# Identificación de Locutor usando Vectores Acústicos basados en Cuantiles

José-Martín Olguín-Espinoza[1], Pedro Mayorga-Ortiz[2], Luis Vizcarra[3].

[1,3]Universidad Autónoma de Baja California, Mexicali, B.C., México
[1]molguin@uabc.edu.mx, [3]luivi@uabc.edu.mx
[2]Instituto Tecnológico de Mexicali, Mexicali, B.C., México
[2]pedromayorga@hotmail.com
*Paper received on 16/07/12, Accepted on 29/08/12.*

**Resumen.** En los sistemas de reconocimiento de locutor, es importante la representación de la señal de voz en términos de vectores acústicos, ya que éstos la caracterizan. Aquí se propone el uso de vectores cuantílicos para sistemas de identificación de locutor. Para evaluar esta propuesta, se efectuaron experimentos creando modelos mezclados gaussianos de distintos tamaños usando vectores cuantílicos de distinta dimensión. Los experimentos fueron realizados con dos bases de datos en español: CEM y AHUMADA. Los resultados fueron alentadores ya que mostraron un muy buen desempeño con modelos de 10 densidades gaussianas.

**Palabras Clave:** Reconocimiento de Locutor, Vectores Cuantílicos.

## 1 Introducción

Los sistemas de reconocimiento automático de locutor modelan las características articulatorias y fisiológicas del tracto vocal, con el fin de extraer y realzar sus parámetros característicos. Entre las técnicas de extracción de parámetros, destacan *Mel Frequency Cepstral Coefficients* (MFCC), *Linear Filter Cepstral Coefficients* (LFCC), *Perceptual Linear Predictive Analysis* (PLP) y *Relative Spectral PLP* (RASTA PLP) [1, 2]. Cada técnica busca modelar el tracto vocal, o la forma en la que el oído humano funciona o como el tracto vocal genera la voz.

Una técnica escasamente encontrada en la literatura de reconocimiento automático de locutor es una que trata de aprovechar información concerniente a la capacidad pulmonar a partir de medidas acústicas. Dicho enfoque ha sido aplicado para fortalecer diagnósticos médicos, midiendo las tasas de flujo y volumen de la respiración en términos de cuantiles [3, 4].

En el presente trabajo se propone la construcción de vectores acústicos obtenidos a partir del análisis de los cuantiles en el dominio de la frecuencia como una alternativa a las técnicas ya citadas.


*Advances in Intelligent and Information Technologies.*
*Research in Computing Science 60, 2012, pp. 65 -75.*

## 2 Vectores Acústicos para Voz

El análisis de la señal de voz comúnmente se realiza sobre segmentos elementales cuasi-estacionarios llamados *ventanas de análisis* (o tramas); en este proceso se toma una señal de voz y se particiona en ventanas de cierto tamaño (típicamente en el orden de los milisegundos), las cuales están traslapadas entre sí. A cada ventana se le aplica análisis espectral para generar un vector acústico correspondiente.

Una aproximación exitosa es la deconvolución Cepstral [5], la cual permite aislar las frecuencias fundamentales de la voz de aquellas que son generadas por el conducto vocal. Esto se obtiene aplicando la transformada discreta cosenoidal (*cosine discrete transform*, CDT) a los vectores espectrales previamente calculados mediante la transformada rápida de Fourier (FFT).

Una extensión de los principios cepstrales y su paso a un espacio frecuencial no lineal relacionado con la audición humana y muy exitoso son los Mel-Frequency Cepstral Coefficients (MFCC) [1, 5], como se muestra en la Fig. 1. Una variante de MFCC es *Linear Frequency Cepstral Coefficients* (LFCC), donde los filtros son repartidos uniformemente sobre una escala lineal de frecuencias [5, 6].

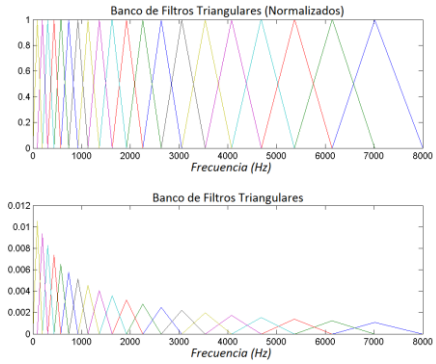


**Figura 1**. Distribución de las frecuencias en la escala Mel

Una aproximación muy difundida es la codificación lineal predictiva LPC (*Linear Predictive Coding*). Este método se basa en la hipótesis de que la voz puede ser modelada por un proceso lineal predictivo [7, 8].

Existen otros métodos, como PLP, que modifican el espectro de potencia de la voz antes de obtener una aproximación por un modelo auto-regresivo [2]. Además, hay otras variantes de esta metodología que son más adaptadas a un canal de comunicación tales como Relative Spectral PLP (RASTA PLP) [2, 7].

## 3 Antecedentes en Vectores basados en Cuantiles

Considerando evidencias en otros trabajos [3, 4], se propone analizar señales de voz, pero con un tratamiento basado en vectores cuantílicos, con el propósito de distinguir elementos que permitan caracterizar y reconocer a los locutores.

### 3.1 Cuantiles

En teoría estadística, las medidas de tendencia no-central permiten conocer aspectos particulares de una distribución (como en el caso de vectores acústicos). Dentro de estas medidas, unas de las más importantes son los *cuantiles*. En estas variables los datos son ordenados de forma creciente, dividiendo la función de distribución en partes, de tal forma que cada una contiene la misma cantidad de área.

Los cuantiles se basan en la función de distribución acumulativa (CDF). El cuantil $q_p$ de una variable aleatoria está definido como el número $q$ más pequeño tal que la función de distribución acumulativa es mayor o igual a algún valor $p$, donde $p$ se encuentra entre $0<p<1$. Esto puede ser calculado para el caso de una función de distribución continua con su función de densidad $f(x)$ resolviendo la Ec. 1:

$$p = \int_{-\infty}^{q_p} f(x)dx \tag{1}$$

Como ejemplo, los *cuartiles* pueden representarse por la notación $q_{0.25}$, $q_{0.5}$, y $q_{0.75}$, respectivamente. En esencia, estos dividen la distribución en cuatro segmentos de igual probabilidad bajo la curva (Fig. 2). El segundo cuartil es conocido como la *mediana* y es muy utilizado en estadística, el cual satisface la Ec. 2:

$$0.5 = \int_{-\infty}^{q_{0.5}} f(x)dx \tag{2}$$

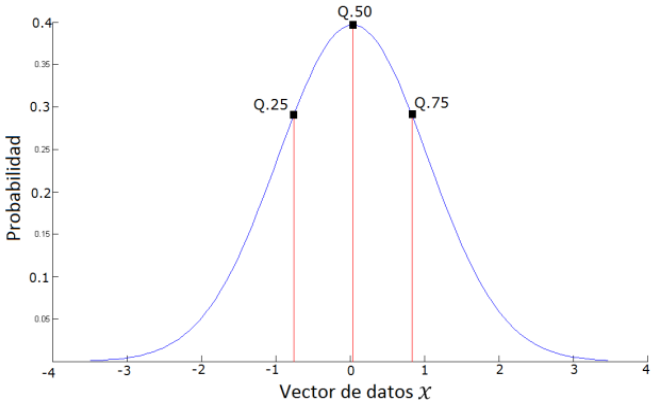


**Figura 2**. Representación gráfica del concepto de Cuartil

Una de las funciones de distribución de probabilidad (PDF) más conocidas es la distribución Gaussiana (o Normal). Donde la CDF continua está dada por la Ec. 3.

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \tag{3}$$

Donde el valor cuantílico de interés sería $x$. Cuando no es conveniente suponer una PDF específica, podemos aplicar la CDF, la cual se obtiene a partir de histogra-

mas de frecuencias de los vectores acústicos, sobre los cuales se pueden ajustar modelos mezclados gaussianos (GMM).

### 3.2 Vectores Acústicos basados en Cuantiles

Lo que se propone en el presente trabajo es, a partir de los datos de la señal de voz *X* obtener su FFT, normalizarla y tomarla como la función de distribución de frecuencias o una función de densidad de frecuencia. De esta forma se puede obtener un vector $Q = (q_1 , q_2 , ... , qn)$ sobre los valores frecuenciales, donde cada $q_i$ es el valor de la frecuencia asociado al valor porcentual acumulado bajo la CDF específico del cuantil *i*. En otras palabras, *Q* representará un vector cuantílico para la señal *X*. De esta forma podemos hablar de vectores cuartílicos, cuando sus elementos representan los valores frecuenciales para el 25%, 50% y 75% del área bajo la curva de la FFT de *X* normalizada; o vectores octílicos en el caso de obtener estos valores para 12.5%, 25%, 37.5%, 50%, 62.5%, 75% y 87.5% del área.

El vector cuantílico se tomará como el vector acústico que representa un segmento, dígase estacionario, de la señal original *X* para efectos del modelado del locutor. Bajo esta idea los vectores se pueden obtener analizando la señal de voz de un locutor en dos modalidades: tiempo largo y tiempo corto.

El análisis en tiempo largo consiste en obtener un sólo vector cuantílico resultante de la señal de un registro completo de entrada (Fig. 3). Por otra parte, en el análisis en tiempo tiempo corto, se definen ventanas de cierta longitud de tiempo *w* (regularmente en el orden de los milisegundos y dentro del rango estacionario de la señal), de tal forma que por cada ventana se genera un vector. Además, se establece un valor de traslape $o < w$, de tal forma que el análisis se va realizando por tramas traslapadas, a fin de hacer la extracción de características en segmentos estacionarios. En análisis de tiempo corto se tendrán $T/(w-o)$ vectores para cada señal, donde *T* es el tiempo total de duración de la señal, *w* es el tamaño de la ventana y *o* representa el tiempo de traslape. Esta modalidad se esquematiza en la Fig. 4.

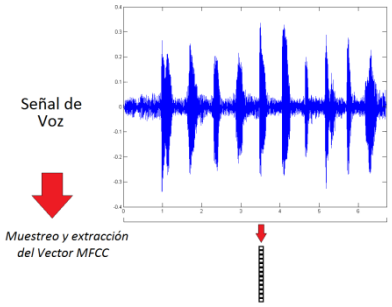


**Figura 3.** Vectores en tiempo largo.

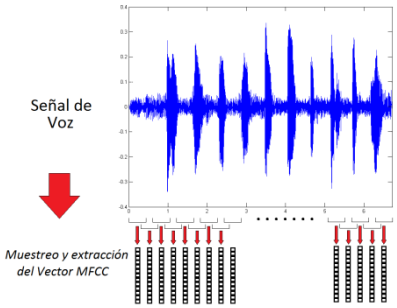


**Figura 4**. Vectores en tiempo corto.

# 4 Reconocimiento Automático de Locutor con GMM

El reconocimiento automático de locutor (RAL) es un término genérico que denota la identificación y verificación de locutor, siendo la primera el objetivo de este trabajo.

## 4.1 Identificación Automática del Locutor

La Identificación Automática del Locutor consiste en determinar de entre una población de locutores conocidos, la persona a la que pertenece cierta señal de voz dada como entrada. En la identificación se proponen dos modos: *Identificación en conjunto cerrado*, para el cual se asume que la señal de voz es pronunciada por un locutor conocido por el sistema. La salida del sistema de identificación en este modo será el identificador del locutor con la mayor similitud a la señal de voz de entrada. Por otro lado, en *Identificación en conjunto abierto*, para la cual cabe la posibilidad de que el locutor pueda no pertenecer al conjunto de locutores conocidos por el sistema, es decir que el locutor sea un impostor. En identificación en conjunto abierto, el sistema de identificación debe decidir la fiabilidad de su juicio aceptando o rechazando la identidad que encontró. Si el sistema la acepta debe además establecer el identificador del locutor al que pertenece la señal de voz de entrada.

## 4.2 Modelado de Locutor con Modelos Mezclados Gaussianos

Los GMM [9] consisten en una suma ponderada de *M* componentes de densidades, como lo muestra la Ec. 4:

$$p(\vec{x} \mid \lambda) = \sum_{i=1}^{M} m_i b_i(\vec{x}) \tag{4}$$

Donde $\vec{x}$ es un vector aleatorio d-dimensional; $b_i(\vec{x})$ son las densidades, $m_i$ representan las ponderaciones para cada mezcla o densidad Gaussiana, con la restricción de $\sum_{i=1}^{M} m_i = 1$ lo que es característico en una verdadera función de densidad de probabilidad. Cada componente de densidad es una función Gaussiana de dimensión D descrita por:

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \overrightarrow{\mu_i})' \Sigma_i^{-1} (\vec{x} - \overrightarrow{\mu_i})\right\} \tag{5}$$

Donde $\overrightarrow{\mu_i}$corresponde al vector de medias y $\Sigma_i$ es la matriz de covarianza. Un modelo GMM completo está parametrizado por $m_i$, $\overrightarrow{\mu_i}$ y $\Sigma_i$. En reconocimiento, cada locutor *r* está representado por un modelo GMM de la siguiente forma: $\lambda_r = \{m_i, \overrightarrow{\mu_i}, \Sigma_i\}$; $i = 1, \ldots, M$ donde *M* representa el número de componentes gaussianas utilizadas para modelar al locutor *r*.

### 4.3 Identificación de Locutor con GMM y UBM

Los sistemas de identificación se dividen en dos etapas: entrenamiento y evaluación. En el entrenamiento se toman las señales de cada locutor para construir su modelo GMM correspondiente. Para realizar esta tarea, el algoritmo de Máxima Expectación (EM) es uno de los más exitosos [5, 10, 11, 12]. EM parte de un modelo $\lambda$ y mediante iteraciones sucesivas trata de encontrar un $\lambda$ tal que:

$$p(X \mid \bar{\lambda}) \geq p(X \mid \lambda) \tag{6}$$

El nuevo modelo resultante de cada iteración se toma como modelo inicial para la siguiente, y el proceso se repite hasta que se alcanza un umbral de convergencia basado en la maximización de la probabilidad. En cada iteración EM, se deben reestimar nuevos parámetros del modelo $\bar{\lambda}$. La elección del modelo inicial $\lambda$ se puede hacer de varias maneras: Una es calcularlo aleatoriamente, la otra es utilizar lo que se conoce como el Universal Background Model (UBM). El UBM es un modelo que se construye a partir de las señales combinadas de todos los locutores [12], de esta forma se obtiene un modelo cuyos parámetros se usarán como valores iniciales en el algoritmo EM en el cálculo de cada modelo cliente.

La evaluación comprende seleccionar al modelo $\lambda_i$ que resulte con la probabilidad más alta de una señal *x* desconocida, como se muestra en la Ec. 7:

$$\prod_{t=1}^{T} p(\overrightarrow{x_t} \mid \lambda_i) > \prod_{t=1}^{T} p(\overrightarrow{x_t} \mid \lambda_r)\,; r = 1, \ldots, I \tag{7}$$

## 5 Experimentos

Se construyó un Sistema de Identificación de Locutor, el cual fue evaluado con diferentes bases de datos de voz. Los vectores acústicos fueron construidos en base a cuartiles, octiles y deciles, eligiendo los que presentaran mejor nivel de reconocimiento.

### 5.1 Base de Datos de señales de Voz

El sistema de identificación fue validado con dos diferentes bases de datos: Corpus en Español Mexicano [13] y AHUMADA [14], las características de cada una se describen en los apartados siguientes.

#### 5.1.1 Corpus en Español Mexicano (CEM)

Esta base de señales de voz contiene grabaciones de 33 personas (20 hombres y 13 mujeres), y fue desarrollada por la Universidad Autónoma de Baja California (UABC) [13]. Cada locutor grabó a lo largo de tres sesiones, con al menos 15 días de separación, tres frases siguiendo la distribución fonética del español mexicano y

un texto fijo; además durante cada sesión se grabaron elocuciones tanto con micrófono como con teléfono y utilizando un sistema de voz sobre IP (VoIP), totalizando 36 grabaciones por persona. Cada señal fue muestreada a 8Khz, 16 bits por dato, un canal (mono) y almacenados en formato WAV.

### 5.1.2 AHUMADA

Corpus en español Ibérico [14], consiste de registros de 25 locutores (hombres) los cuales grabaron dígitos, frases fonéticamente equilibradas, texto fijo, texto específico para cada locutor y además conversación espontánea. Cada tipo de elocución se grabó por medio de micrófono y teléfono. Cada elocución de las frases se repitió en tres sesiones al menos, con 15 días de separación entre ellas.

## 5.2 Obtención de Vectores Cuantílicos

En la obtención de los vectores se parte de un archivo WAV, seguida opcionalmente de un preprocesamiento de preénfasis y antitraslape [5]. Posteriormente, se aplica la transformada rápida de Fourier (FFT). Para cumplir con el principio básico para una distribución de probabilidad, se normaliza el área bajo la curva, garantizando que sea igual a 1. Enseguida, se buscan los valores para los cuantiles. Por ejemplo, para octiles se calculan los valores frecuenciales $f_{0.125}$, … , $f_{0.875}$, aplicando la Ec. 8. En los experimentos realizados se utilizaron cuartiles, octiles y deciles para comparar el porcentaje de identificación con diferentes configuraciones.

$$A = .125 = \int_{-\infty}^{f_{.125}} F_N(f)df, \ldots, A = .875 = \int_{-\infty}^{f_{.875}} F_N(f)df \qquad (8)$$

## 5.3 Construcción de los Modelos de Locutor

Una vez obtenidos los vectores cuantílicos para todos los locutores, se construyeron los modelos GMM de cada locutor con dos configuraciones. La primera utilizando un modelo aleatorio como modelo inicial para el algoritmo Expectación-Maximización (EM). Para la segunda configuración se creó un Universal Background Model (UBM) utilizando los datos de entrenamiento de todos los locutores combinados en un solo modelo GMM [12], posteriormente los modelos para cada locutor se crearon utilizando el UBM como modelo inicial para el algoritmo EM.

Los modelos GMM tambien fueron creados con configuraciones de 10 y 32 componentes gaussianas para las distintas combinaciones de tamaños de vectores (cuartiles, octiles y deciles).

# 6 Resultados y Discusión

Debido a que las bases de datos no tienen el mismo tamaño no fueron realizadas siguiendo el mismo protocolo, los experimentos tuvieron variaciones en cuanto al

número de archivos utilizados para entrenamiento y evaluación. En consecuencia, los experimentos serán explicados por separado para cada base de datos. Sin embargo, lo importante dentro de nuestra contribución está en mostrar la potencialidad de los vectores cuantílicos con distintas bases de datos las cuales fueron creadas bajo distintas condiciones de grabación e idioma.

Para el caso particular de la base de datos CEM se usaron los 20 locutores masculinos, tomando 40 segundos de entrenamiento, correspondiendo a las dos primeras sesiones de las frases 1,2 y 3, que son fonéticamente equilibradas. Para la evaluación se utilizaron 60 segundos correspondientes a la frase 4 (texto fijo) de la sesión 3. Debido a que nuestro sistema está orientado a la identificación de locutor, la medición de eficiencia fue efectuada en función de las clasificaciones correctas de locutor. La Tabla 1 muestra los resultados de eficiencia de reconocimiento considerando entrenamiento y evaluación de señales adquiridas por el mismo medio micrófono (M1), teléfono (T1) y VoIP (T3). La parte superior de la tabla indica el tipo de cuantil, el número de componentes gaussianas en los modelos GMM y los medios de adquisición/evaluación. El número de componentes de los vectores fueron 3, 7 y 9 para cuartiles, octiles y deciles. En este experimento se obtuvieron los mejores resultados con los vectores de mayor dimensión (deciles).

**Tabla 1**. Experimentos con CEM sin pre-procesamiento.

| Cuantil | GMM | MI | T1 | T3 |
|---|---|---|---|---|
| Octil | 10 | 35 % | 32.00 % | 36.66 % |
| Cuartil | 10 | 25 % | 8.33 % | 31.66 % |
| Decil | 10 | 47 % | 20.00 % | 40.00 % |
| Octil | 32 | 20 % | 17.00 % | 23.00 % |
| Cuartil | 32 | 17 % | 8.33 % | 26.66 % |
| Decil | 32 | 33 % | 12.00 % | 35.00 % |

Lo anterior pone de manifiesto que el vector cuantílico con más dimensiones captura con mayor precisión la estructura fina y/o densidad espectral importante del aparato fonatorio del locutor. Asimismo se observa que los resultados no mejoran cuando se aumenta el número de componentes gaussianas en los modelos GMM, denotándose mejores resultados con mezclados de 10 componentes gaussianas. Lo anterior no significa que modelos más pequeños mejoren los resultados, puesto que aquí no mostramos otras evaluaciones menos satisfactorias con modelos de menor tamaño. Lo que sí se puede observar es que modelos con más de 10 gaussianas pueden conducir a errores de otra naturaleza como sobreentrenamiento o redondeo para este sistema en particular. Otro aspecto interesante que nos permite emitir un juicio es que el mejor resultado fue con deciles y directamente con micrófono. Este hecho pone de manifiesto por un lado que el decil capturó mejor la estructura fina del locutor y que el usar directamente micrófono implicaba menos degradación en los modelos debida a ruido o a efectos de canal.

En base a estos resultados, se decidió construir modelos con vectores decílicos y tratando de mejorar los resultados se aplicó preprocesamiento que consiste en preénfasis en segmentos de tiempo corto de 30ms y traslape de 20ms. Además se construyó el modelo impostor aplicando UBM. En la Tabla 2 se puede observar una me-

joría evidente en la eficiencia de identificación de locutor. De nuevo los modelos con 10 GMM arrojaron resultados ligeramente mejores que los de 32 GMM. Lo anterior refuerza la idea de que modelos más complejos pueden conducir a errores de redondeo o problemas de sobre-entrenamiento y que los vectores decílicos son más consistentes con modelos de 10 gaussianas. Otro aspecto relevante es que tanto el preénfasis como el uso de UBM lograron paliar en gran medida los problemas derivados de los medios de transmisión o canal como se observa para el caso de la señal de teléfono (T1). Asimismo estos resultados refuerzan la hipótesis de que los cuantiles para señales orientadas a la identificación de locutor tiene potencial de aplicación.

**Tabla 2.** Experimentos con CEM y modelos GMM usando UBM

| Cuantiles | GMM | MI | T1 |
|---|---|---|---|
| Decil | 10 | 75 % | 80.00 % |
| Decil | 32 | 75 % | 75.00 % |

Debido a que este trabajo presenta dos variantes con respecto a la mayoría de los trabajos en identificación de locutor, las cuales son, el uso de un corpus en español mexicano y la aplicación de vectores cuantílicos, a manera de comparación se realizaron experimentos con la bases de datos AHUMADA, la cual contiene grabaciones en español Ibérico. Se trató de establecer una partición entrenamiento/prueba con condiciones similares a las utilizadas en los experimentos con CEM. Para este caso se tomaron las grabaciones de 20 locutores, el entrenamiento se realizó con las 10 elocuciones fonéticamente equilibradas de las sesiones 1 y 2 (identificadas con el código C), para tener un total aproximado de 40 segundos de señal. Para la evaluación se usó la elocución fonéticamente equilibrada (identificada con el código D) de la sesión 3 de aproximadamente 60 segundos de duración.

En la Tabla 3 se muestran los resultados de estos experimentos, notándose la similitud con los mostrados en la Tabla 2. Los resultados aquí obtenidos con micrófono son iguales, lo cual demuestra consistencia en los métodos de grabación, sin embargo las grabaciones efectuadas por teléfono evidencian un mejor control en el caso de CEM. Lo cual es consistente con lo reportado por los autores de AHUMADA [14]. Un aspecto interesante a resaltar es que los modelos acústicos con 32 gaussianas lograron mejorar la eficiencia de reconocimiento con respecto a los de 10, es decir de alguna manera se adaptaron más a la distorsión debida al canal de transmisión. Pero lo más importante es que nuevamente los resultados refuerzan el potencial de los vectores cuantílicos.

**Tabla 3.** Experimentos con AHUMADA y modelos GMM usando UBM

| Cuantiles | GMM | MI | T1 |
|---|---|---|---|
| 10 | 10 | 75 % | 25.00 % |
| 10 | 32 | 75 % | 30.00 % |

## 7 Conclusiones

Los vectores acústicos cuantílicos como una propuesta novedosa en el campo de vectores acústicos, cuya representación está sustentada en trabajos relacionados con medicina sobre la capacidad de flujo respiratorio, que para el caso de voz siguió un tratamiento de transformada de Fourier en tiempo corto para tomar en cuenta la cuasi-estacionariedad de la señal.

Dichos vectores proporcionan la oportunidad de relacionar la energía con los valores frecuenciales más importantes de la señal de voz. Hasta el momento se obtuvo una tasa de identificación correcta del 80% en análisis de tiempo corto para elocuciones obtenidas de CEM. Los vectores cuantílicos fueron aplicados sobre señales con un mínimo de preprocesamiento (preénfasis y antialiasing) lo cual deja la posibilidad de explorar variantes mas elaboradas de esta técnica e intentar obtener mejores resultados.

Como trabajo futuro se plantea mejorar el sistema aplicando diversas técnicas tanto en el preprocesamiento, tales como suprimir información no correspondiente a voz (VAD) y técnicas de normalización como las reportadas en [5]; así como en el cálculo de modelos GMM (entrenamiento con MAP), todo orientado a mejorar la eficiencia de la tarea de identificación de locutor.

### Referencias


1. Faundez-Zanuy, M.; Monte-Moreno, E.: State-of-the-art in speaker recognition, Aerospace and Electronic Systems Magazine, IEEE , vol.20, no.5, pp.7-12, May 2005 doi: 10.1109/MAES.2005.1432568.
2. Hermansky H. and Fousek P.: Multi-resolution RASTA filtering for TANDEM-based ASR, in Proceedings of the European Conference on Speech Communication and Technology, Lisbon, Portugal, 2005.
3. Mayorga P., Druzgalski C., González O.H., Zazueta A., Criollo M.A.: Expanded Quantitative Models for Assessment of Respiratory Diseases and Monitoring , PAHCE 2011 IEEE, Rio de Janeiro. March 2011; ISBN 978-1-4244-6291-9; DOI: 10.1109/PAHCE.2011.5871938
4. Mayorga P., Druzgalski C., González O.H.: Quantile Vectors based Verification of Normal Lung Sounds, PAHCE 2012 IEEE, Miami, USA. March 2012; ISBN 978-1-4244-6291-9; DOI:
5. Bimbot F., Bonastre J-F., Corinne Fredouille, Guillaume Gravier, Ivan Magrin-Chagnolleau, Sylvain Meignier, Teva Merlin, Javier Ortega-Garcia, Dijana Petrovska-Delacretaz, and Douglas A. Reynolds.: A tutorial on text-independent speaker verification, EURASIP J. Appl. Signal Process. 2004 (January 2004), 430-451. DOI=10.1155/S1110865704310024 http://dx.doi.org/10.1155/S1110865704310024
6. Istrate D.M.: Detection et Reconnaissance des Sons pour la Surveillance Médicale, These pour obtenir le grade de docteur de l'INPG, spécialité Signal, Image, Parole, Télécoms, le 16 decembre 2003, 183 p.
7. Solé-Casals J., Zaiats V.: Advances in Nonlinear Speech Processing, International Conference on Nonlinear Speech Processing, Nolisp 2009, Vic, Spain, June 25-27, 2009, Springer, 2010, ISBN 364211508X, 9783642115080.
8. Milner B. and James A.: Robust Speech Recognition Over Mobile and IP Networks in Burst-Like Packet Loss, IEEE Transactions On Audio, Speech, And Language Pro-

cessing, Vol. 14, No. 1, January 2006, ISSN: 1558-7916, DOI:10.1109/TSA.2005.852997.

9. Reynolds D. A.: An Overview of Automatic Speaker Recognition Technology, ICASSP 2002 (IEEE International Conference on Acoustics, Speech and Signal Processing) , Orlando, Florida, USA, May 13 - 17, 2002.
10. Dempster A. P., Laird N. M. and Rubin D. B.: Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. (B), vol. 39, no. 1, pp. 1–38, 1977
11. Martinez W.L. and Martinez A.R.: Computational Statistics Handbook with Matlab, Second Edition, Chapman & Hall/CRC, 2008, ISBN 1-58488-566-1.
12. Reynolds D.A., Quatieri T.F., and Dunn R.B.: Speaker Verification Using Adapted Gaussian Mixture Models Digital Signal Processing 10, 19–41 (2000) doi:10.1006/dspr.1999.0361
13. Olguín J.M. and Mayorga P.: Corpus de Voz en Español Mexicano Para Experimentación en Reconocimiento Automático de Locutor, RESEARCH IN COMPUTING SCIENCE, CIC-IPN, 2010, Vol. 50, ISSN 18770-4069, México.
14. Ortega-García J., González-Rodríguez J., Marrero-Aguiar V., Díaz-Gómez J., García-Jimenez M., Lucena-Molina J., Sánchez-Molero J.: AHUMADA: A large speech corpus in spanish for speaker Characterization and Identification, Speech Communication, Vol. 3, pp. 255-264, Junio 2000.