

DISEÑO SIMULTÁNEO DE UNA ETAPA DE EXTRACCIÓN DE CARACTERÍSTICAS Y UN CLASIFICADOR BASADO EN HMM

Simultaneous design of a feature extractor and a HMM-based detector

RESUMEN

Se presenta una metodología de diseño simultáneo de una etapa de extracción de características y un clasificador basado en modelos ocultos de Markov (HMM), por medio del algoritmo de mínimo error de clasificación (MCE). La extracción de características depende de los estados del modelo y es optimizada utilizando el mismo criterio de ajuste de parámetros del HMM. La metodología es validada en reconocimiento de patologías de voz. Los resultados muestran que el entrenamiento por medio de MCE mejora la eficiencia en comparación con el entrenamiento clásico por máxima verosimilitud. La metodología propuesta disminuye la similitud entre modelos, mejorando el desempeño.

PALABRAS CLAVES: Extracción dinámica de características, mínimo error de clasificación, modelos ocultos de Markov, reconocimiento automático de patologías de voz.

ABSTRACT

A hidden Markov model (HMM) – based methodology for simultaneous design of extraction and classification stages is presented. Such a methodology is based on the minimum classification error (MCE) algorithm. The feature extraction is model state – dependent and is optimized using the same criterion of parameter estimation of the HMM. Validation is carried out over an automatic detection of pathological voices problem. The result shows that the MCE training improves the accuracy against the classical maximum likelihood training. The proposed methodology diminished the similarity between models of different classes and improves the performance system.

KEYWORDS: *Automatic detection of pathological voices, dynamic feature extraction, minimum classification error, hidden Markov models.*

1. INTRODUCCIÓN

El objetivo principal de los sistemas de reconocimiento de patrones es clasificar datos de entrada dentro de un número definido de clases. Convencionalmente los sistemas de reconocimiento de patrones tienen dos componentes: análisis de características y clasificación de patrones [1]. El análisis de características se alcanza en dos pasos: extracción de parámetros (EP) y extracción de características (EC). Típicamente, la etapa de EC se emplea para reducir el tamaño del espacio de representación de los datos, proporcionado en la etapa de EP, de tal manera que se utilicen únicamente las variables que mayor información aportan al proceso de clasificación y se reduzca el espacio de representación. Utilizar un espacio de representación de alta dimensión presenta principalmente dos problemas: a) Es conocido que el clasificador más simple requiere que el número de observaciones sea una función exponencial de la dimensión del espacio de características (lo que se conoce como la maldición de la dimensión) [2]. Además,

diversos problemas de reconocimiento de patrones deben ser abordados en condiciones de baja estadística, debido a deficiencias en los procesos de recolección de muestras o en el caso por ejemplo de sistemas de detección de patologías, a la dificultad en la identificación y posterior obtención de los registros de pacientes con determinada alteración, razón por la cual, el número de observaciones con que se cuenta es reducido. b) En segundo lugar, porque reduciendo la dimensionalidad del espacio de características se disminuye la complejidad del clasificador [3]. En diversos problemas de reconocimiento de patrones, se encuentra que los parámetros resultantes de la medición de variables son parámetros que presentan una dinámica temporal, y que esta dinámica en sí misma, es la que contiene la mayor parte de la información discriminante que debe ser utilizada por el sistema (éste es el caso del problema de detección automática de patologías de voz, en el cual los parámetros comúnmente utilizados, presentan un comportamiento dinámico en el tiempo y su dinámica es altamente relevante en la valoración médica de los

JULIÁN D. ARIAS-LONDOÑO

Ingeniero Electrónico. M.Sc.
G. Control y Procesamiento Digital
de Señales.
Universidad Nacional de Colombia
sede Manizales.
jdariasl@unal.edu.co

GERMÁN CASTELLANOS- DOMÍNGUEZ

Ingeniero de Radiocomunicaciones,
Ph.D. Profesor.
G. Control y Procesamiento Digital
de Señales.
Universidad Nacional de Colombia
sede Manizales.
cgcastellanosd@unal.edu.co

JUAN I. GODINO-LLORENTE

Ingeniero de Telecomunicación,
Ph.D. Profesor.
Universidad Politécnica de
Madrid, España.
igodino@ics.upm.es

pacientes [4]). Las técnicas típicamente utilizadas en la etapa de EC, están diseñadas para variables estáticas, es decir, variables que no presentan ningún tipo de dinámica (este es el caso de técnicas como PCA y LDA), lo que impide su utilización sobre características dinámicas. Simultáneamente, es conocido que las técnicas clásicamente utilizadas para realizar EC, tienen la desventaja que su criterio de optimización es diferente al criterio de diseño de la etapa de clasificación, que consiste en obtener mínimo error de reconocimiento [1]. Este hecho puede causar inconsistencia entre las etapas de EC y clasificación de un sistema de reconocimiento de patrones y consecuentemente, degradar el desempeño del sistema [5].

Por otro lado, los modelos ocultos de Markov (*Hidden Markov Models (HMM)*) son una clase de procesos estocásticos que permiten modelar series temporales [6, 7, 8]. El método convencional de entrenamiento de los HMM es el algoritmo de *Esperanza y Maximización (EM)* que es un método de estimación de parámetros que cae dentro del campo general de estimación de máxima verosimilitud (*maximum likelihood - ML*) basado en la teoría clásica de decisión Bayesiana. Sin embargo, si se evalúan las suposiciones fundamentales y limitaciones de esta aproximación, se puede encontrar que existen diferencias entre el problema de estimar una distribución óptima y el problema de diseñar un sistema de reconocimiento óptimo [9]. Considerar entonces, un sistema de reconocimiento de patrones que emplee en la etapa de extracción de características un método tradicional (con las restricciones mencionados antes) y adicionalmente utilice en la etapa de clasificación un HMM entrenado a partir del criterio ML, es un sistema que no está diseñado para obtener óptimo desempeño de reconocimiento.

Una forma directa de superar el problema de la inconsistencia entre la etapa de EC y la etapa de clasificación, es conducir las conjuntamente utilizando un criterio común de optimización [1]. Dado que en muchos problemas de reconocimiento de patrones y específicamente en el procesamiento de bioseñales, se han obtenido mejores resultados a partir de la inclusión de los HMM como método de clasificación, y teniendo en cuenta los problemas comentados antes, en este trabajo se presenta una metodología de entrenamiento que permite reducir el espacio de los datos de entrada y ajustar los parámetros de los modelos ocultos de Markov de manera conjunta, utilizando para esto, un criterio común de reducción de error.

2. MEDICIÓN / EXTRACCIÓN DE CARACTERÍSTICAS

2.1 Parametrización

En este trabajo la señal de voz es parametrizada utilizando análisis de tiempo corto, empleando ventanas

de señal 40ms de tal manera que contenga al menos dos periodos consecutivos de pitch y traslapada en un 50 %. De acuerdo con el modelo usual de la voz [10], ésta está compuesta de una secuencia de excitación en convolución con la respuesta al impulso del sistema vocal. Para modelar esta respuesta, en las tareas de procesamiento de señales de voz es común el empleo de los coeficientes derivados del análisis de predicción lineal *Linear Predictive Coefficients (LPC)* y de los *Mel-Frequency Cepstrum Coefficients (MFCC)* [10]. Los MFCCs pueden estimarse usando una aproximación paramétrica derivada de los LPC o de manera no paramétrica basados en la FFT. Sin embargo, la aproximación no paramétrica permite modelar los efectos de las patologías en la excitación (pliegues vocales) y en el sistema (tracto vocal), mientras que el enfoque paramétrico presenta problemas debido a que las patologías introducen no linealidades en el modelo [11]. Por tal motivo, en este trabajo se emplean los coeficientes MFCC derivados del cálculo de la FFT. La escala de frecuencias *mel*, en la que esta basada la representación perceptual de los MFCC, es una unidad de medida de la frecuencia percibida y no corresponde linealmente a la frecuencia física de la señal, debido a que el sistema auditivo humano aparentemente no percibe las frecuencias de manera lineal [10]. Además de los MFCC, se han considerado dentro de los vectores de características, la energía medida por ventana de la señal y parámetros relacionados con mediciones de la componente de ruido relativo en las señales de voz. En particular se utilizó la relación armónico ruido (*Harmonic-to-Noise Ratio - HNR*) [12], la energía de ruido normalizada (*Normalized Noise Energy - NNE*) [12] y la relación excitación glottal al ruido (*Glottal to Noise Excitation Ratio - GNE*) [13], debido a que estas medidas dan una idea de la calidad y grado de normalidad de la voz.

2.3. Extracción de características

Los métodos EC emplean diversos criterios para mapear los datos originales a un nuevo espacio. En algunos casos se tiene en cuenta la información de etiquetado de las clases, para encontrar una proyección que mejore la discriminación, y en otras se buscan las componentes de mayor información en todo el conjunto de datos para realizar el mapeo. Los métodos de EC más empleados en la literatura se describen a continuación.

2.3.1 Análisis de componentes principales - PCA

El análisis de componentes principales es una técnica estadística cuyo propósito es condensar la información de un gran conjunto de variables correlacionadas, en otro conjunto con menos variables (“las componentes principales”) [3]. Los parámetros del modelo pueden ser calculados directamente de la matriz de datos centralizada \mathbf{X} bien sea por descomposición en valores

singulares o la diagonalización de la matriz de covarianza (positiva semidefinida). Sea \mathbf{x}_i el i -ésimo vector de observación columna de longitud d , siendo $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$, la matriz de rotación \mathbf{U} con la que se calculan las d' componentes principales denotadas por \mathbf{z} que resumen \mathbf{x} .

$$\mathbf{z} = \mathbf{U}^T \mathbf{x} \quad (1)$$

\mathbf{U} puede ser calculada a partir de los primeros valores propios de $\mathbf{X}^T \mathbf{X}$, esto es,

$$\mathbf{X}^T \mathbf{X} \mathbf{U} = \mathbf{U} \Lambda \quad (2)$$

De esta forma los vectores originales son proyectados sobre las direcciones de mayor varianza del espacio inicial.

2.3.2 Análisis discriminante múltiple - MDA

La representación de baja dimensionalidad que entrega PCA, emplea la población de datos como un único conjunto, es decir, no tiene en cuenta la información de etiqueta en los datos [3]. El objetivo de MDA es encontrar una proyección lineal del espacio, donde la dispersión entre-clases es maximizada mientras que la dispersión intra-clases es minimizada. Una forma de encontrar tal proyección, es maximizar el radio:

$$J(\mathbf{W}) = \frac{|\mathbf{W}^T S_B \mathbf{W}|}{|\mathbf{W}^T S_W \mathbf{W}|} \quad (3)$$

Donde, \mathbf{W} es la matriz de transformación del espacio, S_B es la "matriz de dispersión entre-clases" y S_W es la "matriz de dispersión intra-clases". El problema puede ser definido como la maximización de $|\mathbf{W}^T S_B \mathbf{W}|$ que es,

$$\max_{\mathbf{W}} |\mathbf{W}^T S_B \mathbf{W}| \quad (4)$$

sujeto a,

$$|\mathbf{W}^T S_W \mathbf{W}| = 1 \quad (5)$$

Los extremos condicionales pueden ser obtenidos empleando multiplicadores de Lagrange; la solución para (4) corresponde a los $N-1$ mayores eigenvectores generalizados de S_B y S_W que equivalen a los mayores eigenvectores de $S_W^{-1} S_B$.

3. MODELOS OCULTOS DE MARKOV-HMM

Un HMM corresponde a una cadena de Markov en la cual la observación de salida es una variable aleatoria generada de acuerdo a una función de probabilidad asociada a cada estado. Un HMM está definido por el siguiente conjunto de parámetros:

$$\lambda = \{\boldsymbol{\Pi}, \mathbf{B}, \mathbf{p}_{\theta_i}\} \quad (6)$$

donde, $\boldsymbol{\Pi}$ es la matriz probabilidad de transición de estados, $\mathbf{B} = \{b_j(\cdot)\}$, $b_j(\varphi_n) = \sum_{r=1}^M c_{jr} \mathcal{N}(\varphi_n | \mu_{jr}, \Sigma_{jr})$ es la matriz de probabilidad de salida, donde M es el número

de mezclas y \mathbf{p}_{θ_i} es el vector probabilidad de estado inicial.

3.1 Entrenamiento por mínimo error de clasificación

El método de entrenamiento empleando el criterio de error de clasificación mínimo, busca minimizar la probabilidad de error a través de una representación suavizada de la función de pérdida (*loss function*), para la cual cualquier miembro de la familia sigmoidal es un obvio candidato [9]:

$$\ell(d_i(\boldsymbol{\varphi})) = \frac{1}{1 + \exp(-\gamma d_i(\boldsymbol{\varphi}) + \alpha)} \quad (7)$$

donde normalmente α es igual a cero, $\gamma \geq 1$ y,

$$d_i(\boldsymbol{\varphi}) = -g_i(\boldsymbol{\varphi}; \lambda_i) + \dots$$

$$\log \left[\frac{1}{N-1} \sum_{j \neq i} \exp[g_j(\boldsymbol{\varphi}; \lambda_j) \eta] \right]^{1/\eta} \quad (8)$$

es llamada la medida del error de clasificación, que representa simplemente una medida de la distancia entre la probabilidad de una decisión correcta y otras decisiones. N es el número de clases, η es un número positivo, $\boldsymbol{\varphi}$ es una secuencia de observaciones y $g_i(\boldsymbol{\varphi}; \lambda)$ es la función de verosimilitud condicional para la clase i , que para el caso de HMM puede ser utilizada la probabilidad conjunta estado-observación [9]. La minimización de la función de pérdida se logra mediante técnicas de gradiente descendente; en particular el algoritmo descendente probabilístico generalizado (GPD) [5]. El algoritmo GPD puede ser definido de la siguiente forma:

$$\lambda_{n+1} = \lambda_n - \varepsilon_n U_n \nabla \ell(\boldsymbol{\varphi}_n, \lambda_n) |_{\lambda=\lambda_n} \quad (9)$$

donde U_n es una matriz definida positiva y ε es la tasa de aprendizaje. En particular, el algoritmo GPD es un esquema de minimización sin restricciones que necesita modificaciones para resolver un problema de minimización con restricciones, como es el caso del entrenamiento de los HMM. Las ecuaciones de re-estimación de los parámetros de los HMM, pueden ser encontradas en [9].

4 EXTRACCIÓN DE CARACTERÍSTICAS BASADA EN HMM: Método propuesto.

De manera similar a los métodos de extracción de características expuestos en la sección 2.3, es posible calcular una matriz de transformación del espacio original, mediante la definición de la función de pérdida (7), de tal forma que se minimice la pérdida empírica [1]. Sin embargo, en el caso de secuencias de datos temporales, utilizar una única matriz de transformación, desconoce el comportamiento estocástico del proceso modelado. Una forma directa de tomar en cuenta el comportamiento dinámico del proceso para la extracción

de características, es utilizar el modelo dinámico proporcionado por la secuencia de estados del HMM. De esta manera, se puede entrenar una matriz de transformación para cada uno de los estados del modelo. Así, dada una secuencia de observaciones, y estimada una secuencia de estados más probable en el modelo $\bar{\theta}$, si la secuencia de observación en el tiempo t , se encuentra en el estado $\bar{\theta}_t$, la observación t de la secuencia será transformada utilizando la transformación asociada al estado $\bar{\theta}_t$. Formalmente, existe una transformación $\mathcal{W} = \{\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_{n_\theta}\}$, que es en forma general, el conjunto de transformaciones relacionadas con cada estado del modelo. La estimación de la transformación asociada al modelo de la clase i en el estado j , puede ser derivada de (9) como:

$$\mathbf{W}_j^{(i)}(n+1) = \mathbf{W}_j^{(i)}(n) - \varepsilon \left. \frac{\partial \ell_i(\boldsymbol{\xi}; \lambda)}{\partial \mathbf{W}_j^{(i)}} \right|_{\lambda=\lambda_n} \quad (10)$$

Donde $\boldsymbol{\xi} = \mathcal{W}^T \boldsymbol{\varphi}$ es la secuencia de observación transformada por \mathcal{W} . De la misma forma como son establecidas las restricciones para los parámetros, en el problema de estimación de los HMM [9], debe ser establecida una restricción para la transformación \mathbf{W} . Sin embargo, en este caso las restricciones no se derivan de la necesidad de satisfacer condiciones probabilísticas, sino de la necesidad de mantener acotados los parámetros transformados. Una forma de lograr este acotamiento, es limitar la norma de los vectores transformados a 1. Para el caso en el cual se tiene una secuencia de observaciones, se puede definir la siguiente función de pérdida restringida:

$$f_i(\boldsymbol{\xi}_n; \lambda) = \ell_i(\boldsymbol{\xi}_n; \lambda) - \kappa \left(\frac{1}{n_\varphi} \sum_{t=1}^{n_\varphi} \left(\left\| \left(\mathbf{W}_t^{(i)} \right)^T \boldsymbol{\varphi}_t \right\| - 1 \right) \right) \quad (11)$$

donde κ es constante. La derivada en la formula de reestimación de la transformación (Ec. (10)), se convertirá en:

$$\frac{\partial f_i(\boldsymbol{\xi}; \lambda)}{\partial \mathbf{W}_j^{(i)}} = \frac{\partial \ell_i(\boldsymbol{\xi}; \lambda)}{\partial \mathbf{W}_j^{(i)}} - \kappa \left(\frac{1}{n_\varphi} \sum_{t=1}^{n_\varphi} \delta(\bar{\theta}_t - j) \frac{\partial}{\partial \mathbf{W}_j^{(i)}} \left(\left\| \left(\mathbf{W}_t^{(i)} \right)^T \boldsymbol{\varphi}_t \right\| - 1 \right) \right) \quad (12)$$

De esta manera, cuando se calcula la secuencia de estados más probable, es posible transformar la secuencia de observaciones, asociando cada observación a un estado y transformándola.

La ventaja de esta aproximación se encuentra en que la transformación \mathcal{W} se estima a partir de la maximización de la distancia definida en (8), con lo cual el criterio de optimización es compartido entre las etapas de extracción de características y clasificación. Los parámetros del clasificador se estiman empleando el algoritmo MCE, en el mismo procedimiento iterativo en el cual se ajusta la transformación.

5. EXPERIMENTOS Y RESULTADOS

5.1 Base de datos

La base de datos utilizada es perteneciente al Grupo de Control y Procesamiento Digital de Señales de la

Universidad Nacional de Colombia sede Manizales, contiene 90 registros de la pronunciación de la vocal /a/ de forma sostenida, para diferentes individuos, de los cuales 40 corresponden a pacientes con voz normal y 50 pacientes con voz disfónica. Los registros fueron adquiridos con una frecuencia de muestreo de 22050Hz y una duración aproximada de 2.3 seg.

5.2 Toma de decisión

A partir del teorema de decisión de Bayes, es posible calcular una puntuación (o *score*) para cada una de las muestras, que permita estimar un umbral de decisión óptimo. La puntuación para el caso de HMMs, puede ser calculada como el logaritmo del cociente entre las probabilidades de generación de la muestra de ambos modelos, conocido como *likelihood ratio*. Para la evaluación de los sistemas se empleará la metodología propuesta en [14], la cual establece como primera medida el empleo de una base de datos disponible para cualquier investigador, y el empleo dentro de ésta de registros que tengan diagnóstico.

Para determinar las capacidades de generalización de los sistemas se adoptará un esquema de validación cruzada, con diferentes conjuntos de entrenamiento-validación (*k-fold*), aleatoriamente escogidos del conjunto completo de datos. En este trabajo se emplean 11- conjuntos, utilizando para el entrenamiento el 70% de los ficheros y para la validación el 30% restante. Los resultados finales serán presentados por medio de matrices de confusión (Ver Fig. 1). Para construir la matriz de confusión se deben calcular los siguientes parámetros [14]:

- *Tasa de verdaderos positivos - (tp) o sensibilidad*: es el cociente entre el número de registros patológicos correctamente clasificados y el número total de voces patológicas.
- *Tasa de falsos negativos - (fn)*: es el cociente entre el número de registros patológicos incorrectamente clasificados y el número total de voces patológicas.
- *Tasa de verdaderos negativos - (tn) o especificidad*: es el cociente entre el número de registros normales correctamente clasificados y el número total de voces normales.
- *Tasa de falsos positivos - (fp)*: es el cociente entre el número de registros normales incorrectamente clasificados y el número total de voces normales.

		Diagnóstico actual	
		Patológico	Normal
Decisión del clasificador	Patológico	<i>tp</i>	<i>fp</i>
	Normal	<i>fn</i>	<i>tn</i>

Figura 1. Aspecto general de una matriz de confusión o contingencia con dos clases.

Cabe anotar que, $tp + fn = 1$ y $tn + fp = 1$.

La tasa de acierto final del sistema se calcula como el cociente entre el número de registros correctamente clasificados y el número total de registros.

A partir de los valores de las puntuaciones entregadas por cada clasificador, podrán ser construidas las curvas de evaluación de desempeño DET y ROC. La curva ROC es una herramienta popular en tareas de decisión médicas, expresa el rendimiento en términos de la *sensibilidad* y *1-especificidad* o *fp*. La Curva DET ha sido usada ampliamente en la valoración del desempeño en sistema de identificación de hablante. La curva DET gráfica las tasas de error en ambos ejes (*fp* y *fn*), dando tratamiento uniforme a ambos tipos de error [11].

5.3 Resultados

La Tabla 1 muestra los mejores resultados obtenidos empleando diferentes criterios de entrenamiento para los HMM y los diferentes métodos de EC expuestos. Fueron realizadas varias pruebas variando el número de estados del modelo (en el intervalo [1,5]) y el número de Gaussianas (en el intervalo [2,5]). La evaluación para los métodos de EC se realizó variando el número de componentes en el espacio transformado; el número entre paréntesis indica la dimensión del espacio para el cual se logró el resultado.

Esquema	N° de Estados	N° de Gaussianas	Matriz de Confusión	
ML	3	2	82,47%	12,73%
			17,53%	87,27%
			Efi. 84,47% ± 5,4	
ML_PCA (14)	2	3	88,96%	9,09%
			11,04%	90,91%
			Efi. 86,77% ± 5,2	
ML_MDA (15)	4	3	79,87%	8,19%
			20,13%	91,81%
			Efi. 84,84% ± 8,3	
MCE	3	4	90,26%	7,27%
			9,74%	92,73%
			Efi. 91,29% ± 3,9	
MCE_PCA (14)	4	5	93,88%	14,29%
			6,12%	85,71%
			Efi. 90,48% ± 4,6	
MCE_MDA (16)	4	4	91,07%	10,00%
			8,93%	90,00%
			Efi. 90,63% ± 4,5	
MCE_ECD (13)	4	3	96,43%	7,5%
			3,57%	92,50%
			Efi. 94,79% ± 3,2	

Tabla 1. Mejores resultados obtenidos para diferentes configuraciones.

De la Tabla 1 se puede observar que el empleo del criterio de entrenamiento MCE aplicado a HMMs, reduce en un 6,82% el error de detección en comparación con el criterio convencional ML. De la misma forma, es posible observar, que el empleo de las técnicas de extracción de características PCA y MDA, no tiene una influencia positiva en el rendimiento.

Por otro lado, si se compara la reducción del error lograda por la metodología de extracción de características y entrenamiento simultáneo propuesta (MCE_ECD), se puede observar que la disminución en el error es de 10,32%. La tasa de acierto alcanzada del 94,79%, reduce en un 11,19% el error de clasificación obtenido por otros métodos sobre esta misma base de datos [15].

De igual manera, es posible observar que la desviación estándar de la eficiencia en todos los casos en los cuales se empleo el criterio MCE, para ajustar los parámetros de los modelos, es menor que la desviación para el caso en el cual se utilizó el criterio ML.

Las Figuras 2 y 3 muestran de manera comparativa las curvas ROC y DET, para el sistema empleando el criterio convencional ML, el criterio MCE y la metodología propuesta.

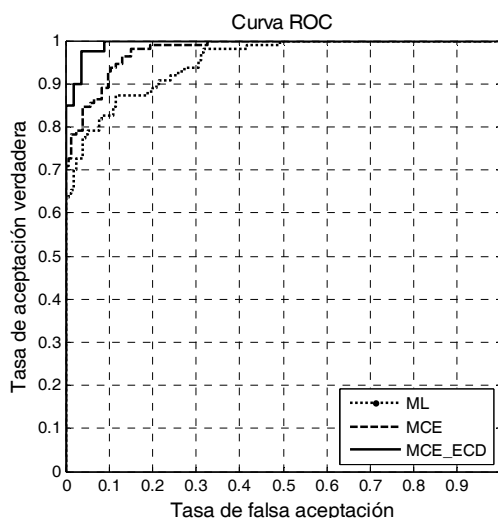


Figura 2. Curvas ROC para el sistema empleando diferentes criterios de entrenamiento de HMMs.

El área bajo la curva ROC (AUC) es una medida escalar que en algunos trabajos se ha empleado como un buen predictor de la eficiencia del sistema [16]. EL AUC para el esquema ML es 0.9507, para MCE es 0.9780 mientras que para MCE_ECD es 0.9942. Este hecho ratifica que el mejor desempeño ha sido mostrado por la metodología propuesta. Una vez más se puede observar de la Figura 3, que la curva correspondiente al esquema MCE_ECD muestra mejor rendimiento al estar más cerca del vértice inferior izquierdo en comparación con las demás curvas.

6. CONCLUSIONES

La metodología de extracción de características y entrenamiento simultáneo de HMMs mostró un mejor desempeño en comparación con el criterio convencional de entrenamiento de los HMM, sobre la base de datos utilizada. Además a diferencia de las técnicas clásicas de extracción de características, logró reducir el espacio de entrenamiento a la vez que se aumento la eficiencia del

sistema. Este hecho es debido a que las técnicas clásicas asumen las observaciones de cada registro de manera independiente, mientras que la metodología propuesta tiene en cuenta la información dinámica de las características a partir de la dependencia de las observaciones.

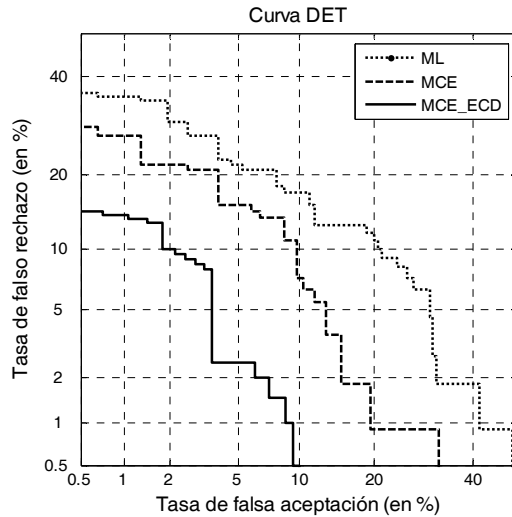


Figura 3. Curvas DET para el sistema empleando diferentes criterios de entrenamiento de HMMs

7. AGRADECIMIENTOS

Este trabajo se realiza en el marco del proyecto “Identificación automatizada de hipernasalidad en niños con LPH por medio de análisis acústico del habla”, financiado por la Dirección de Investigaciones DIMA, de la Universidad Nacional de Colombia Sede Manizales. Y cofinanciado bajo el proyecto con referencia TEC2006-12887-C02-02 (DAREVOZ) financiado por el Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica de España.

8. BIBLIOGRAFÍA

[1] X. Wang and K. K. Paliwal, “Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition”, *Pattern Recognition*, 36, pp 2429-2439. 2003

[2] A. K. Jain, P.W. Duin and J. Mao, “Statistical pattern recognition: A review”, *IEEE Transactions On Pattern Analysis and Machine Intelligence*, vol 22, no 1, pp 4-37. 2000.

[3] R. O. Duda, M. E. Hart and D. G. Stork, *Pattern Classification*, John Wiley & Sons, Second edition. 2001.

[4] M. Álvarez, R. Henao, G. Castellanos, J. Godino-Llorente, and A. Orozco, “Kernel principal component analysis through time for voice disorder classification,” in *Proc. of The 28th International*

Conf. of the IEEE Engineering in Medicine and Biology Society, New York, USA, August 2006.

[5] B.-H. Juang and S. Katagiri, “Discriminative learning for minimum error classification,” *IEEE Transactions on Signal Processing*, vol. 40, no. 12, pp. 3043–3053, 1992.

[6] M. Wester, “Automatic classification of voice quality: Comparing regression models and hidden Markov models,” *Proc of VOICEDATA98, Symposium on Databases in Voice Quality Research and Education*, 1998.

[7] J. D. Arias-Londoño, M. Álvarez, G. Castellanos-Domínguez, and J. I. Godino-Llorente, “Caracterización dinámica de señales de ECG con infarto agudo de miocardio usando HMM,” *Congreso anual de la sociedad española de ingeniería biomédica - CASEIB*, 2005.

[8] A. Dibazar and S. Narayanan, “A system for automatic detection of pathological speech,” in *Proc. of the 36th Asilomar Conf. Signals, Systems & Computers*, 2002.

[9] B.-H. Juang, W. Chou, and C.-H. Lee, “Minimum classification error rate methods for speech recognition,” *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 3, pp. 257–265, 1997.

[10] X. Huang, A. Acero and H.W. Hon, “*Spoken language processing*”, Prentice Hall, 2001.

[11] J.I. Godino-Llorente, P. Gómez-Vilda, and M. Blanco-Velasco, “Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters,” *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 10, pp. 1943–1953, 2006.

[12] D. G. Childers. “*Speech Processing and Synthesis Toolboxes*”, John Wiley & Sons, 2000.

[13] 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, pp. 700–706, 1997.

[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, no 2, pp. 120-128. 2006.

[15] G. Daza-Santacoloma, “Metodología de reducción de dimensión para sistemas de reconocimiento automático de patrones sobre bioseñales,” M. Sc. dissertation, Universidad Nacional de Colombia, 2006.

[16] A. Bradley, “The use of the area under the roc curve in the evaluation of machine learning algorithms,” *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.