

Corpus de Voz en Español Mexicano para Experimentación en Reconocimiento Automático de Locutor

José-Martín Olguín-Espinoza¹ y Pedro Mayorga-Ortiz²

¹ Universidad Autónoma de Baja California, Mexicali, B.C., México
molguin@uabc.edu.mx

² Instituto Tecnológico de Mexicali, Mexicali, B.C., México
pedromayorga@hotmail.com

Paper received on 29/03/10 , Accepted on 20/09/10.

Resumen. Un elemento indispensable en la construcción de un sistema de Reconocimiento Automático de Locutor (RAL) es el corpus de voz, el cual es necesario para la experimentación y evaluación. Las condiciones que debe cumplir un corpus para RAL es que contenga grabaciones de frases fonéticamente equilibradas de varios locutores usando distintos medios de grabación y a lo largo de varias sesiones. En este trabajo se presenta la metodología y resultados de la construcción de un corpus de voz para RAL con locutores de español mexicano, así como la comparación con estudios previos.

Palabras Clave: Reconocimiento Automático de Locutor, Corpus de Voz.

1 Introducción

Debido a que la voz es el instrumento de comunicación preferida de los humanos, utilizar Reconocimiento Automático de Locutor (RAL) es de gran importancia [10]. Las modalidades del RAL se subdividen en dos tareas [9]: la primera consiste en determinar si el emisor es quien dice ser, esto se conoce como Verificación Automática de Locutor (VAL); la segunda consiste en determinar quién es el emisor, conocida como Identificación Automática de Locutor (IAL). Esto es, para la VAL el locutor provee la señal de voz e información adicional (como un número de identificación personal o PIN) y el sistema determina si la voz corresponde a ese locutor. En el caso de la IAL, sólo se tiene como entrada la señal de voz, siendo la salida del sistema de reconocimiento el identificador del locutor a quien corresponde esa señal. Recientemente Patil y Basu [16] añaden a estas tareas la Clasificación Automática de Locutor (CAL), la cual trata de agrupar a los locutores de acuerdo a una característica en común, como el acento o manera de pronunciar ciertos fonemas. Esta tarea es importante cuando se trata de identificar por ejemplo la zona geográfica de donde proviene cierto locutor (regionalismos en el habla), por lo que en este caso en

particular es indispensable contar con modelos específicos para cada región geográfica que se desea ubicar.

Para la construcción de los sistemas de RAL es necesario tener señales de voz de los locutores que serán reconocidos, esta señal se tiene que procesar con la finalidad de extraer los elementos que permitan caracterizar de manera única a cada locutor (o conjunto de locutores). Al conjunto de señales de voz de varios locutores se le denomina Corpus de Voz y para tener sistemas RAL robustos, es necesario que el corpus cumpla con varios aspectos entre los que destacan: a) Incluir todos los fonemas del lenguaje; y b) Preservar la distribución fonética del lenguaje [17].

En relación al español mexicano, la literatura encontrada es escasa y orientada al Reconocimiento Automático del Habla (RAH), por lo mismo los corpora encontrados han sido construidos con este fin [17],[11], razón por la cual la contribución principal del presente trabajo es la construcción de un corpus en español mexicano para experimentación con sistemas de RAL.

2 Corpus de Voz

Para propósitos del presente trabajo, los corpora de voz para experimentación se pueden clasificar en dos tipos: los orientados a RAH y los orientados a RAL. Los primeros deben permitir analizar las diferencias fonéticas del idioma independientemente del locutor del cual provenga la señal, para lograr esto es necesario grabar las mismas frases para una gran cantidad de locutores, no es tan importante el número de sesiones por locutor, lo que importa es caracterizar los fonemas para un amplio rango de locutores. Por otro lado, los corpora para RAL deben ser multisesión para que el sistema pueda discriminar los elementos de variabilidad intralocutor a lo largo del tiempo y resaltar la variabilidad interlocutor para poder caracterizar a las personas. Esto impone una mayor complejidad en la construcción de este tipo de corpora porque es necesario llevar un seguimiento del locutor a lo largo de la etapa de grabación, debido al tiempo que debe transcurrir entre una sesión y otra, el cual puede ser de semanas a meses.

2.1 Corpora para RAL en Idiomas Distintos al Español

Polycost. Frases en inglés de 134 locutores de 13 países europeos (de habla inglesa y habla no inglesa), más de 5 sesiones por cada locutor grabadas por medio de aparatos telefónicos a través de líneas digitales ISDN [8].

YOHO. Diseñado para evaluar sistemas de verificación de locutor en situaciones dependientes del texto. Consiste de 138 locutores (106 hombres, 32 mujeres) que grabaron frases compuestas de dígitos en idioma inglés [2].

Switchboard I-II. Un corpus en inglés con más de 500 locutores, las grabaciones fueron realizadas mediante un sistema telefónico automático que conectaba a un participante con otro y grababa la conversación. Un subconjunto de este corpus es utilizado en las evaluaciones de reconocimiento de locutor que regularmente organiza el Instituto Nacional de Estándares y Tecnología de Estados Unidos (NIST) [18].

SIVA. Base de señales en italiano, contiene 691 locutores (335 hombres, 336 mujeres), grabado en múltiples sesiones utilizando líneas telefónicas [3,8].

XM2VTSDB. Corpus multimodal que incluye cerca de 300 locutores ingleses en grabaciones de audio y video a lo largo de cuatro sesiones. Es un producto en constante evolución cuyos orígenes fueron M2VTS en francés y XM2VTS en inglés [13].

2.2 Corpora para RAL en Español

AHUMADA. Corpus en español ibérico, consiste de 28 locutores hombres. Fue desarrollado en el contexto de un proyecto de investigación forense [14]. En lo que respecta al alcance del presente trabajo de investigación, este es el único corpus encontrado para RAL, que ha sido desarrollado en un país de habla hispana.

Otro corpus en español ibérico encontrado en la literatura es el denominado *Albayzin*, pero está orientado al reconocimiento del habla [4,21]

2.3 Corpora de Voz en Español Mexicano

Hasta donde tienen conocimiento los autores, no existe un corpus para RAL que haya sido desarrollado en México, los siguientes corpora encontrados están orientados a reconocimiento del habla:

DIMEX-100. Desarrollado en el Instituto de Investigaciones en Matemáticas Aplicadas de la Universidad Nacional Autónoma de México (IIMAS-UNAM). Es un corpus en español mexicano orientado al reconocimiento del habla, incluye grabaciones de 100 locutores del centro del México, incluye un estudio detallado de los fonemas y sus correspondientes alófonos para el español mexicano [17].

TLATOA. Desarrollado por la Universidad de las Américas en Puebla (UDLA). Consiste de grabaciones de 550 adultos principalmente del centro de México, está orientado a experimentación con reconocimiento del habla [11].

3. Distribución Fonética del Español Mexicano

En [1,6] se establece la ventaja de asignarles peso a los fonemas para mejorar el reconocimiento, por lo tanto es importante que un corpus de voz contenga al menos los fonemas más utilizados en el lenguaje nativo de los locutores. Por esta razón, el primer paso para la construcción del corpus es la definición de una o más frases que los locutores deben grabar. Sin embargo, para definir estas frases es necesario conocer la distribución fonética del lenguaje y en base a ésta construir las frases fonéticamente equilibradas, tal y como se reporta en la construcción del corpus AHUMADA [14]. Pérez [15] presenta la distribución fonética del español latinoamericano obtenida mediante el análisis de grabaciones de audio correspondientes a noticieros chilenos, Villaseñor-Pineda [20] la obtuvo a partir de fuentes de texto de páginas web, ambos trabajos concuerdan en que la diferencia entre el español ibérico y el latinoamericano (chileno y mexicano) es la frecuencia de uso de los fonemas /a/ y /e/; en el español ibérico se utiliza más el fonema /a/ que el fonema /e/ y en el

español latinoamericano es al contrario, tiene una mayor frecuencia de uso el fonema /e/ que el fonema /a/.

3.1 Recolección de Datos

Para este trabajo se utilizó un enfoque similar al presentado por [20], esto es, se utilizaron páginas web como única fuente de información. La principal diferencia consistió en el tipo de fuentes de datos utilizadas y el volumen total de información recolectada (se requirió alrededor de 90% menos datos). Las páginas web fueron seleccionadas de acuerdo a los siguientes criterios:

1. Páginas web en español de periódicos mexicanos de diferentes áreas geográficas. Para esto se utilizó la distribución de zonas propuesta por una editorial de periódicos mexicana cuyo portal de Internet publica los diarios que tiene a lo largo de todo el país (7 zonas: Norte, Golfo, Noroeste, Centro, Bajío, Sur, Este).
2. Con la finalidad de capturar las palabras utilizadas específicamente en esa zonas geográficas (regionalismos), sólo se utilizaron las páginas correspondientes a las noticias locales, las cuales se asume fueron escritas por reporteros locales.

En total se recabó información de 39 periódicos en línea, durante un lapso de 20 días, acumulando un total de 157MB de texto. Se generaron 6 bases de datos, una por región, en cada una se vació la información recabada separando las palabras y su correspondiente número de ocurrencias. La información fue almacenada por separado para poder hacer un análisis por áreas geográficas y así poder identificar diferencias en el uso de los fonemas entre zonas.

3.2 Análisis de Datos y Resultados

Los fonemas analizados corresponden a la definición básica en [15], la cual consiste de 22 fonemas básicos para el español y sus correspondientes alófonos, aunque para este trabajo no se consideran estos últimos. Se aplicó un algoritmo de detección de fonemas a la información recabada en cada una de las zonas, los resultados obtenidos no mostraron diferencias significativas de la frecuencia de los fonemas entre las diversas zonas. Al final se agregó toda la información y se obtuvieron las frecuencias finales, en la Tabla 1 se muestran los porcentajes finales.

3.3 Validación con Trabajos Previos

Los resultados obtenidos en la distribución fonética muestran el mismo comportamiento reportado en [15,20] para el español latinoamericano, el análisis de correlación entre lo encontrado y esos trabajos se muestra en la Tabla 2.

Tabla 1. Frecuencia encontrada para cada uno de los Fonemas.

Fonema	Frecuencia	Fonema	Frecuencia	Fonema	Frecuencia
/i/	7.63%	/e/	13.85%	/a/	12.69%
/o/	9.37%	/u/	2.96%	/p/	2.81%
/t/	4.60%	/k/	4.07%	/b/	2.05%
/d/	5.41%	/g/	0.44%	/f/	0.78%
/s/	9.66%	/j/	0.87%	/ch/	0.18%
/ll/	0.45%	/m/	2.66%	/n/	7.21%
/ñ/	0.16%	/r/	6.59%	/rr/	0.18%
/l/	5.37%				

Tabla 2. Coeficiente de correlación con trabajos previos.

Coeficiente de correlación	
Villaseñor-Pineda (2003)	0.9960
Pérez (2003)	0.9974

4 Desarrollo del Corpus

Una vez establecida la distribución fonética, se diseñaron las frases fonéticamente equilibradas y se realizaron las grabaciones.

4.1 Protocolo de Grabación

Una vez establecida la distribución fonética se definió el protocolo de grabación, el cual incluyó las siguientes características:

1. Por cada locutor, grabar 3 frases fonéticamente equilibradas (4, 6 y 10) segundos respectivamente y un texto adicional no balanceado (aprox. 60 segundos).
2. Realizar tres sesiones por locutor.
3. Utilizar dos tipos de dispositivo de adquisición: micrófono y aparato telefónico.
4. Tres modalidades: sistema telefónico analógico, grabación directa de micrófono y sistema telefónico de VoIP.

4.2. Software y Hardware utilizado

Para las grabaciones telefónicas se configuró el software Asterisk, el cual es un servidor de comunicaciones de voz para líneas telefónicas analógicas y líneas de VoIP, bajo el sistema operativo Linux. El hardware utilizado fue una computadora

con procesador Pentium 4 de 3GHz de velocidad, 2GB de RAM, una tarjeta de puertos telefónicos Digium modelo 1TDM422EF. Las grabaciones de VoIP se realizaron a través de una red local de datos inalámbrica 803.11g utilizando el softphone Zoiper y un micrófono tipo cardiod. Para las grabaciones analógicas y las de VoIP se configuró un sistema de contestación automática en el servidor Asterisk, el cual consistió de un menú interactivo para establecer el número de locutor, tipo de medio, número de frase y número de sesión. Todas las grabaciones telefónicas se almacenaron en el sistema de archivos del servidor Linux. Para las grabaciones directas se usó un micrófono de escritorio tipo cardiod conectado por puerto USB a una computadora Pentium 4 de 3GHz, 2GB de RAM con sistema operativo Windows XP SP3 y el software Audacity para procesar las señales de audio.

4.3 Sesiones de Grabación

Todas las grabaciones se realizaron en un cuarto especialmente acondicionado para tal fin, con dimensiones de 3x3 mts. y las paredes cubiertas con espuma fono-absortora de poliuretano con cuñas anecoicas para disminuir la reverberación acústica. Las sesiones se llevaron a cabo a lo largo de 18 meses, participaron 50 personas, 31 hombres y 19 mujeres. 23 hombres completaron dos sesiones y 18 completaron las tres sesiones, en el caso de las mujeres, 14 completaron dos sesiones y 13 completaron las 3 sesiones. En total el corpus contiene 31 locutores con las tres sesiones completas. El tiempo transcurrido entre sesiones varía de 1 a 5 semanas. El tiempo promedio que cada locutor le dedicó por sesión fue de 20 minutos.

4.4 Organización del Corpus

El corpus está organizado por directorios, uno para cada locutor, el nombre del directorio es el identificador del locutor e.g. L01 indica al locutor 01. Cada archivo del corpus se nombró utilizando la siguiente nomenclatura:

L<ID de locutor><Tipo de Grabación>F<ID de Frase>S<ID de Sesión>
<ID de locutor>: Número consecutivo asignado a cada locutor con fines de seguimiento e identificación en el corpus.
<Tipo de Grabación>: Identificador del tipo de adquisición de la señal, MI para micrófono, T1 para teléfono analógico, T2 para VoIP.
<ID de Frase>: Identificador de la frase, puede tomar los valores de 1, 2, 3 ó 4.
<ID de Sesión>: Identificador de la sesión en que fue grabada esa señal, puede ser 1, 2 ó 3.
Por ejemplo, el archivo *L01MIF2S3* corresponde a la grabación directa de micrófono del locutor 01, frase 2, sesión 3.

4.5 Experiencias Obtenidas durante la Grabación

Durante la etapa de grabación se identificaron dos aspectos a considerar en los proyectos de este tipo:

1. Cuando se construyeron las frases fonéticamente equilibradas, no se consideró importante que fueran frases coherentes, lo único que se buscaba es que cumplieran con la distribución de frecuencia de fonemas, se trató de seguir la idea de las frases de XM2VTSDB [13], las cuales son una secuencia de palabras sin significado. Sin embargo los locutores tendían a confundirse fácilmente en las grabaciones, por lo que se optó por cambiar las frases por otras más coherentes, esto disminuyó el número de errores que el locutor cometió por sesión y por lo tanto disminuyó el tiempo promedio por sesión.
2. Dado que los participantes no tienen claro el objetivo de las grabaciones, es necesario estar constantemente explicando los alcances de la investigación, ya que sienten cierta incomodidad al no tener certeza del uso que se les dará a sus datos. La certeza es indispensable para conseguir de ellos las sesiones subsecuentes.

5 Evaluación del Corpus

Para la evaluación del corpus se implementó un sistema de IAL basado en Modelos Mezclados Gaussianos (GMM) [19], utilizando las herramientas open-source utilizadas en [7], la arquitectura se muestra en la Fig. 1.

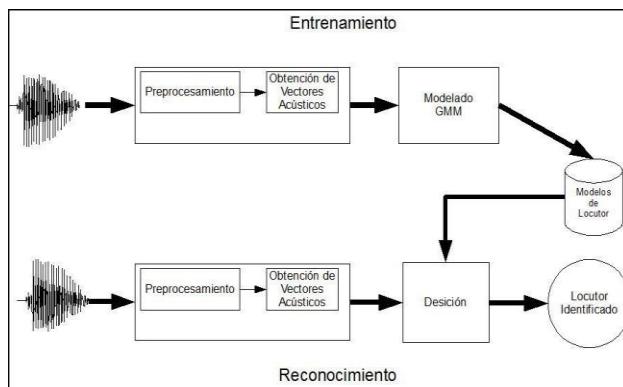


Fig. 1. Arquitectura del Sistema de IAL utilizado para la evaluación del Corpus.

Para la construcción de vectores acústicos se utilizó una ventana de 30 ms con un traslape de 10 ms, se aplicó un banco de filtros de 26 canales para obtener un vector de 41 elementos (13 coeficientes MFCC, 13 valores de la primera derivada, 13 de la segunda derivada y 2 coeficientes de energía) además se aplicó sustracción de medias cepstrales (CMS) para atenuar el ruido del medio. Para el modelado de locutores primero se construyó un modelo impostor con 32 mezclados gaussianos con matriz de covarianza diagonal, el entrenamiento se llevó a cabo con el algoritmo Expectación-Maximización (EM); cada modelo cliente se obtuvo mediante adaptación máxima a posteriori (MAP).

5.1 Evaluación

El corpus se particionó en dos conjuntos, uno de entrenamiento y otro para prueba, quedando de la siguiente forma:

1. Frases 1, 2 y 3 de las sesiones 1 y 2 (aprox. 40 segundos de señal).
2. Frase 4 de la sesión 3 (aprox. 60 segundos de señal).

La evaluación del corpus se realizó sólo con las señales de los locutores hombres, con la finalidad de poder comparar los resultados con otros corpora disponibles, los cuales sólo contienen señales de locutores masculinos.

El resultado de las pruebas se muestra en la Tabla 3. El mejor reconocimiento se obtuvo con las grabaciones de micrófono, tal y como se esperaba ya que este tipo de datos son los que mayor calidad presentan en términos del medio de adquisición, en segundo lugar los de VoIP los cuales también fueron adquiridos por medio del mismo micrófono. Por otro lado, el menor reconocimiento corresponde a los datos adquiridos con un aparato telefónico, como era de esperarse, ya que este tipo de medio es el que tiene menor calidad de captura.

Tabla 3. Resultados de la evaluación del corpus.

	Micrófono	Teléfono	VoIP
% Reconocimiento	95.24	71.43	90.48
% Identificación errónea	0.00	0.00	0.00
% Falsos Positivos	0.00	28.57	9.42
% Falsos Negativos	4.76	0.00	0.00
TOTAL	100.00	100.00	100.00

5.2. Comparación con AHUMADA

Además de la evaluación anteriormente presentada, se realizó otra con el mismo sistema ASR pero utilizando los datos del corpus en español ibérico AHUMADA para micrófono y teléfono ya que este corpus no contiene señales de VoIP. Las particiones del conjunto de datos para AHUMADA se eligieron de manera semejante a las utilizadas con nuestro corpus:

1. Diez frases fonéticamente equilibradas grabadas en dos sesiones distintas (Aprox. 40s de señal).
2. Lectura de un texto de una tercera sesión, un texto distinto al usado en el entrenamiento (Aprox. 60s de señal). De la misma manera que con nuestro corpus, sólo se utilizaron señales de locutores masculinos. La comparación de resultados se presenta en la Tabla 4. Se observa que para el caso de la señal de micrófono, el reconocimiento de los datos de AHUMADA fue ligeramente mayor, mientras que nuestra señal telefónica obtuvo un mejor reconocimiento.

Una posible explicación para la gran diferencia en el reconocimiento de la señal telefónica es que las condiciones de grabación en nuestro caso fueron más cuidadas

que lo reportado por [14] y otra posibilidad radica en la diferencia de calidad del aparato telefónico.

Tabla 4. Comparación de resultados de reconocimiento usando el corpus AHUMADA.

	Micrófono	Micrófono AHUMADA	Teléfono	Teléfono AHUMADA
% Reconocimiento	95.24	100.00	71.43	47.62
% Identificación errónea	0.00	0.00	0.00	4.76
% Falsos Positivos	0.00	0.00	28.57	33.33
% Falsos Negativos	4.76	0.00	0.00	14.29
TOTAL	100.00	100.00	100.00	100.00

6 Conclusiones

En el presente trabajo se construye un corpus de voz en español mexicano orientado al reconocimiento de locutor, es importante contar con una base de señales de este tipo ya que algunas tareas de reconocimiento, especialmente la clasificación, requieren de señales de voz que contengan características fonéticas específicas de los habitantes de cierta ubicación geográfica. También se muestra que es posible obtener la distribución fonética del español mexicano analizando texto de páginas web de periódicos locales de varias regiones de México. Se comprobó que la distribución fonética no varía de región en región. Por otro lado, se establece la importancia de que las frases que se vayan a grabar tengan cierto grado de coherencia, aún y cuando expresen ideas incompletas, ya que los locutores tienden a equivocarse menos cuando las frases a pronunciar son menos complejas.

El corpus construido permitirá continuar con investigaciones en reconocimiento y clasificación de locutores de la zona noroeste de México.

Referencias

1. Auckenthaler, R., Parris, E. S., Carey, M. J.: Improving a GMM speaker verification system by phonetic weighting. In Proceedings of the Acoustics, Speech, and Signal Processing, 1999. Volume 01 March 15 - 19, (1999).
2. Campbell, J.P. Jr.: Testing with The YOHO CD-ROM voice verification corpus. In Proceedings of the international conference on acoustics, speech and signal processing, ICASSP'95, Detroit, 341-344 (1995)
3. Campbell, J.P. Jr., Reynolds, D.A.: Corpora for the evaluation of speaker recognition systems. In Proceedings of international conference on acoustics, speech and signal processing, ICASSP'99. 2, 829-832 (1999).
4. Casacuberta, F., García, R., Llisterri, J., Nadeu, C., Pardo, J.M., Rubio, A.: Desarrollo de Corpus para Investigación en Tecnologías del Habla (ALBAYZIN). Procesamiento del Lenguaje Natural. 12. 35-42 (1992)
5. Faúndez-Zanuy, M.: State-of-the-art in speaker recognition. IEEE A&E Systems Magazine, 20(5). 7-12 (2005)

6. Faltthauser, R., Ruske, G.: Improving Speaker Recognition Performance Using Phonetically Structured Gaussian Mixture Models. In EUROSPEECH-2001, 751-754 (2001)
7. Fauve, B., Matrouf, D., Scheffer, N., Bonastre, J., Mason, J.: State-of-the-Art Performance in Text-Independent Speaker Verification Through Open-Source Software. IEEE Transactions on Audio, Speech, and Language Processing. 15 (7). 1960-1968 (2007)
8. Hennebert, J., Melin, H., Petrovska, D., Genoud, D.: Polycost: A Telephone-Speech Database For Speaker Recognition. Speech Communication. 31 (2-3), 265-270 (2000)
9. Juang, B.H., Chen, T.: The past, present, and future of speech processing. Signal Processing Magazine, IEEE. 15(3), 24-48 (1998)
10. Keshet, J., Bengio, S.: Automatic Speech and Speaker Recognition: Large Margin and Kernel Methods. Wiley. (2009)
11. Kirschning, I.: TLATOA: Developing Speech Technology & Applications for Mexican Spanish. In 2nd. Intl.Workshop on Spanish Language Processing and Language Technologies. September 14-15, 2001, Jaén, Spain. (2001)
12. Li, Q., Reynolds, D.A.: Corpora for the evaluation of speaker recognition systems. In Proceedings of the Acoustics, Speech, and Signal Processing, ICASSP'99, 829-832. March 15-19, (1999)
13. Messer, K., Matas, J., Kittler J., Jonsson, K.: XM2VTSDB: The Extended M2VTS Database. In Proceedings of the 2nd International Conference on Audio and Video Based Biometric Person Authentication (AVBPA'99). March 22-24, 1999. Washington DC, USA. (1999)
14. Ortega-García, J., González-Rodríguez, J., Marrero, V., Díaz-Gómez, J., García-Jiménez, R., Lucena-Molina, J., Sánchez-Molero, J.: AHUMADA: A Large Speech Corpus in Spanish for Speaker Identification and Verification. Speech Communication. 3. 255-264. (2000)
15. Pérez, H. E.: Frecuencia de fonemas. Revista Electrónica de la Red Temática en Tecnologías del Habla, 1, Marzo 2003. http://lorien.die.upm.es/~lapiz/erthabla/numeros/N1/N1_A4.pdf (2003)
16. Patil, H. A., Basu, T.K.: Development of speech corpora for speaker recognition research and evaluation in Indian languages. International Journal of Speech Technology. 11(1). 17-32 (2008)
17. Pineda ,A. L., Castellanos, Cuétara, J., Galescu, L., Juárez, J., Llisterri, J., Pérez, P., Villaseñor, L.: The Corpus DIMEx100: Transcription and Evaluation. Language Resources and Evaluation (DOI: 10.1007/s10579-009-9109-9) . (2009)
18. Przybocki, M. A., Martin, A. F.: NIST Speaker Recognition Evaluation Chronicles. In Proceedings of The Speaker and Language Recognition Workshop Odyssey 2004. <http://www.nist.gov/speech/publications/papersrc/ody2004NIST-v1.pdf>
19. Reynolds, D.A., Rose, R.C.: Robust text-independent speaker identification using Gaussian mixture speaker model. IEEE Transactions on Speech and Audio Processing. 3 (1). 72-83 (1995)
20. Villaseñor-Pineda, L., Montes-y-Gómez, M., Vaufreydaz, D., Serignat J.: Elaboración de un Corpus Balanceado para el Cálculo de Modelos Acústicos usando la Web. XII Congreso Internacional de Computación CIC-2003. ISBN:970-36-0098-0. pp 198-200. Mexico City, México. (2003)
21. Zamalloa, M., Bordel, G., Rodríguez, L.J., Peñagarikano, M., Uribe, J.P.: Selección y Pesado de Parámetros Acústicos Mediante Algoritmos Genéticos para el Reconocimiento del Locutor. In Proceedings of IV Jornadas en Tecnologías del Habla. 349-354. November 8-10, 2006. Zaragoza, Spain. (2006) Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195-197 (1981)