# EXPERIMENTOS PRELIMINARES DE VERIFICACIÓN DE LOCUTORES CON UNA BASE DE DATOS REALISTA

José Antonio Rubio García , José Manuel Pardo Muñoz, Ricardo de Córdoba Herralde, Javier Macías Guarasa

Grupo de Tecnología del Habla, Departamento de Ingeniería Electrónica, Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, e-mail: pardo@die.upm.es, jarubio@die.upm.es ,Web: www-gth.die.upm.es.

**RESUMEN**

El presente sistema es capaz de realizar tareas de verificación de locutores de forma automática e independiente del texto. El sistema se basa en el modelado de locutores mediante mezclas gaussianas y reconocimiento de los mismos usando técnicas de cálculo de probabilidades en aritmética entera [2].

Se ha evaluado el sistema de forma completa para distintos tipos de habla, conversación, frases leídas normal y rápido, y para distinto número de gaussianas utilizadas en la creación del modelo. También se han obtenido resultados sobre el efecto que tiene entrenar los modelos de cada locutor con mayor o menor cantidad de habla.

Todos estos resultados han de ser vistos teniendo en cuenta la significación estadística de los mismos, derivada de la realización de un determinado número finito de pruebas, ya que son ciertos para un 95% de los casos y siempre considerando el resultado obtenido junto con el margen de variación del mismo.

Se han probado diversos métodos de mejora del sistema, desarrollados éstos en el dominio de distancias, con objeto de mejorar las tasas de EER entre pruebas y modelos generados con distintos tipos de habla o instantes de tiempo.

## 1. PLANTEAMIENTO Y OBJETIVOS

Los objetivos de nuestro sistema son el desarrollo de una potente herramienta de reconocimiento de locutores independiente de texto basada en modelos multigaussianos y en técnicas de mejora a posteriori (normalización de probabilidades a posteriori [3]). Las técnicas probadas sobre el sistema en el *dominio de distancias* son:

- "A Posteriori Probability Normalization" basada en Log-likelihood Ratios [3].
- "A Posteriori Probability Normalization" utilizando un modelo global de background[3].
- Medidas de parecido entre las matrices covarianza del habla de prueba y del modelo[4]. Distancia robusta.

## 2. LA BASE DE DATOS

La base de datos contiene varios tipos de habla, habla de conversación, frases leídas normalmente y frases leídas rápidamente, con 2 grabaciones en distintos instantes de tiempo separados unos 6 meses para cada uno de ellos. Todo ello con 26 locutores. Esta base de datos ha sido proporcionada por la Dirección Gral de la Guardia Civil. El tiempo medio de los fragmentos de habla con los que se ha probado, así como el tiempo medio de entrenamiento, para los casos de mayor y menor tiempo de

entrenamiento, de los modelos se muestra en las siguientes tablas:

| TIPO DE HABLA | TIEMPO MEDIO DE PRUEBA |
|---|---|
| Conversación | 31 seg. |
| Frases Normal | 7,9 seg. |
| Frases Rápido | 6,5 seg. |

**Tabla 1.-** Duración media de los fragmentos de prueba para cada tipo de habla.

| TIPO DE HABLA MODELO | Tiempo Medio Entrenamiento | Tiempo Medio Entrenamiento (mayor) |
|---|---|---|
| Conversación | 31 seg | 248 seg |
| Frases Normal | 7,9 seg | 31,6 seg |
| Frases Rápido | 6,5 seg | 26 seg |

**Tabla 2**.- Tiempo medio de entrenamiento para cada caso y tipo de habla.

## 3. TÉCNICAS DE MEJORA

A continuación se enumeran los principios teóricos de las técnicas de mejora usadas en el sistema, cuyos resultados se verán en el apartado de *Resultados y Conclusiones*.

### 3.1. Técnicas en el dominio de distancias.

Los métodos utilizados en esta línea de mejora consisten en una normalización de medidas en el dominio de distancias (o probabilístico), frente a las técnicas que suponen una normalización en el dominio de parámetros (tipo CMN, Rasta, etc...).

Se han implementado y probado tres métodos basados en la normalización de distancias, los cuales se exponen a continuación.

*3.1.1. APosteriori Probability Normalization basada en Log-Likelihood Ratios.*

Según este método, se puede calcular la probabilidad a posteriori en verificación independiente de texto como,

$$p(Sc/x) \approx \frac{p(x/Sc)}{\sum_i p(x/Si)}$$

siendo Si un locutor de la base de datos, y Sc el locutor auténtico. La probabilidad de que aparezca el locutor i, p(Si) se supone constante e igual para todos los locutores de la base de datos. El término p( x / Sc ) es la probabilidad de pertenencia del habla de prueba x al locutor auténtico Sc.

La anterior fórmula se puede escribir, considerando log-probabilidades como,

$$\log(p(Sc/x)) = \log(L(X)) = \log[p(X/S=Sc)] - \log\left[\sum_{\forall S \in Base\ Datos} p(X/S)\right]$$

Aplicando este primer método de normalización, la distancia resultante se calcula como:

$$\log(L(X)) = \log[p(X/S=Sc)] - \log\left[\sum_{\forall S \in Ref} p(X/S)\right]$$

En nuestro caso el segundo término de la ecuación, también denominado término de normalización, se calcula considerando las probabilidades de que el habla de prueba pertenezca a cada uno de los modelos de los locutores existentes en la base de datos, incluido el locutor al cual pertenece dicho habla de prueba, por tanto, *Ref* contiene a todos los locutores de la base de datos.

*3.1.2. A Posteriori Probability Normalization utilizando un Modelo de Background.*

Al igual que el método anterior, este tiene su base teórica en el método de normalización propuesto por Matsui y Furui [3], aunque la forma de calcular el término de normalización difiere del anterior considerablemente.

Matsui y Furui [3] proponen dos métodos para calcular el término de normalización de probabilidades. Ambos consisten en un modelo global que ha sido entrenado con el habla de los locutores de la base de datos.

En el primer caso, el modelo global es de 256 gaussianas y se construye a partir de habla de varios locutores que no pertenecen a la base de datos (no registrados), este modelo inicial así construido se ve modificado por el habla de entrenamiento correspondiente a los locutores registrados en la base de datos, reestimando de esta forma los factores de peso de las gaussianas. Cuando un nuevo locutor entra a formar parte de la base de datos el modelo global es actualizado (sin necesidad de volverlo a construir), volviendo a reestimar los factores de peso del modelo usando el habla de entrenamiento del nuevo locutor incorporado.

En el segundo caso (el implementado en nuestro sistema), la construcción del modelo de background se realiza con los locutores registrados, y se repite cada vez que un nuevo locutor se registra, entrenando un modelo de 64 gaussianas con todo el habla de prueba de los locutores registrados.

Así pues, una vez calculado este modelo global de background, se calcula la probabilidad de pertenencia del habla de prueba a un locutor como:

$$p(Sc/x) \approx \frac{p(x/Sc)}{\sum_i p(x/Si)}$$

Donde el sumatorio del denominador se calcula como la probabilidad de pertenencia de ese habla de prueba al modelo global. Este método se ha implementado en nuestro sistema, arrojando unos resultados mejores que los del sistema original pero aproximadamente iguales que los del método de normalización de ratios de log-likelihood (apartado anterior).

*3.1.3. Medida de parecido entre las matrices covarianza del habla de prueba y del modelo. Distancia Robusta.*

Este método de cálculo de parecido entre el habla de prueba y el modelo es el propuesto por Herbert Gish en [1]. A diferencia de los métodos anteriores (incluido el método de cálculo de distancias del sistema original) éste método no requiere modelado previo del habla de entrenamiento. Simplemente requiere el cálculo previo de ciertas características del habla de entrenamiento, supuesta una distribución gaussiana de la misma, siendo tan solo necesario el cálculo de la matriz de covarianza de dicho habla .

En la fase de prueba, el proceso el similar al del entrenamiento, y tan solo es necesario conocer la matriz de covarianza del habla de prueba (igualmente supuesta gaussiana) para realizar las medidas de parecido con los modelos.

Esto supone un enorme ahorro de tiempo en el reconocimiento, pues se evita el cálculo de los modelos a partir del habla de entrenamiento (el proceso más costoso en tiempo) y en el test de prueba el cálculo de la similitud entre ambas matrices covarianza (la del modelo y la de prueba), bajo la suposición –aceptablemente correcta- de parámetros Mel independientes entre sí, se reduce a la "comparación" de los autovalores de ambas matrices, lo cual reduce mucho el tiempo de prueba.

La forma de calcular la similitud entre el habla de prueba y de modelo viene dado por la siguiente fórmula.

$$COV(\sum_j, S) = -\frac{N-1}{2} \times \log\left|\sum_j\right| - \frac{N}{2} \times tr(\sum_j^{-1} * S)$$

Esta distancia es proporcional a la log-likelihood de que el habla de prueba, supuesta gaussiana y con una matriz covarianza S, pertenezca al modelo del locutor j representado dicho modelo con una matriz covarianza $\Sigma$. N representa el número de vectores de parámetros o tramas que componen el habla de test. $|\Sigma|$ es el determinante de la matriz covarianza del habla de modelo y tr ( $\Sigma^{-1} \bullet S$) representa la traza de la matriz producto (inversa de la covarianza del habla de modelo por la covarianza del habla de prueba).

Desarrollando la ecuación anterior se obtiene, en función de los autovalores de

$\Sigma^{-1} \bullet S$, una medida de distancia derivada de ésta, y que denominaremos COV *,

$$COV^* = \frac{N \cdot d}{2}\left(\frac{1}{d}\cdot \sum_{i=1}^{d}(\log \boldsymbol{l}_i - \boldsymbol{l}_i)\right)$$

donde d es el número de parámetros utilizados (en nuestro caso 11 parámetros MEL) y $\lambda_i$ representa el i-ésimo autovalor de $\Sigma^{-1} \bullet S$. Puesto que $\log \lambda_i \leq \lambda_i - 1$ podemos dar una cota superior a esta distancia,

$$COV^* \leq -\frac{N \cdot d}{2}$$

Derivada de ésta distancia se obtiene otra robusta, basada únicamente en los autovalores de $\Sigma^{-1} \bullet S$ . Así pues, para el caso de que las matrices $\Sigma$ y S sean iguales (las características del habla de prueba idénticas a la del habla de modelo) los autovalores de $\Sigma^{-1} \bullet S$, $\lambda_i$ son iguales a la unidad $\lambda_i = 1$ . Basándonos en esta observación podemos definir la siguiente distancia robusta,

$$dist_{ROBUSTA}(k1,k2) = \sum_{i=k1}^{k2} \left| \boldsymbol{l}_i - 1 \right|$$

en la cual k1 y k2 son dos valores comprendidos entre 1 y d (d es la dimensión de nuestros vectores de parámetros, 11 ), y los autovalores $\lambda_i$ están ordenados de menor a mayor.

Esta métrica tiene dos fuentes de robustez, la primera de ellas es que eligiendo k1>1 y/o k2<11 podemos no incluir en la métrica autovalores excesivamente pequeños o excesivamente grandes (pertenecientes a alguno de los parámetros Mel que difiere mucho entre el modelo y el habla de prueba). La segunda es la utilización del valor absoluto de la desviación de cada autovalor respecto del valor 1.

Esta técnica ha sido implementada en el sistema junto con otra modificación, consistente en sumar los valores de $\lambda_i > 1$ como $1/\lambda_i$ , limitando la contribución de autovalores muy grandes a la distancia. Los resultados conseguidos con esta técnica mejora los conseguidos con la distancia COV*.

## 4. RESULTADOS Y CONCLUSIONES

A continuación se muestra la tabla con los casos de mejora sustancial del sistema con cada uno de los métodos y variando el tiempo de entrenamiento.

| PRUEBA | MODELOS | ENTRENANDO MENOS TIEMPO | RATIOS | BACKGROUND | DIST.ROBUSTA |
|---|---|---|---|---|---|
| T1 | T1 | SI* | SI | NO | NO |
| T1 | T2 | NO | SI | SI | NO |
| T1 | FL1 | SI* | SI | NO | NO |
| T1 | FL2 | NO | SI | SI | NO |
| T1 | FR1 | SI* | SI | NO | NO |
| T1 | FR2 | NO | SI | SI | NO |
| T2 | T1 | NO | NO | SI | NO |
| T2 | T2 | SI* | NO | NO | NO |
| T2 | FL1 | NO | NO | NO | SI * |
| T2 | FL2 | SI* | SI | SI | NO |
| T2 | FR1 | NO | SI | NO | NO |
| T2 | FR2 | SI* | SI | SI | NO |
| FL1 | T1 | SI* | SI | SI | NO |
| FL1 | T2 | NO | SI | SI | NO |
| FL1 | FL1 | SI* | NO | NO | SI * |
| FL1 | FL2 | NO | NO | NO | SI * |
| FL1 | FR1 | NO | NO | NO | NO |
| FL1 | FR2 | NO | NO | NO | NO |
| FL2 | T1 | NO | SI | SI | NO |
| FL2 | T2 | SI* | SI | SI | NO |
| FL2 | FL1 | NO | NO | NO | NO |
| FL2 | FL2 | SI* | NO | NO | SI * |
| FL2 | FR1 | NO | NO | NO | NO |
| FL2 | FR2 | SI* | SI | NO | NO |
| FR1 | T1 | NO | SI | NO | NO |
| FR1 | T2 | NO | SI | SI | NO |
| FR1 | FL1 | NO | NO | NO | NO |
| FR1 | FL2 | NO | NO | NO | SI * |
| FR1 | FR1 | SI* | NO | NO | SI * |
| FR1 | FR2 | SI* | NO | NO | SI * |
| FR2 | T1 | NO | NO | NO | SI * |
| FR2 | T2 | SI* | SI | SI | SI * |
| FR2 | FL1 | NO | NO | NO | SI * |
| FR2 | FL2 | NO | NO | NO | SI * |
| FR2 | FR1 | NO | NO | NO | NO |
| FR2 | FR2 | SI* | NO | NO | NO |

* En estos casos la variación es sustancial pero PEOR que el caso Normal

A la vista de estos resultados, se puede concluir que:

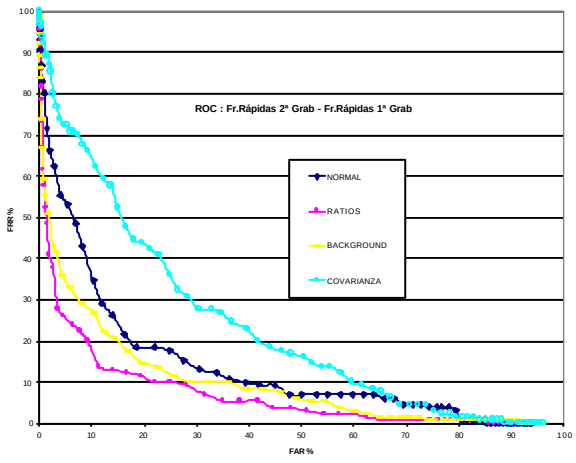
1) El efecto de entrenar con mayor cantidad de habla produce una mejora en los valores medios del intervalo de variación, para 32 de 36 casos probados y una mejora significativa (incluidos márgenes de variación) para 15 de ellos.

2) El efecto de aplicar normalización consistente en ratios de log-likelihood produce unos mejores resultados que el sistema normal, siendo estas mejoras sustanciales para 17 de los 36 casos y dándose mejoras en los valores medios del intervalo para todos los casos.

3) El efecto de aplicar normalización con modelo de background ofrece mejores resultados en los valores medios de los márgenes de variación en todos los casos (36 pruebas), siendo estas mejoras estadísticamente considerables (sin solapamiento de márgenes de variación) en 12 de esos 36 casos.

4) En cuanto a los resultados de las distancias de matrices covarianza, y en concreto la distancia robusta derivada de esta medida, se puede concluir que, a diferencia de lo que parecían apuntar las pruebas realizadas en el sistema funcionando con menor tiempo de entrenamiento, el uso de la distancia robusta no mejora el funcionamiento del sistema normal, es más, se observa un empeoramiento general para todos los valores medios, y sustancial para 11 de los 36 casos. A pesar de éstos resultados peores, este método es de interés pues el tiempo de calculo de modelos y de prueba es mínimo.

El efecto de las normalizaciones sobre el sistema se puede ver en el siguiente gráfico ROC, en el cual se aprecia el comportamiento global del sistema, dato este que no nos proporciona el valor de EER.




Las siguientes tablas de EER muestran el comportamiento del sistema funcionando con el tiempo total de entrenamiento y para los distintos tipos de habla y número de gaussianas de los modelos. La notación seguida es T1-T2 para habla de conversación en 1ª y 2ª grabación, FL1-FL2 para frases leídas y FR1-FR2 para las frases leídas rápido.

| | MODELO | T1 | | | T2 | | |
|---|---|---|---|---|---|---|---|
| PRUEBA | Gauss-> | 4 | 16 | 64 | 4 | 16 | 64 |
| T1 | NORMAL | 17,13 +/- 4,83 | 9,4 +/- 3,74 | 5,55 +/- 2,93 | 33,95 +/- 2,02 | 30,91 +/- 1,97 | 28,44 +/- 1,93 |
| | RATIOS | 4,37 +/- 2,62 | 1,282 +/- 1,44 | 0,42 +/- 0,83 | 27,06 +/- 1,9 | 17,89 +/- 1,64 | 17,39 +/- 1,62 |
| T2 | NORMAL | 29,1 +/- 1,94 | 27,01 +/- 1,9 | 25,15 +/- 1,85 | 9,82 +/- 3,81 | 5,555 +/- 2,93 | 3,39 +/- 2,32 |
| | RATIOS | 20,56 +/- 1,73 | 21,27 +/- 1,75 | 23,02 +/- 1,8 | 2,56 +/- 2,02 | 0,1538 +/- 0,5 | 3,35 +/- 2,31 |
| FL1 | NORMAL | 24,7 +/- 2,47 | 18,04 +/- 2,2 | 14,75 +/- 2,03 | 32,99 +/- 2,69 | 31,96 +/- 2,67 | 28,91 +/- 2,6 |
| | RATIOS | 15,2 +/- 2,06 | 8,29 +/- 1,58 | 5,29 +/- 1,28 | 23,24 +/- 2,42 | 20,82 +/- 2,33 | 18,03 +/- 2,2 |
| FL2 | NORMAL | 31,11 +/- 2,65 | 29,74 +/- 2,62 | 27,33 +/- 2,55 | 26,66 +/- 2,53 | 20,04 +/- 2,29 | 14,79 +/- 2,03 |
| | RATIOS | 23,75 +/- 2,44 | 22 +/- 2,37 | 21,11 +/- 2,34 | 12,58 +/- 1,9 | 3,93 +/- 1,11 | 1,605 +/- 0,72 |
| FR1 | NORMAL | 21,45 +/- 2,35 | 17,92 +/- 2,2 | 14,44 +/- 2,01 | 33,72 +/- 2,71 | 28,94 +/- 2,6 | 27 +/- 2,54 |
| | RATIOS | 15,58 +/- 2,08 | 9,67 +/- 1,69 | 7,86 +/- 1,54 | 25,29 +/- 2,49 | 21,02 +/- 2,33 | 19,44 +/- 2,27 |
| FR2 | NORMAL | 31,88 +/- 2,67 | 28,8 +/- 2,59 | 27,77 +/- 2,57 | 21,05 +/- 2,34 | 15,29 +/- 2,06 | 11,79 +/- 1,85 |
| | RATIOS | 29,04 +/- 2,6 | 24,86 +/- 2,48 | 23,43 +/- 2,43 | 12,84 +/- 1,92 | 6,75 +/- 1,44 | 3,05 +/- 0,99 |

| | MODELO | FL1 | | | FL2 | | |
|---|---|---|---|---|---|---|---|
| PRUEBA | Gauss-> | 4 | 16 | 64 | 4 | 16 | 64 |
| T1 | NORMAL | 25,03 +/- 2,48 | 19,74 +/- 2,28 | 18,03 +/- 2,2 | 34,52 +/- 2,72 | 31,86 +/- 2,67 | 29,82 +/- 2,62 |
| | RATIOS | 18,44 +/- 2,22 | 13,4 +/- 1,95 | 12,14 +/- 1,87 | 30,25 +/- 2,63 | 23,12 +/- 2,42 | 20,07 +/- 2,3 |
| T2 | NORMAL | 30,34 +/- 2,63 | 28,14 +/- 2,58 | 28,16 +/- 2,58 | 21,02 +/- 2,33 | 14,52 +/- 2,02 | 12,73 +/- 1,91 |
| | RATIOS | 25,98 +/- 2,51 | 25,73 +/- 2,5 | 24,91 +/- 2,48 | 15,12 +/- 2,05 | 5,64 +/- 1,32 | 4,12 +/- 1,14 |
| FL1 | NORMAL | 9,13 +/- 4,95 | 2,96 +/- 2,91 | 1,35 +/- 1,98 | 29,03 +/- 7,8 | 25,52 +/- 7,49 | 23,84 +/- 7,32 |
| | RATIOS | 3,2 +/- 3,03 | 0,769 +/- 1,5 | 0 +/- 0,15 | 26,3 +/- 7,57 | 17,78 +/- 6,57 | 15,38 +/- 6,2 |
| FL2 | NORMAL | 28,46 +/- 7,76 | 23,98 +/- 7,34 | 23,07 +/- 7,24 | 13,8 +/- 5,93 | 5,3 +/- 3,85 | 2,3 +/- 2,58 |
| | RATIOS | 21,53 +/- 7,07 | 18,56 +/- 6,68 | 18,46 +/- 6,67 | 5,38 +/- 3,88 | 0,769 +/- 1,5 | 0,1846 +/- 0,74 |
| FR1 | NORMAL | 12,29 +/- 5,64 | 5,27 +/- 3,84 | 4,61 +/- 3,6 | 27,03 +/- 7,63 | 24,61 +/- 7,4 | 21,53 +/- 7,07 |
| | RATIOS | 6,15 +/- 4,13 | 2,3 +/- 2,58 | 1,58 +/- 2,14 | 27,53 +/- 7,68 | 14,61 +/- 6,07 | 13 +/- 5,78 |
| FR2 | NORMAL | 24,75 +/- 7,42 | 20,92 +/- 6,99 | 23,07 +/- 7,24 | 11,53 +/- 5,49 | 5,41 +/- 3,89 | 5,38 +/- 3,88 |
| | RATIOS | 23,07 +/- 7,24 | 19,23 +/- 6,77 | 16,92 +/- 6,45 | 9,13 +/- 4,95 | 2,3 +/- 2,58 | 2,3 +/- 2,58 |

| | MODELO | FR1 | | | FR2 | | |
|---|---|---|---|---|---|---|---|
| PRUEBA | Gauss-> | 4 | 16 | 64 | 4 | 16 | 64 |
| T1 | NORMAL | 22,49 +/- 2,39 | 20,89 +/- 2,33 | 20,08 +/- 2,3 | 33,76 +/- 2,71 | 30,94 +/- 2,65 | 29,74 +/- 2,62 |
| | RATIOS | 16,43 +/- 2,12 | 13,33 +/- 1,95 | 12,73 +/- 1,91 | 30 +/- 2,63 | 23,2 +/- 2,42 | 21,7 +/- 2,36 |
| T2 | NORMAL | 32,77 +/- 2,69 | 30,07 +/- 2,63 | 30,41 +/- 2,64 | 18,16 +/- 2,21 | 14,7 +/- 2,03 | 14,017 +/- 1,99 |
| | RATIOS | 28,47 +/- 2,59 | 23,5 +/- 2,43 | 24,44 +/- 2,46 | 15,98 +/- 2,1 | 8,9 +/- 1,63 | 5,43 +/- 1,3 |
| FL1 | NORMAL | 13,7 +/- 5,91 | 10,76 +/- 5,33 | 10 +/- 5,16 | 27,69 +/- 7,69 | 27,69 +/- 7,69 | 23,24 +/- 7,26 |
| | RATIOS | 6,92 +/- 4,36 | 4,64 +/- 3,62 | 3,84 +/- 3,3 | 23,49 +/- 7,29 | 14,61 +/- 6,07 | 11,53 +/- 5,49 |
| FL2 | NORMAL | 30,69 +/- 7,93 | 26,92 +/- 7,62 | 24,61 +/- 7,4 | 15,38 +/- 6,2 | 9,2 +/- 4,97 | 6,89 +/- 4,35 |
| | RATIOS | 23,12 +/- 7,25 | 17,69 +/- 6,56 | 15,3 +/- 6,19 | 8,3 +/- 4,74 | 2,3 +/- 2,58 | 0,8 +/- 1,53 |
| FR1 | NORMAL | 4,67 +/- 3,63 | 1,53 +/- 2,11 | 0,7692 +/- 1,5 | 24,7 +/- 7,41 | 22,12 +/- 7,13 | 19,81 +/- 6,85 |
| | RATIOS | 3,84 +/- 3,3 | 0,769 +/- 1,5 | 0,0923 +/- 0,52 | 23,84 +/- 7,32 | 13,84 +/- 5,94 | 10,87 +/- 5,35 |
| FR2 | NORMAL | 24,61 +/- 7,4 | 21,43 +/- 7,05 | 18,46 +/- 6,67 | 8,29 +/- 4,74 | 3,84 +/- 3,3 | 1,53 +/- 2,11 |
| | RATIOS | 21,36 +/- 7,05 | 15,38 +/- 6,2 | 13,07 +/- 5,79 | 4,6 +/- 3,6 | 0,6769 +/- 1,41 | 0,184 +/- 0,74 |

## REFERENCIAS


[1]- H.Gish, "*Robust Discrimination in Automatic Speaker Identification*", Proc. ICASSP, pp. 289-292, Abril 1990.

[2]- J.Ferreiros López, "*Aportación a los métodos de entrenamiento de los modelos de Markov para reconocimiento de habla contínua*", Tesis Doctoral ETSIT 1996.

[3]- Tomoko Matsui and Sadaoki Furui. "*Similarity Normalization Method for Speaker Verification Based on A Posteriori Probability*". Esca Workshop on Automatic Speaker Recognition, Identification and Verification 1994.

[4]- Herbert Gish. *"Robust Discrimination in Automatic Speaker Identification*". Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing, pp 289-292.

[5]- José Antonio Rubio García. "*Evaluación y optimización del sistema de verificación automática de locutores independiente del texto sobre modelos multigaussianos*". Proyecto Fin de Carrera ETSIT 2000.