



# APLICACION DE LA CUANTIZACION VECTORIAL EN LA VERIFICACION AUTOMATICA DEL HABLANTE

---

*CARLOS ALBERTO RUEDA V.*

Grupo de Investigación en Imagen y Sonido

Facultad de Ingeniería de Sistemas

Universidad Autónoma de Manizales

E

ste trabajo aplica la técnica de la cuantización de vectores para fines de la autenticación del hablante en la modalidad dependiente del texto. Se discute principalmente el problema de la determinación de umbrales de decisión en la tarea de clasificación. Utilizando una base de datos de cuatro hablantes-cuatro palabras, se examina el planteamiento básico de verificación y la forma cómo pueden mejorarse con miras a una experimentación de mayor escala.

A N F O R A

## L. INTRODUCCION

---

La voz es el medio natural por excelencia para la comunicacion humana, cuyo contenido no solo conlleva el mensaje directamente expresado, sino también otros tipos de informacion como son el acento, el estado de animo y, en particular, la misma identidad de la persona. En un marco de comunicacion exclusivamente oral -una llamada telefónica, por ejemplo- pueden distinguirse dos modalidades de reconocimiento a los que se enfrenta un oyente con respecto a su interlocutor: *identificar* quien le contesta cuando inicia la llamada; o bien, *verificar* si quien contesta es la persona que él espera. Este trabajo se concentra en el problema de la verificación mediante el computador. En uso, el sistema recibe como entrada la identidad de la persona y la pronunciaci3n de una palabra predeterminada. Confrontando esta informacion con respecto a una cierta caracterizaci3n previamente obtenida para la persona, el sistema responde si el hablante es genuino o es impostor.

Entre los principales campos de aplicaci3n de los sistemas automáticos de verificación del hablante encontramos el control de acceso a recursos restringidos, por ejemplo, bases de datos confidenciales, así como también la supervisi3n de transacciones remotas como transferencia electrónica de di-

nero, telemarketing, etc. Los usos más especializados incluyen sistemas de verificación en el rastreo de llamadas.

Las primeras características utilizadas para la verificación del hablante incluyeron el tono, la intensidad, los formantes, la coarticulaci3n en sonidos nasales y algunos coeficientes asociados con el análisis de predicci3n lineal.

<sup>11</sup> Los estudios posteriores se han basado en diversas variantes o en nuevas características así como en diferentes métodos para la medici3n de distorsiones. Un elemento importante fue la inclusi3n de características transicionales -es decir, en relaci3n con las variaciones temporales de la seña de la voz- toda vez que dicha informaci3n debe tener gran dependencia del hablante. Furui <sup>12</sup> obtiene un alto grado de desempeño con sus sistemas de verificación utilizando informaci3n espectral tanto instantánea como transicional. El sistema estudiado por Burton <sup>13</sup> se basa en una técnica de cuantizaci3n vectorial por multisecciones. Su mejor variante logra un 0.7% de aceptaci3n equivocada y un 0.6% de rechazo equivocado sobre una poblaci3n de 16 hablantes genuinos y 111 impostores casuales. El trabajo de Soong y Rosenberg <sup>14</sup> contempla también la informaci3n espectral instantánea y transicional pero orientado a la identificaci3n del hablante. Sus resultados muestran que la representaci3n instantánea tiene más po-

der discriminador que la representación transicional, y que ambas informaciones, por mostrar relativamente poca correlación, mejoran en conjunto el desempeño. En Chen *et al.*<sup>17</sup> se utiliza la técnica de la cuantización matricial en un sistema de identificación del hablante. En esta técnica se cuantizan grupos de vectores en vez de vectores individuales, lográndose altos desempeños al no solo capturar las características espectrales de corto tiempo, sino también la información temporal global asociada a cada hablante. El trabajo de Tishby<sup>18</sup> utiliza los modelos escondidos de Markov autorregresivos como medio de caracterización de hablantes en un sistema de verificación. Su desempeño arroja un error global menor del 3%. Entre sus conclusiones, Tishby destaca que la mejora lograda por este modelamiento con respecto al enfoque más simple de la cuantización vectorial no compensa el esfuerzo adicional requerido solo para fines de reconocimiento del hablante. El mismo esquema de caracterización es utilizado por Jimenez<sup>19</sup> orientado hacia la identificación de hablantes. El sistema de Jimenez obtiene un grado de acierto por encima del 90% aplicado sobre una población de cuatro hablantes.

Este estudio se basa concretamente en la cuantización de vectores como forma básica de caracterización de los hablantes. Se desarrolla de la si-

guiente manera. En la sección 2 se formaliza el problema de la verificación y se enuncia la estrategia utilizada para solucionarlo. Después de revisar brevemente la operación de cuantización, el documento discute más detenidamente el problema de la determinación de umbrales de decisión en la tarea de clasificación. La sección 2 describe la experimentación llevada a cabo. Con el ánimo de ilustrar el progreso del trabajo, esta sección muestra los resultados obtenidos desde el planteamiento preliminar hasta mostrar los ajustes involucrados para mejorar el desempeño del sistema.

## 2. FORMULACIÓN DEL PROBLEMA

---

El problema de la verificación del hablante dependiente del texto se puede plantear de la siguiente manera. En la fase de operación como tal, la persona introduce en el sistema dos elementos de información: un código de su identidad personal (mediante el teclado, por ejemplo) y la pronunciación de una palabra perteneciente a un cierto vocabulario predeterminado (a través de algún dispositivo de adquisición de datos). A continuación, el sistema procede a confrontar estas dos informaciones para determinar si hay suficiente grado de consistencia entre ambas, lo que equivale a decidir si la persona

es quien dice ser. Esto requiere una fase previa de entrenamiento durante la cual el sistema almacena la información necesaria para caracterizar adecuada-

mente a cada hablante a verificar. La Figura 1 muestra el diagrama correspondiente a la fase de verificación.

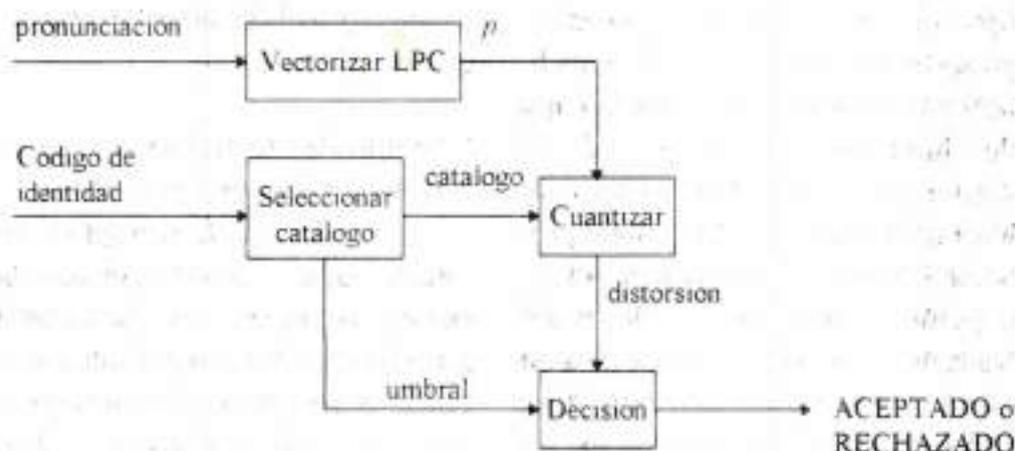


Figura 1 Esquema de la fase de verificación del sistema

Para cada posible hablante  $i$ , se determina una caracterización  $C_i$  adecuada, se define una medición  $d(i, p)$  del grado de distorsión o incompatibilidad de la pronunciación  $p$  con respecto al hablante  $i$ , y se determina un umbral de la decisión  $\theta_i$  para cada  $C_i$ . El hablante será aceptado sólo si  $d(i, p) \leq \theta_i$ .

## 2.1 Caracterización

Se ha establecido que la voz humana se mantiene aproximadamente estacionaria, es decir, sin cambiar los rasgos estadísticos, en periodos de 25-

75 ms. Esta propiedad permite aplicar sobre cada tramo un análisis de predicción lineal (8). Dicho análisis produce una secuencia temporal de vectores de coeficientes predictores. Si denotamos con  $\mathbf{a}_t$  el vector de predicción obtenido sobre el  $t$ -ésimo tramo, y  $T$  es el número de tramos extraídos y analizado de la pronunciación, podemos representar la pronunciación como

$$p = \mathbf{a}_1 \mathbf{a}_2 \dots \mathbf{a}_T$$

La caracterización de un hablante se lleva a cabo mediante una tec-

nica de codificación de fuentes de información llamada *cuantización vectorial* (CV) (9). En la cuantización vectorial, cada hablante es caracterizado con una colección de vectores representativos. Esta colección se constituye en el *catálogo* de vectores prototipo con los que mejor se pueden representar los vectores de predicción lineal obtenidos de las pronunciaciones del hablante. El objetivo de la cuantización vectorial es entonces obtener una colección de  $M$  vectores prototipo  $C_i = [c_1, c_2, \dots, c_M]$  que mejor representen a todos los posibles  $a$  del hablante  $i$  al menos con respecto a un cierto vocabulario de referencia. Esta mejor representación significa que los  $c_m$  deben minimizar la distorsión promedio esperada cuando cada  $a$  es cuantizado como  $c_m$ . La cuantización para  $a$  dará como resultado el vector  $c_m$  si:

$$d(a, c_m) \leq d(a, c_l) \quad \forall l$$

## 2.2 Medida de Distorsión

La medición de distorsión de una pronunciación  $p$  con respecto al hablante  $i$ ,  $d(i, p)$ , puede especificarse como la distorsión promedio de cuantización de la pronunciación con respecto al catálogo  $C_i$ , expresado como:

$$d(i, p) = \frac{1}{T} \sum_{t=1}^T \min_m d(a_t, c_m)$$

Por lo tanto, a menor distorsión promedio, mayor será el grado de consistencia entre la identidad y la pronunciación.

## 2.3 Umbrales de decisión

En tanto que disponemos de una caracterización explícita por cada hablante a verificar, no se cuenta por lo general con información precisa sobre hablantes impostores. Por lo tanto, los umbrales de decisión deben ser determinados empíricamente con base en una muestra disponible de pronunciaciones de la población de impostores. Esto se hace teniendo en cuenta los dos posibles errores en la fase de verificación: aceptación equivocada (AE) y rechazo equivocado (RE).

Sea  $D_i$  el conjunto de las distorsiones obtenidas al codificar las pronunciaciones del hablante  $i$  con el catálogo  $C_i$ , y sea  $D_j$  el conjunto de las distorsiones obtenidas al codificar las pronunciaciones de los hablantes disponibles diferentes de  $i$  también con respecto al catálogo  $C_i$ . Tómese la media y la desviación estándar de estos conjuntos de datos:

Un primer criterio consiste en aproximar las densidades de probabilidad para pronunciaciones impostoras y correctas relativas al hablante y calcular el umbral tal que se igualen las probabilidades para las dos equivocaciones. Entonces, bajo el supuesto de distribuciones gaussianas para estos datos, el umbral de decision sera

$$\theta_j = \frac{\mu_i \bar{\sigma}_j + \bar{\mu}_i \sigma_j}{\sigma_j + \bar{\sigma}_j}$$

Como segunda alternativa tendremos en cuenta tambien el criterio de Furui<sup>11</sup>, cuya formula considera directamente sólo la distribucion para los hablantes impostores disponibles.

$$\theta_j = a(\bar{\mu}_i - \bar{\sigma}_j) + b$$

donde  $a$  y  $b$  son dos parametros constantes determinados experimentalmente que se aplican para todos los hablantes. La fórmula muestra que a mayor media y menor desviacion standard, mayor es el umbral resultante.

### 3. EXPERIMENTACIÓN

---

#### 3.1 Condiciones de trabajo

---

Para este trabajo se utilizo el banco de datos construido en el proyecto *Mira quién habla: Sistema de reco-*

*nocimiento del hablante*<sup>12</sup>. Esta base de datos esta conformada con las voces de cuatro personas (3 hombres, 1 mujer) pronunciando cuatro palabras: *acuna, biologia, tricasó e identifiqume*. De cada hablante se tienen 44 ocurrencias por cada una de las cuatro palabras, lo que da un total de 176 pronunciaciones por persona. Para la obtencion de esta base de datos se llevaron a cabo cinco sesiones de captura separadas entre si por intervalos entre 2 y 4 semanas. En las primeras 4 sesiones se hicieron 10 capturas por palabra, y en la ultima 4 capturas.

Las pronunciaciones se digitalizaron a una frecuencia de 6.670 muestras por segundo, por el método de aproximacion sucesiva con una resolucion de 12 bits. Se utilizo un factor de preenfasis de 0.95 y un análisis de prediccion lineal de 8 polos por el método de autocorrelacion cada 15 ms sobre secciones de 45 ms con ponderacion de Hamming. Tanto este procesamiento, como la generacion de catalogos de cuantizacion, se hace utilizando el software desarrollado por este autor en un proyecto sobre reconocimiento de palabras pronunciadas de manera aislada<sup>13</sup>.

#### 3.2 Entrenamiento y pruebas

---

Se efectuaron varios tipos de pruebas con el fin de obtener una bue-

na configuración para el sistema de verificación. La primera considera todos los datos disponibles y tiene como objetivo básico examinar la operación para determinar los umbrales de decisión, con el fin de despejar el camino hacia las siguientes pruebas. La siguiente fase consistió en tomar una mitad de la base de datos para hacer la caracterización y el cálculo de umbrales, y dejar la otra mitad exclusivamente para ensayos abiertos. En todos los casos que se describen en este trabajo, se utilizaron catálogos de cuantización de tamaño 64.

### 3.2.1 Pruebas preliminares

El primer ensayo consistió en tomar todas las 176 pronunciaciones disponibles para uno de los hablantes tanto para la generación cuantizadores, como para la obtención de las propias medidas estadísticas. En el mismo sentido, se tomaron todas las 528 pronunciaciones disponibles de los otros tres hablantes en el cálculo de los estadísticos para los impostores. La Tabla 1 muestra los resultados obtenidos.

<b>Tabla 1</b> Prueba cerrada total para un hablante-todas las palabras <i>umbral = 142</i>						
Intentos genuinos	Rechazos equivocados		Intentos impostores	Aceptaciones equivocadas		Error global
176	12	6.8%	528	9	1.7%	4.25%

La última columna es el promedio de los dos tipos de error. Si bien la fórmula para el umbral busca igualar las probabilidades para los dos errores posibles, la diferencia resulta significativa.

A continuación, siguiendo con

la modalidad de pruebas cerradas, o sea, utilizando todas las pronunciaciones disponibles en cada caso, se hace entrenamiento discriminado por palabra. La Tabla 2 ilustra los resultados obtenidos. La última fila muestra los totales.

**Tabla 2**

Prueba cerrada total para un hablante—por palabra

Palabra	Umbral	Intentos genuinos	Rechazos equivocados		Intentos impostores	Aceptaciones equivocadas		Error global
<i>identifiqueme</i>	120	44	6	13.6%	132	0	0.0%	6.80%
<i>biología</i>	171	44	4	9%	132	2	1.5%	5.25%
<i>fricase</i>	131	44	6	13.6%	132	2	1.5%	7.55%
<i>aeton</i>	130	44	0	0%	132	0	0.0%	0.00%
Totales		176	14	7.9%	528	4	0.7%	4.30%

Con esta variante se obtiene un incremento de rechazos equivocados y una disminución de las aceptaciones equivocadas, y el error global permanece prácticamente igual.

El siguiente paso fue particionar el repertorio de datos en dos conjuntos aproximadamente iguales: uno (el más pequeño) utilizado para entrenamientos y el otro destinado exclusivamente para pruebas abiertas.

La Tabla 3 muestra los resultados obtenidos para uno de los cuatro hablantes con pruebas abiertas. Se utilizaron 80 pronunciaciones tanto para la generación del cuantizador, como para la obtención de las medidas estadísticas del propio hablante. Se tomaron 240 pronunciaciones de los otros tres hablantes en el cálculo de los estadísticos para los impostores.

**Tabla 3**

Pruebas para un hablante—todas las palabras  
*umbral = .122*

Tipo de prueba	Intentos genuinos	Rechazos equivocados		Intentos impostores	Aceptaciones equivocadas		Error global
Cerrada	80	3	3.7%	240	0	0.0%	1.85%
Abierta	96	44	47.9%	288	0	0.0%	23.96%

Los resultados en prueba cerrada mejoran aunque no de una manera significativa con respecto a las pruebas sobre el conjunto total. Sin embargo, el desempeño ante pronunciaciones nuevas sufre una degradación considerable. Nótese que el error recae completamente sobre el rechazo equivocado. Este comportamiento es consecuencia de un umbral de decisión demasiado bajo, lo que hace tender a un rechazo generalizado.

La Tabla 4 muestra los resultados obtenidos en prueba abierta cuando se hace entrenamiento discriminado por palabra pero utilizando el mismo cuantizador del caso anterior. Se utilizan 20 pronunciaciones para el cálculo de estadísticos en el caso del hablante genuino, y 60 pronunciaciones de los otros tres hablantes en el cálculo de los estadísticos para los impostores. La última fila muestra los totales.

**Tabla 4**

Prueba abierta para un hablante—por palabra

Palabra	Umbral	Intentos genuinos	Rechazos equivocados		Intentos impostores	Aceptaciones equivocadas		Error global
<i>identifiqueme</i>	114	24	22	91.6%	72	0	0.0%	45.80
<i>biologia</i>	152	24	2	8.0%	72	1	1.4%	4.70
<i>fricase</i>	117	24	5	20.8%	72	0	0.0%	10.40
<i>aetou</i>	112	24	20	83.3%	72	0	0.0%	41.65
<b>Totales</b>		96	49	51.0%	288	1	0.3%	25.65

En esta tabla se manifiesta igualmente el fenómeno de los umbrales demasiado bajos. No obstante, el comportamiento más bien errático del rechazo equivocado en función de la palabra, obliga a pensar en otras causas que posiblemente tienen que ver con la insuficiencia de los datos de entrenamiento. Es evidente que para mejo-

rar el desempeño del sistema se hace necesario refinar el criterio para la determinación de los umbrales de decisión. En la siguiente sección se examinan algunas posibilidades.

### 3.2.2 Mejoramiento del sistema

En estadística, es bien sabido

que mientras la media muestral puede constituirse en una buena estimación de la media de la población correspondiente, no sucede lo mismo con respecto a la variabilidad. En efecto, la variabilidad en una muestra tiende a ser menor que la de la población de la cual proviene ya que es poco probable que una muestra relativamente pequeña contenga los extremos de la población, y, por tanto, presentara menos variabilidad<sup>14</sup>. Por lo tanto, alterar de alguna manera la desviación estandar para el hablante de modo que se haga mayor, parece ser una opción a considerar. Por otra parte, también puede arguirse una justificación razonable para restimar la media:

recuerdese que el catalogo con el cual se hace la operación de cuantización se construye buscando *minimizar* la distorsión promedio. De hecho, la promedio aumentara si se toman repertorios de entrenamiento cada vez mas grandes. El objetivo es entonces encontrar formulas validas para hacer que los umbrales de decision sean mayores.

En primera instancia, pueden examinarse los resultados fijando los umbrales de forma manual. La Tabla 5 ilustra el desempeño como función del umbral. La primera fila corresponde al umbral calculado mediante la fórmula inicial.

Este comportamiento del error

**Tabla 5**

Desempeño en función del umbral  
Pruebas abiertas para un hablante—todas las palabras  
Intentos genuinos 96 — Intentos impostores 288

Umbral	Rechazos equivocados		Aceptaciones equivocadas		Error global
.122	44	47.92%	0	0.0%	23.96%
.140	26	27.08%	0	0.0%	13.54%
.160	10	10.42%	5	1.73%	6.07%
.180	3	3.12%	11	3.81%	3.47%
.200	1	1.04%	30	10.38%	5.71%
.220	0	0.0%	58	20.07%	10.03%
.240	0	0.0%	83	28.72%	14.36%
.300	0	0.0%	161	55.71%	27.85%

Este comportamiento del error global en función del umbral es típico.

La Figura 2 contiene la grafica sobre un rango más amplio.

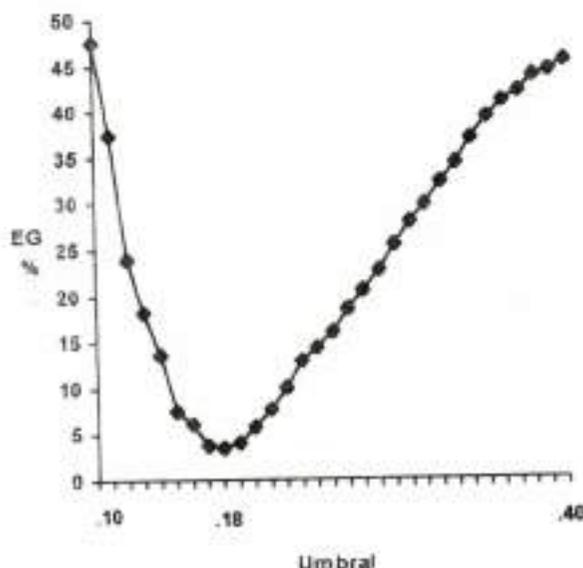


Figura 2 Error global como función del umbral

Estos resultados hacen evidente que las dos clases están bien diferenciadas, es decir, el cuantizador logra caracterizar muy aceptablemente al hablante, y que el problema radica en la escogencia del umbral de decisión (observe el notable desempeño para  $\theta=0.18$ ).

Consideremos ahora la fórmula  $\theta_i = a(\bar{\mu}_i - \bar{\sigma}_i) + b$  la cual depende solo de los estadísticos para los impostores. La fórmula es intuitivamente razonable ya que las distorsiones promedio para los impostores deben presentar gran variabilidad y, por lo tanto  $\bar{\mu}_i - \bar{\sigma}_i$  debe ser una buena frontera inferior para el grueso de estas distorsiones. Las constantes, al

ser determinadas empíricamente, ayudarían a ajustar esta primera aproximación. La Tabla 6 muestra los resultados obtenidos utilizando la fórmula directa  $\theta_i = \bar{\mu}_i - \bar{\sigma}_i$ . Esto ha logrado una mejoría radical. Además, notese el comportamiento consistente tanto entre palabras independientemente examinadas, como con respecto a las cuatro palabras consideradas sin discriminación. El mínimo error global obtenido en este último caso reitera la importancia de contar con buena cantidad de datos de entrenamiento.

Con la claridad lograda hasta este punto, se procedió a la aplicación de las diferentes alternativas examina-

cuatro hablantes (sin diferenciar entre palabras). Las pruebas mostraron que el uso del último criterio para el cálculo del umbral,  $\theta_j = \mu_j - \sigma_j$ ,

comporta de manera consistentemente superior. La Tabla 7 muestra los resultados obtenidos en este último caso.

Observese que solo un hablan-

**Tabla 6**

Prueba abierta para un hablante

Palabra	Umbral	Intentos genuinos	Rechazos equivocados		Intentos impostores	Aceptaciones equivocadas		Error global
(todas)	180	96	3	3.12%	288	11	3.81%	3.47%
<i>identifiqueme</i>	151	24	2	8.33%	72	0	0.0%	4.17%
<i>biologia</i>	226	24	0	0.0%	72	7	9.72%	4.86%
<i>fricase</i>	161	24	1	4.17%	72	7	9.72%	6.95%
<i>aeton</i>	205	24	1	4.17%	72	3	4.17%	4.17%
Totales palabras		96	4	4.17%	288	17	5.9%	5.03%

**Tabla 7**

Pruebas abiertas generales por hablante - umbral  $\theta_j = \tilde{\mu}_j - \tilde{\sigma}_j$

Código hablante	Umbral	Intentos genuinos	Rechazos equivocados		Intentos impostores	Aceptaciones equivocadas		Error global
CAR	180	96	3	3.12%	288	11	3.81%	3.47%
HFM	177	96	49	51.04%	288	73	25.26%	38.15%
JJJ	205	96	3	3.12%	288	31	10.76%	6.93%
LAR	214	96	0	0.0%	288	51	17.65%	8.82%
Totales		384	55	14.32%	1152	166	14.40%	14.36%

te tuvo un desempeño comparativamente pobre, provocando una sensible degradación en el desempeño total. Esto no es sorpresa, pues se sabe que por lo general aparece una pequeña fracción de usuarios «difíciles», sobre los cuales se deberían intensificar los entrenamientos. Con respecto a los demás hablantes, la situación es más o menos uniforme. Cabe anotar, además, que los resultados de la última columna son consistentes si se tiene en cuenta el nivel previo de experiencia de los hablantes participantes ante sistemas de reconocimiento similares.

#### 4. CONCLUSIONES

Los resultados de este trabajo han validado la aplicabilidad de la técnica de la cuantización vectorial como método de caracterización de la voz para fines de la autenticación de hablantes. Con respecto a un solo hablante, los resultados obtenidos alcanzaron hasta un 3.5% de error global ante 96 intentos genuinos y 288 intentos de tres impostores en relación a un vocabulario predeterminado de cuatro palabras. Los resultados por palabra fueron levemente menos efectivos. Con respecto al total de los cuatro hablantes, se obtuvo un error global del 14.4% ante 384 intentos genuinos y 1.152 intentos impostores. Este resultado es notable teniendo en cuenta que uno de

los cuatro participantes se comportó como hablante difícil. Los otros tres hablantes estuvieron por debajo del 9% de error global.

En este estudio se evidenció la relativa poca dimensión de la base de datos sobre la que se hicieron las pruebas, toda vez que debió recurrirse a la revisión y ajuste de las fórmulas para los umbrales de decisión que mejorarán el desempeño del sistema. Es de suponer que en pruebas a mayor escala se haga más aplicable la fórmula original que asume distribuciones gaussianas. En este sentido hay varias observaciones que hacer. Primero, según Burton, el logaritmo de la distorsión promedio presenta un comportamiento más gaussiano que la misma distorsión promedio como tal. Además otra alternativa válida consiste en confrontar los propios histogramas de las distorsiones del hablante genuino y de los impostores para determinar los umbrales de manera directa. Estas consideraciones pueden tenerse en cuenta para experimentación futura. Otra variante que merece examinarse, también utilizada exitosamente por Burton, consiste en tomar, digamos,  $N$  pronunciaciones del hablante para obtener el catálogo de cuantización, y calcular los estadísticos tomando estas mismas pronunciaciones más otras tantas. El argumento es que las pronunciaciones no utilizadas para entrenar el catálogo pue-

den contribuir a modelar mejor las pronunciaciones nuevas del hablante. ■

## 5. REFERENCIAS

BURTON, David K. *Text-Dependent Speaker Verification Using Vector Quantization Source Coding*. IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 35, No. 2, pp. 132-143, 1987.

CHEN, Ming-Shih, LIN, Pei-Hwa y WANG, Hsiao-Chuan. *Speaker Identification Based on Matrix Quantization Method*. IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 41, No. 1, pp. 398-405, Enero, 1993.

FURU, Sadaoki. *Spectral Analysis Technique for Automatic Speaker Verification*. IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 29, No. 2, pp. 254-272, 1981.

GRAY, Robert M. *Vector Quantization*. IEEE Transactions on Acoustics, Speech, and Signal processing Magazine, pp.4-28, Abril, 1984.

JIMENEZ, John Jairo. *Mira quién habla... Sistema de reconocimiento del hablante*. Proyecto de grado. Facultad de Ingeniería de Sistemas, Universidad Autónoma de Manizales, 1996.

JUANG, Bing-Hwang, WONG, David Y., GRAY, Augustine H. *Distortion Performance of Vector Quantization for LTP Lattice Coding*. IEEE Transactions on Acoustics, Speech, and Signal processing,

Vol. 30, No. 2, pp. 294-303, 1982.

MAKHOU, John. *Linear Prediction. A Tutorial Review*. Proceedings of the IEEE, pp. 561-580, Abril, 1975.

MAKHOU, John, ROUCOS, Salim, GISH, Herbert. *Vector Quantization in Speech Coding*. Proceedings of the IWWW, Vol. 73, No. 11, pp. 1551-1588, 1985.

MOOD, Alexander, GRAYBILL, Franklin A. y BOES, Duane C. *Introduction to the Theory of Statistics*. MacGraw-Hill, 1974.

NAIK, Jayant M. *Speaker Verification: A Tutorial*. IEEE Communications Magazine, pp. 42-48, Enero 1990.

RUEDA, Carlos Alberto. *Experimentación en reconocimiento de voz -dígitos aislados- mediante cuantización de vectores de predicción lineal y modelos escondidos de Markov*. Proyecto de Grado. Facultad de Ingeniería de Sistemas, Universidad Autónoma de Manizales, 1993.

SCHIEFFER, Richard L. y MacClave, James T. *Probabilidad y estadística para Ingeniería (1990)*. Grupo Editorial Iberoamerica, Mexico, 1993.

SOONG, Frank K. y ROSENBERG, Aaron E. *On the Use of Instantaneous and Transitional Spectral Information in Speaker Recognition*. IEEE Transactions on Acoustics, Speech, and Signal processing, Vol. 36, No. 6, pp. 871-879, Junio 1991.

TISHBY, Naftaly Z. *On the Application of Mixture AR Hidden Markov Models to Text Independent Speaker Recognition*. IEEE

Transactions on Signal Processing, Vol. 239, No. 3, pp. 563-570, Marzo 1991.

- (1) Naik, *Speaker Verification: A Tutorial*, 1990.
- (2) Furi, *Cepstral Analysis Technique for Automatic Speaker Recognition*, 1981.
- (3) Burton, *Text-Dependent Speaker Verification Using Vector Quantization Source Coding*, 1987.
- (4) Soong y Rosenberg, *On the Use of Instantaneous and Transitional Spectral Information in Speaker Recognition*, 1988.
- (5) Chen et al., *Speaker Identification Based on Matrix Quantization Method*, 1993.
- (6) Tishby, *On the Application of Mixture AR Hidden Markov Models to Text Independent Speaker Recognition*, 1991.
- (7) Jiménez, *Mira quién habla!*, Sistema

- de reconocimiento del hablante, 1996.
- (8) Makhoul, *Linear Prediction: A Tutorial Review*, 1975.
- (9) Gray, *Vector Quantization*, 1984; Makhoul et al., *Vector Quantization in Speech Coding*, 1985; Juang et al., *Distorsion performance of Vector Quantization for LPC Voice Coding*, 1982.
- (10) Burton, *Op. Cit.*
- (11) Furi, *Op. Cit.*
- (12) Jiménez, *Op. Cit.*
- (13) Rueda, *Experimentación en reconocimiento de voz -dígitos aislados- mediante cuantización de vectores de predicción lineal y modelos escondidos de Markov*, 1993.
- (14) Scheaffter y Jones, *Probabilidad y estadística para ingeniería*, pag. 219.

