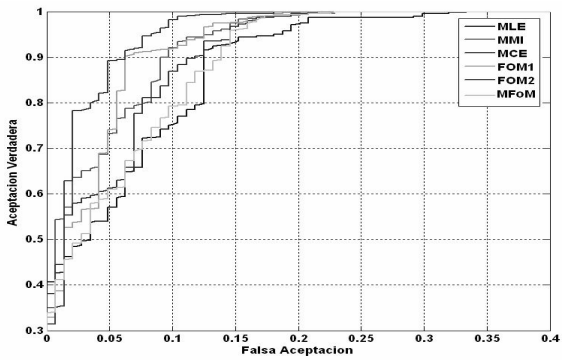

Figura 3. Curva ROC para los diferentes criterios de entrenamiento
Figure 3. ROC Curve for all different training criteria

Los resultados obtenidos son concluyentes demostrando de forma clara que las técnicas de entrenamiento discriminativo son superiores y pueden lograr una mayor capacidad de generalización en un sistema de clasificación basado en HMM, que la técnica de entrenamiento estándar, y además que el desempeño de un sistema de clasificación puede mejorarse significativamente al emplear como criterio de entrenamiento la maximización de una medida de distancia entre las distribuciones de las clases, con el fin de incrementar el área que encierra la curva ROC.

5. CONCLUSIONES

Se mejora el desempeño de clasificación del método básico de entrenamiento MLE, mediante el uso de un criterio de entrenamiento discriminativo, para el cual se sugiere el empleo de una función de costo que relaciona indirectamente el área que encierra una curva de desempeño, en particular se propone la curva ROC, con una distancia entre modelos de clases.

La función de costo empleada es la distancia de Mahalanobis, sin embargo se abordan dos aproximaciones para lograr su optimización. Mostrando en los dos casos de forma satisfactoria la estrecha relación que existe entre la medida de distancia empleada y el área de la curva ROC. Las pruebas realizadas presentan como resultado un desempeño satisfactorio empleando una arquitectura HMM relativamente simple, mejorando no solo el desempeño del método de entrenamiento estándar, sino también, los otros criterios de entrenamiento discriminativo que se tienen en cuenta. Esto demuestra que para mejorar el desempeño de un sistema de detección de patologías de voz, además de ser muy necesario contar con un buen conjunto de características, también se debe tener un criterio de entrenamiento adecuado que se enfoque en la generación de una frontera de decisión óptima, para que de esta forma no sea necesario incrementar la complejidad del modelo, y esto permita que la etapa de entrenamiento sea más eficiente

Como trabajo futuro se propone emplear una etapa de reducción de espacios de características mediante transformaciones lineales que tengan en cuenta la información cambiante en el tiempo como DPCA, para reducir el coste computacional en la etapa de entrenamiento. Además llevar esta comparación a otro tipo de señales biomédicas como PCG, EEG y ECG. Adicionalmente, se propone emplear como medida de desempeño la curva DET (*Detection Error Tradeoff*) para tener un marco de comparación más amplio.

6. AGRADECIMIENTOS

Este trabajo se enmarca dentro del proyecto 1127-40520232 “Identificacion de posturas labiales en pacientes con labio y/o paladar hendido corregido”, financiado por Colciencias y el programa Jóvenes Investigadores.

REFERENCIAS

[1] RABINER, L. A TUTORIAL ON HIDDEN MARKOV Models and selected applications in speech recognition. PROCEEDINGS OF THE IEEE, vol. 77 (2), 257-286 (1989).

[2] JIANGLIN WANG, CHEOLWOO JO. Vocal Folds Disorder Detection using Pattern Recognition Methods. EMBS’07 29th Annual International Conference of the IEEE. 3253 – 3256 (2007).

[3] P. GÓMEZ, J. I. GODINO, F. RODRÍGUEZ, F. DÍAZ, V. NIETO, A. ÁLVAREZ, V. RODELLAR. Evidence of Vocal Cord Pathology From the Mucosal Wave Cepstral Contents. Acoustics, Speech, and Signal Processing, vol 5, pp 437 – 440. (2004).

[4] GENARO DAZA-SANTACOLOMA, Julián David Arias-Londoño, Juan Ignacio Godino-Llorente, Nicolás Sáenz-Lechón, Víctor Osma-Ruíz, and César Germán Castellanos-Domínguez. Dynamic feature extraction: an application to voice pathology detection. Intelligent Automation and Soft Computing, 2009. To appear.

[5] BLIMES, J. A gentle tutorial of the EM algorithm and its applications to parameter estimation for Gaussian mixture and Hidden Markov Models. International Computer Science Institute, Bekerly CA, USA. (1998).

[6] BAHL L.R., BROWN, P.F., SOUZA, P. V. and MERCER, R.L. Maximum mutual information estimation of Hidden Markov Models parameters for speech recognition. Proceedings ICASSP, vol. 11, 49- 52 (1986).

[7] JUANG, B.H., CHOU W. and LEE, C.H. Minimum classification error rate methods for speech recognition. IEEE transaction on Speech and Audio Processing, vol. 5 (3), 257-265, (1997).

[8] HANLEY, J.A. and MCNEIL, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, vol. 143 (1), 29–36, (1982).

[9] LI, X., CHANG, E. and DAI, B.. Improving speaker verification with figure of Merit training. Acoustics, Speech, and Signal Processing, Proceedings. (ICASSP'02), vol. 1, 693- 696, (2002).

[10] GAO, S., WU., W., LEE, C. H. and CHUA, T.S. A Maximal Figure-of-Merit Learning Approach to Text Categorization. Annual ACM Conference on Research and Development in Information Retrieval. 174-181, (2003).

[11] JONATHAN L. ROUX AND ERIK MCDERMOTT, Optimization methods for discriminative training, Interspeech, septiembre 4 – 5, lisboa portugal (2005).

[12] B.-H. JUANG AND S. KATAGIRI, Discriminative learning for minimum error classification, IEEE Transactions on Signal Processing, vol. 40 (12), 3043 - 3053, (1992).

[13] V. PARSA and D.G. JAMIESON, Identification of pathological voices using glottal noise measures, Journal of Speech, Language and Hearing Research, vol 43(2), 469-485, (2000)

[14] N. SÁENZ-LECHÓN, J. I. GODINO-LLORENTE, V. OSMA-RUIZ and P. GÓMEZ-VILDA, Methodological issues in the development of automatic systems for voice pathology detection, Biomedical Signal Processing and Control, vol. 1(2), 120-128. (2006).

[15] G. DE KROM, A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals, Journal of Speech and Hearing Res.,vol 36(2), 254-266, (1993).

[16] H. KASUYA, S. OGAWA, K. MASHIMA, and S. EBIHARA, Normalized noise energy as an acoustic measure to evaluate pathologic voice, Journal of the Acoustical Society of America, vol. 80 (5), 1329–1334, (1986).

[17] D. MICHAELIS, T. GRAMMS, and H. W. Strube, Glottal-to-noise excitation ratio – a new measure for describing pathological voices, Acustica/Acta acustica, vol. 83, 700–706, (1997).

[18] L. RABINER AND B. JUANG, Fundamentals of Speech Recognition. PTR Prentice Hall, (1993).

[19] J. I. GODINO-LLORENTE, P. GÓMEZ-VILDA, N. SÁENZ-LECHÓN, M. BLANCO-VELASCO, F. CRUZ-ROLDÁN, and M. A. FERRER-BALLESTER, Discriminative methods for the detection of voice disorders. Proceedings of the 3th International Conference on Non-Linear speech processing, Barcelona, Spain, (2005).

[20] Saénz-Lechon, N., Osma-Ruiz, V., Godino-Llorente, J.I., Blanco-Velasco, M., Cruz-Roldán, F. and Arias-Londoño, J.D., Effects of Audio Compression in Automatic Detection of Voice Pathologies, IEEE Transactions on Biomedical Engineering, vol. 55 (12), 2381-2385, (2008).