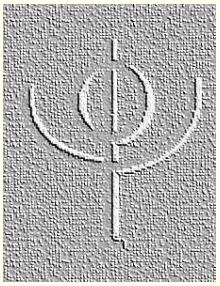
ISSN: 0325-2043



Laboratorio de Investigaciones Sensoriales (LIS)

Informe XLIV–2011

I N I G E M

CONICET

U B A

Instituto de Inmunología, Genética y Metabolismo
Córdoba 2351, Piso 9, (1121), Buenos Aires
Tel/Fax: 5950-9024
lis@fmed.uba.ar — http://www.lis.secyt.gov.ar

Índice

1. Introducción

Desde su creación en al año 1968, el LIS publica un informe anual en donde se consignan las publicaciones realizadas, los trabajos en curso, la actividad docente y el intercambio científico.

Los Informes LIS están registrados bajo ISSN 0325-2043 (International Standard Serial Number), a través de Latindex[1], reconocido internacionalmente para la identificación de las publicaciones seriadas. La serie comienza con el Informe I-1968, Laboratorio de Investigaciones Sensoriales, CONICET.

En los informes aparecen siglas que referencian las sedes del LIS, primero en el Hospital Escuela (HE), luego en la Facultad de Medicina (FM) y, actualmente, en el Hospital de Clínicas (HC) de la Universidad de Buenos Aires.

Desde el año 1997, los informes también están disponibles a través del sitio web del laboratorio: http://www.lis.secyt.gov.ar/.

El 14 de septiembre de 2011, el LIS y otros laboratorios del Hospital de Clínicas-UBA constituyeron el *Instituto de Inmunología, Genética y Metabolismo* (INIGEM), dependiente del CONICET y de la Universidad de Buenos Aires.

2. Personal

Investigadores

- GUIRAO Miguelina, Prof. Filosofía, Dra. en Psicología Experimental.
- GURLEKIAN Jorge A., Ing. Electrónico. Responsable del LIS.
- TORRES Humberto, BioIngeniero, Dr. en Ingeniería.

Investigadores que participan en proyectos que se desarrollan en el LIS:

- CALVIÑO Amalia M., Farmacéutica, Dra. en Bioquímica.
- GRAVANO Agustín, Licenciado y Dr en Ciencias de la Computación.
- TOLEDO Guillermo, Lingüista, Dr. en Filosofía y Letras.
- VACCARI María Elena, Lic. en Fonoaudiología.

Becarios

- EVIN Diego, Bioingeniero, Dr. en Ciencias de la Computación. Becario Posdoctoral CONICET
- COSSIO MERCADO Christian, Ing. en Informática, Becario FONCYT. Tesista de Doctorado UBA
- MARTINEZ SOLER Miguel, Ing. en Informática, Becario FONCYT. Tesista de Doctorado, UBA

[1] Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal. Sitio: http://www.latindex.unam.mx

Tesistas

- CECCONELLO Luis, Fonoaudiólogo. Tesista de Doctorado UMSA
- ELISEI Natalia, Fonoaudiologa. Tesista de Doctorado UBA
- TRIPODI Mónica, Lingüista. Tesista de Doctorado UBA.
- UNIVASO Pedro, Ing. Electrónico. Tesista de Doctorado UBA.

3. Proyectos de Investigación

3.1. Proyecto Mincyt-BMBF: Extracción y Modelación de los Parámetros Prosódicos para el Análisis, Síntesis y Reconocimiento del habla

Nombre en aleḿan: *Prosodische Parameterextraktion und Modellierung für die Sprachanalyse, -synthese und -erkennung*

Directores: Jorge A. Gurlekian y Hansjörg Mixdorff. Período: 2009-2011.
Unidad de Ejecución: Laboratorio de Investigaciones Sensoriales y Department of Computer Sciences and Media.
Institución de la que depende la Unidad de Ejecución: CONICET y Technische Fachhochschule Berlin (TFH)
Este proyecto se integra con el proyecto Nombre: PAE Nro: 37122, PID 2007. Nro. 094. FONCYT. Desarrollo de un sistema de conversión de Texto a Habla. Director: Jorge A. Gurlekian. Período: 2009-2011. Unidad de Ejecución: Laboratorio de Investigaciones Sensoriales.

3.2. CONICET PIP Nro. 5897/06: Análisis de las sensaciones de dulce, agrio y amargo en soluciones puras y mezcladas en medio acuoso y alcohólico

Dirección: Miguelina Guirao
Codirección: Amalia Mirta Calviño

Efecto del etanol en el gusto con y sin atributos trigeminales

Hasta el momento los estudios sobre la influencia del etanol en el sabor se habían dirigido a bebidas alcohólicas en las que se mezclan diferentes sustancias gustativas. En cambio no se tenía un conocimiento acabado acerca de la interacción etanol-gusto en sustancias puras. Los pocos intentos que se han realizado discrepan en los resultados. Algunos autores han encontrado que el efecto del etanol no es significativo y otros que los resultados dependen del método. En nuestro caso hemos investigado los cambios que se producen en el gusto cuando se le agrega etanol a tres sustancias puras: el dulce de la sacarosa, el agrio del ácido cítrico y el amargo de la cafeína. Para ese fin hemos experimentado con dos dimensiones de la sensación la intensidad del gusto y la duración o persistencia del gusto en la cavidad oral. Además para descartar una posible influencia del procedimiento elegido aplicamos tres métodos psicofísicos diferentes.

En general los resultados revelan que el efecto del etanol depende en gran medida de la concentración de la sustancia y de la graduación del alcohol. Sobre la base de los datos obtenidos se comparara el efecto que tiene el etanol en cada uno de los tres gustos y la posible influencia de los atributos trigeminales en la modificacion del gusto.

4. Proyectos de I+D

4.1. PID 094/2007 - Desarrollo de un Sistema de Conversión de Texto a Habla

PAE Nro: 37122, PID 2007. Nro. 094.FONCYT. Desarrollo de un sistema de conversión de Texto a Habla
Director: Jorge A. Gurlekian
Período: 2009-2011
Unidad de Ejecución: Laboratorio de Investigaciones Sensoriales
Institución de la que depende la Unidad de Ejecución: CONICET
Entidad Acreditadora y/o Financiadora: FONCYT
Financiamiento obtenido: 258.200 pesos. Costo total 780.200 pesos

4.2. PID 35891 - Desarrollo de las técnicas de reconocimiento del hablante para su aplicación a nivel forense

Secretaria de Ciencia y Técnica. Proyectos de Investigación y Desarrollo PID
Entidad adoptante: Policía Científica, Gendarmería Nacional Argentina.
Director: Jorge A. Gurlekian

4.2.1. Trabajos terminados

Aplicaciones vinculadas al estudio forense de la voz. Diego A. Evin y Jorge A. Gurlekian Dentro del grupo de aplicaciones referidas al estudio forense de la voz se desarrolló un segmentador de archivos de audio, un programa para el filtrado acústico, un editor de etiquetas léxicas, y otro programa para la comparación biométrica de registros de voz.

5. Asesorías Tecnológicas

5.1. Asesoria Técnica: Prevencion ART, Grupo SANCOR SEGUROS

Asesor Responsable: Jorge A. Gurlekian
Estado: Terminado

Medición y Caracterización de Ruidos en Ambientes Laborales. Desarrollo de un sistema de análisis de ruido en bandas, análisis de dosis y análisis de ruidos impulsivos

Diego A. Evin y Jorge A. Gurlekian

Resumen

El sistema para el análisis de ruidos laborales desarrollado está conformado por un componente hardware y otro software. El componente hardware consiste de un decibelímetro estándar de bajo costo con salida de audio, y un grabador digital de alta fidelidad. El componente software efectúa las mediciones en bandas de frecuencia, obtiene el nivel de dosis a partir de los cálculos de los niveles sonoros continuos equivalentes —con y sin protectores auditivos—, y realiza el cálculo de los niveles de presión sonora máxima, mínima y el nivel pico en ruidos de impacto e impulsivos. La secuencia de estos componentes ha sido mencionada en la literatura por distintos autores. La empresa que ha solicitado la consultoría es el área de Salud Ocupacional de Prevención ART Sancor Seguros, que ha realizado la grabación de ruidos en fábricas y provisto las normas requeridas para su empleo institucional. El diseño de los datos de entrada y del reporte de salida son exclusivos para esta empresa y no serán presentados en este informe. La comparación de costos de un decibelímetro equivalente al conjunto desarrollado es de 3 a 1 a favor de este híbrido.

6. Docencia

6.1. Cursos de posgrado

Docente: Miguelina Guirao
Para la Carrera de Especialistas en ORL Facultad de Medicina UBA.
Lugar: Asocacioń Médica Argentina, Buenos Aires, Argentina.

- Tema: Fisiologia del Olfato (22 de marzo de 2011)
- Tema: Mecanismos Neurofisiológicos del Sistema Olfatorio (6 de octubre de 2011)

6.2. Seminarios en el laboratorio

- Martes 20 de septiembre: Jorge Gurlekian. “Acentos tonales primarios y no primarios. Análisis del Corpus Amper”.
- Martes 27 de septiembre: Agustín Gravano. “Mimetización de rasgos prosódicos entre interlocutores”
- Martes 15 de noviembre: Diego Evin. “Deep Machine Learning y procesamiento del habla”.

7. Tesis

7.1. Doctorales

Incorporación de Información Suprasegmental en el Proceso de Reconocimiento Automático del Habla

Dr. Diego Alexis Evin

El dia 10 de Mayo 2011 el Dr. D.A. Evin defendió su tesis doctoral sobre el tema “Incorporación de Información Suprasegmental en el Proceso de Reconocimiento Automático del

Habla") en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires.
Director de tesis: Ing. Jorge A. Gurlekian
Consejera de Estudios: Dra. A. Ruedin
Disponible en la biblioteca digital de la FCEyN-UBA: link

Patrones vibratorios de los pliegues vocales en cantantes con diferentes niveles de calidad vocal

Dr. Luis Cecconello

El dia 8 de Noviembre de 2011 el Dr. L. Cecconello defendió su tesis doctoral sobre el tema "Patrones vibratorios de los pliegues vocales en cantantes con diferentes niveles de calidad vocal" en la Universidad del Museo Social Argentino.
Director de tesis: Ing. Jorge A. Gurlekian

Evaluación acústica y perceptual de la voz para la detección y caracterización de los desórdenes vocales

Dra. Natalia G. Elisei

El dia 14 de Diciembre de 2011 la Dra. N.G. Elisei defendió su tesis doctoral sobre el tema "Evaluación acústica y perceptual de la voz para la detección y caracterización de los desórdenes vocales" en la Facultad de Medicina de la Universidad de Buenos Aires.
Director: Ing. Jorge Alberto Gurlekian
Co-Director: Dr. Alberto Chinski
Co-Directora: Dra. Ana Maria Borzone

7.2. Doctorales en curso

Desarrollo de Pruebas de Evaluación de la Inteligibilidad del Habla

Tesista: Ing. Jorge A. Gurlekian
Director: Prof. Dr. Alberto Chinski (FMED-UBA)
Universidad de Buenos Aires, Facultad de Medicina
Resumen:

Este trabajo consta del desarrollo de una prueba de evaluación de la inteligibilidad en condiciones de ruido altamente interferente. Se trata de una nueva prueba rapida y de fácil aplicación en diversos ámbitos donde concurren escolares. Tambien puede utilizarse para evaluar el deterioro de la percepción y de la producción del habla.

Reconocimiento automático de hablantes empleando información de largo plazo

Tesista: Ing. Pedro Univaso
Director: Ing. Jorge A. Gurlekian
Universidad de Buenos Aires, Facultad de Ingeniería.

Diagnostico diferencial de pacientes con movimientos anormales laringeos, complementacion entre el diagnostico neurológico y los resultados que brinda el abordaje otorrino-fonoaudiologico

Tesista: Liliana Sigal
Universidad de Buenos Aires, Facultad de Medicina
Director : Ing. Jorge A. Gurlekian.
Consejero : Profesor Dr. Federico Micheli

El presente trabajo se desarrollará en el Laboratorio de Investigaciones Sensoriales, Conicet, la División Otorrinolaringología, el Departamento de Movimientos Anormales y el Departamento de Neurofisiología del Hospital de Clínicas José de San Martín. Tiene por objeto el estudio de la distonía laríngea, alteración que afecta el movimiento de las cuerdas vocales. Los síntomas son generalmente graduables desde una inestabilidad moderada a cortes incontrolables en la voz y un creciente esfuerzo que repercute en la inteligibilidad del habla.

8. Actividades de Divulgación

Los Sentidos

Divulgadores: Dra. Amalia Calviño, Dra. Miguelina Guirao
Lugar: Programa "Dosis de Radio", como parte el programa de la Facultad de Farmacia y Bioquímica, Radio UBA
Fecha: 9 de mayo de 2011

9. Publicaciones

9.1. Revistas

- TOLEDO, G. Y GURLEKIAN, J.A: Amper-Argentina: relaciones entre los acentos tonales y los acentos primarios y no primarios. Revista Internacional de Lingüística Iberoamericana (RILI), vol IX, nro. 1 (17), pp.101–110, 2011.

9.2. Congresos

- EVIN, D., GURLEKIAN, J.A., TORRES, H.M.: N-Best Rescoring Based on Intonation Prediction for a Spanish ASR System. In Proc. of the 21 Konferenz Elektronische Sprachsignalverarbeitung (ESSV 2010), pp. 234-242, TUD Press Verlag der Wissenschaften, Berlin, Germany, September 8-10, 2010.

9.3. Informes Técnicos

- MARTINEZ-SOLER, M.: Sistema de Identificación de Hablantes Basado en Estadísticas Sobre Formantes, 2011.
- TORRES, H.M.: Etiquetado de clase de palabras. Laboratorio de Investigaciones Sensoriales, Neurociencias, Hospital de Clínicas, UBA. Febrero de 2011.

- TORRES, H.M.: Diseño de un sistema de TTS. Esquema general. Definición de estructuras, aplicaciones y rutinas. Laboratorio de Investigaciones Sensoriales, Neurociencias, Hospital de Clínicas, UBA. Marzo de 2011.
- TORRES, H.M.: Informe técnico Convenio: CONICET-MITROL - Período 1/1/2010 al 31/3/2011. En el marco del convenio de entre CONICET MITROL, Buenos Aires. Marzo de 2011.

Apéndice

A. Resúmenes de Trabajos

A.1. Índice de perturbación, de precisión vocal y de grado de aprovechamiento de energía para la evaluación del riesgo vocal. *Gurlekian, J.A. y Molina, N.*

Jorge A. Gurlekian, Nancy Molina

Resumen

Este trabajo presenta la aplicación de un método adecuado para el análisis del riesgo vocal debido a las alteraciones de voz. Se obtienen tres índices: 1. Un índice de perturbación que agrupa cuatro parámetros clásicos como el jitter, shimmer, la relación armónico ruido y la amplitud del cepstrum. 2. Un índice de precisión vocal vinculado con la estabilidad articulatoria y medido como la inversa de la desviación estándar de los primeros cinco formantes y 3. Un índice asociado al grado de aprovechamiento de energía, que evalúa tanto la coincidencia entre los armónicos con los formantes como las pérdidas de energía que se producen en el tracto vocal, medidas como la inversa de los anchos de banda. Para esta presentación, los índices mencionados se evalúan en 84 voces de docentes con distintos grados de alteración de voz, durante la emisión de la vocal /a/. El índice de perturbación se calcula a partir de las contribuciones parciales sobre una diagonal que va desde valores normales en un extremo hasta valores patológicos en el otro. El índice de precisión vocal se presenta con un grafico de las áreas de formantes normalizadas respecto de la frecuencia fundamental. El índice de aprovechamiento de energía grafica la inversa de los anchos de banda a lo largo de un continuo. La agrupación de las voces de docentes en normales, con riesgo vocal, y alteradas, se presentan en relación a los respectivos diagnósticos laringológicos verificando su utilidad en la evaluación masiva de los profesionales con riesgo vocal.

Palabras clave

Voz; Riesgo vocal; Análisis objetivo; Prevención profesional.

Perturbation index, precision and extra energy gain indexes for the evaluation of vocal risk

Abstract

This paper presents the application of an acoustic analysis method adequate for vocal risk evaluation. Three indexes are calculated: 1. A perturbation index, associated with classical perturbation parameters, such as jitter, shimmer, harmonic to noise relation, and cepstrum amplitude. 2. A precision index related to articulatory stability which is measured by the inverse of the standard deviation of the first five formants, and 3. An extra energy gain index due to both optimal harmonic/formant alignment and losses at the vocal tract which is measured by the inverse of the first five formant bandwidths. For this presentation the voice risk condition of 84 school teachers was evaluated on the basis of different levels of voice quality during the emission of vowel /a/. Graphic representations of the three indexes are available

to have a quick visual feedback. Perturbation index is calculated looking at the contribution scores defined along a diagonal line that goes from normal to altered measurements. Exactly at the diagonal center, normal thresholds are represented for the four perturbation parameters. A formant plot -normalized to fundamental frequency- is presented to see both formant one and formant two contributions to the precision index. Formant bandwidths are measured and their inverse is drawn as points along a non-contribution/contribution line to verify energy management at the vocal tract. Teacher´s voices resulted automatically classified in three groups: normal, risky and altered all of which were successfully compared with their laryngologist diagnosis. The method employed is a promissory application for massive vocal risk evaluation.

Keywords

Voice; Vocal risk; Acoustic analysis; Professional prevention.

A.2. Evaluación de la intensidad y la duración del gusto agrio en la mezcla de ácido cítrico con etanol (HC-Q 177). *Guirao, M. et al.*

Guirao, M., Greco Drianó, E., Calviño, A., y Evin, D.

Resumen

En el presente trabajo se investiga la influencia que las cualidades del etanol (EtOH) tienen en el gusto agrio del ácido cítrico. En el etanol se reconocen varias propiedades trigeminales como irritación, temperatura, (EtOH) anestesia (numbing) y pungencia y también sensoriales como gusto y olor. Esas cualidades pueden cambiar con la graduación del alcohol. A niveles cercanos al umbral se perciben tonos dulces y amargos y a concentraciones más altas las trigeminales. Entre estas la sensación de anestesia parece ser la mas saliente y la de mayor persistencia. A su vez el ácido cítrico a concentraciones altas presenta también propiedades trigeminales como la irritación oral e intranasal y puede también aparecer el gusto amargo. En general son pocos los autores que han estudiado este problema con métodos psicofísicos y no ha habido acuerdo respecto a las conclusiones. Algunos autores han encontrado que el efecto del (EtOH) en los gustos no es significativo y otros que los resultados dependen del método. En este estudio se realizaron tres series de experimentos y en cada una se aplicó un método psicofísico diferente. En todas las sesiones participaron diez panelistas que evaluaron la intensidad y persistencia de la sensación de agrio. Las soluciones de ácido cítrico se presentaron en forma pura, diluidas en medio acuoso, y con dos graduaciones 8 y 15 % de etanol. Las evaluaciones se efectuaron combinando las muestras con y sin etanol, en forma aleatoria, en una misma sesión experimental. En general los procedimientos experimentales fueron similares a los que se describieron antes para estudiar el efecto del etanol en la sacarosa. En la primera serie se determinaron las variaciones en la intensidad del agrio con el método de Estimación de la magnitud (CL). A este fin se evaluaron siete concentraciones de ácido cítrico 3, 5, 10, 15, 30, 45, y 70 mM sin y con etanol En la segunda se aplicó el método de Comparación por Pares en grupos de a tres. Los panelistas ordenaron, compararon y estimaron la intensidad de tres soluciones de ácido cítrico 3, 15 y 70 mM solas y mezcladas y a modo de control se agregó una solución de agua destilada. En la tercera serie se aplicó el Método de Registro de las Curvas Intensidad-Tiempo (IT). Los panelistas usaron una técnica computarizada para registrar la intensidad y la duración (o persistencia) de las respuestas a tres soluciones 5, 15

y 45mM puras y mezcladas. Se hicieron tres registros por cada estimulo. Se obtuvieron tres series de curvas en las que se midieron la intensidad máxima (I max.), tiempo total de la sensación (T tot.) y el área bajo la curva (AUC). A pesar de la diferencia en los métodos, los dos primeros con estimaciones numéricas y el tercero con respuestas motoras, los resultados fueron coincidentes

B. Resúmenes de Tesis

B.1. Incorporación de Información Suprasegmental en el Proceso de Reconocimiento Automático del Habla. *Evin, D.A.*

Diego A. Evin

Resumen

Desarrollar sistemas informáticos capaces de interactuar con sus usuarios de la forma más natural y eficiente posible es uno de los requisitos esenciales para lograr la integración del mundo tecnológico en la sociedad. En ese marco el habla se presenta como una de las formas de comunicación más eficientes y naturales que posee el ser humano. Es por ello que desde el origen mismo de la investigación en ciencias de la computación, el desarrollo de interfaces hombre-máquina a través de la voz ha despertado un gran interés. Uno de los elementos que componen dicha interfaz oral es el Reconocimiento Automático del Habla (RAH), área de la Inteligencia Artificial que busca desarrollar sistemas computacionales capaces de transformar un fragmento de habla en su trascripción textual. El RAH es un problema de gran complejidad, lo que se puede atribuir principalmente a dos factores: en primer lugar a la variabilidad de la señal de habla, que responde a múltiples factores como características particulares del locutor y medio acústico donde se registra, la velocidad y estilos de elocución; y en segundo lugar a la necesidad de encontrar palabras individuales en un continuo acústico, es decir realizar al mismo tiempo las tareas de segmentación y clasificación. Si bien se pueden encontrar en los últimos años avances significativos en el desempeño de los sistemas de RAH, aún hay mucho por mejorar en relación a la capacidad de reconocimiento que presentan los oyentes humanos para las mismas tareas y bajo las mismas condiciones. Varias hipótesis intentan explicar esta diferencia de desempeño: información insuficiente o representada de manera inadecuada en los sistemas automáticos, problemas en el modelado del sistema de reconocimiento, insuficientes cantidades de ejemplos empleados para lograr tasas de reconocimiento similares, etc. Con respecto al primero de estos puntos, los sistemas de RAH no utilizan toda la información acústica disponible en la señal de habla. Dichos sistemas interpretan el habla como secuencias de unidades cuyas duraciones se encuentran a nivel segmental (fonético). Por lo tanto procesan la información acústica en la escala segmental para obtener las hipótesis de secuencias de unidades emitidas. Sin embargo estudios tanto psicoacústicos como psicolingüísticos resaltan el rol crucial que posee la información de una escala temporal mayor: la información suprasegmental, en la percepción humana. Se entiende por información suprasegmental toda aquella que está dada en segmentos de duración superior al fonético, y cuyas propiedades están determinadas principalmente por la prosodia de una frase. Además se argumenta que en la tarea de reconocimiento e interpretación del habla los seres humanos emplean e integran varios niveles de conocimiento lingüístico, muchos de los cuales aún no han sido incorporados o aprovechados eficientemente en el RAH. A partir de esas evidencias resulta interesante in-

vestigar cuál es el aporte que puede brindar la información suprasegmental o prosódica para mejorar el desempeño de los sistemas de RAH estándar. En esta Tesis se investiga el empleo de información suprasegmental como factor de mejora en el desempeño, así como alternativas para su integración en sistemas de RAH estándar.

Palabras Clave:

Prosodia; Entonación; Acentuación; Modelos Ocultos de Markov; Reconocimiento Automático del Habla.

Incorporation of Suprasegmental Information into Automatic Speech Recognition Process

Abstract

The development of computational systems capable of interacting with users in the most natural and efficient way is one of the essential requirements for the integration of the technological world in society. In this context speech is presented as one of the most efficient form of communication mechanisms available for human beings. That is why from the very beginning of research in computer science, the development of human-machine interfaces through voice have gain great interest. One of the elements that compose such interfaces is the Automatic Speech Recognition (ASR). ASR is a field of Artificial Intelligence which searches for the development of computational systems that transform speech segments into text transcriptions. ASR is a very complex problem, which can be attributed mainly to two factors: first, to the huge variability of the speech signal, depending on multiple factors such as the speaker, the acoustic environment, linguistic context, speech rate, emotional states, locution styles, and many others; and secondly to the need of finding isolated words in an acoustic continuum, that is to say solving segmentation and classification problems simultaneously. Even thought we can find significant advances in the performance of ASR systems in recent years, there is still much space for improvement to match human recognition ability for the same tasks under the same conditions. Several hypotheses attempt to explain these differences on performance: insufficient information, inadequate way to represent it, problems in modelling, insufficient quantities of used examples to achieve similar recognition rates, etc. Regarding the first point, ASR systems do not use all available acoustic information in speech signal. These systems interpret the speech as sequences of units whose durations spans in a segmental (phonetic) level. Therefore they process the acoustic information at a segmental scale to obtain the hypotheses of sequences of uttered units. Nevertheless psychoacoustic and psycholinguistic research emphasize the essential role of information at a higher temporal level for the human speech perception: the suprasegmental information. Any information whose duration spans over several phonetic units can be thought as suprasegmental, and its properties are determined principally by the prosody of an utterance. Furthermore, it is argued that during the task of speech recognition and interpretation, various linguistic knowledge are integrated and used. It has been also argued that no much of linguistic knowledge have yet been incorporated or utilized efficiently in the ASR From these evidences it seems relevant to investigate whether the suprasegmental or prosodic information could contribute to improve the performance of standard ASR systems. In this thesis the use of the suprasegmental information is investigated as a factor for improving performance, as well as an alternative for the integration of this information into the architecture of standard ASR systems.

Keywords

Prosody; Intonation; Stress Patterns; Hidden Markov Models; Automatic Speech Recognition.

B.2. Patrones vibratorios de los pliegues vocales en cantantes con diferentes niveles de calidad vocal. *Cecconello, L.*

Luis Cecconello

Resumen

El objetivo del presente trabajo es determinar si existen diferencias entre los patrones vibratorios de los pliegues vocales en cantantes con diferentes niveles de calidad vocal. Para ello se estudiaron las voces de 150 cantantes de 18 a 50 años, siendo 81 de sexo femenino y 69 masculino. Fueron utilizados 3 métodos para la valoración de los patrones vibratorios acústicos y mecánicos: Video-estroboscopia laríngea, Electroglotografía y Análisis acústico. Mediante estos métodos fueron valorados 23 parámetros cualitativos y 42 cuantitativos. Se utilizaron 3 programas de análisis: Dr. Speech 4 módulo Vocal Assessment, Anagraf y VoceVista. La calidad vocal de los cantantes fue categorizada en 5 tipos (mala, regular, buena, muy buena y excelente); un panel conformado por 10 fonoaudiólogos y/o profesores de canto con más de 5 años de experiencia en el área vocal cantada evaluó a cada uno de los cantantes durante la emisión de la vocal /e/ (calidad vocal propiamente dicha) y en un fragmento cantado (calidad de la canción). Una serie de parámetros presentaron diferencias significativas en las diferentes calidades vocales determinadas por la vocal /e/ y por la canción. El análisis de las calidades vocales de la canción presentó menor cantidad de parámetros capaces de diferenciarlas. Esto se debe a que durante la canción, los cantantes suelen producir una serie de adaptaciones, por lo cual la calidad vocal mejora, sumado al factor de que el análisis del nivel de calidad vocal de la canción, se basa en ciertos aspectos, entre ellos, el estilo musical. En cambio, este último, no se incluye en la vocal. Los resultados aquí obtenidos demuestran que diferentes calidades vocales de cantantes pueden ser diferenciadas mediante parámetros acústicos y mecánicos.

B.3. Evaluación acústica y perceptual de la voz para la detección y caracterización de los desórdenes vocales. *Elisei, N.G.*

Natalia G. Elisei

Resumen

El presente proyecto de investigación tiene como objetivo general, obtener la información requerida para el diseño de una herramienta de evaluación objetiva y automática para la detección y caracterización del grado de los desórdenes vocales. En particular se espera realizar:

1. La creación de una base de datos que recopilará las emisiones representativas de distintas alteraciones de la voz.

2. La aplicación de las pruebas perceptuales actualmente en uso, como referencia de ponderación de las mediciones objetivas.

3. La evaluación objetiva mediante algoritmos de cálculo de las alteraciones acústicas de los atributos de la voz. Además se prevé integrar información clínica incluyendo el análisis videoestroboscópico.

Este trabajo es innovador en el sentido que pretende desarrollar una herramienta que a partir de una grabación de voz permita el prediagnóstico y estudio de los desórdenes vocales. Presenta la ventaja de ser una técnica no invasiva, de bajo costo y eficiente que no requerirá hardware y facilitará mediciones objetivas de la función vocal. Supone un notable interés de diferentes sectores: profesionales de la medicina en particular los relacionados con el estudio de la voz, docentes, cantantes, locutores, en definitiva a los profesionales donde la voz es la herramienta de trabajo. Empresas relacionadas con la prevención y diagnóstico clínico así como las aseguradoras del riesgo del trabajo asociadas a callcenters, telemarketers, etc.

C. Informes técnicos

C.1. Aplicaciones vinculadas al estudio forense de la voz. *Evin, D.A. y Gurlekian, J.A.*

Dentro del grupo de aplicaciones referidas al estudio forense de la voz se desarrolló un segmentador de archivos de audio, un programa para el filtrado acústico, un editor de etiquetas léxicas, y otro programa para la comparación biométrica de registros de voz.

Segmentador de archivos de audio

En el ámbito forense es habitual que los registros de voz disponibles para efectuar un análisis biométrico se encuentren dentro de grabaciones de considerable duración, como puede ser una escucha telefónica. Por ello es conveniente contar con un módulo que permita separar el registro original en fragmentos cuyas duraciones hagan más manejables las etapas de análisis y procesamiento subsecuentes. En este sentido se diseñó e implementó un programa que permite seleccionar un archivo de entrada determinado y obtener a la salida fragmentos del mismo en archivos de audio separados. Los formatos de entrada admitidos por el programa son ogg vorvis, mp3, flac, monkey audio, wavpack y wave, y a la salida es posible conservar el tipo de archivo original o transformarlo en formato wave, que es el más adecuado para las fases de análisis posteriores.
Con respecto al modo de funcionamiento del programa, éste efectúa una detección de regiones de silencio que resulten adecuadas para efectuar la segmentación (cuyas amplitudes se encuentren debajo de un umbral y tengan una duración mínima especificada), y considera un umbral de duración mínima y máxima de cada fragmento. El usuario debe seleccionar el umbral de amplitud para considerar un sonido como silencio, la duración mínima de un lapso de silencio, la duración mínima de cada archivo segmentado y si desea acotar la duración máxima de cada uno de ellos.

En la figura 1 muestra una captura de pantalla correspondiente a esta aplicación.

Este programa ofrece como dato de salida, el número, duración y nombres de los fragmentos obtenidos, los cuales se graban en el mismo directorio donde se encuentra el archivo de entrada. El nombre de los archivos de salida son iguales al de entrada pero agregan un sufijo que identifica el número de fragmento correspondiente. El objetivo de esta identificación automática de cada archivo fue reducir la interacción requerida con el usuario.

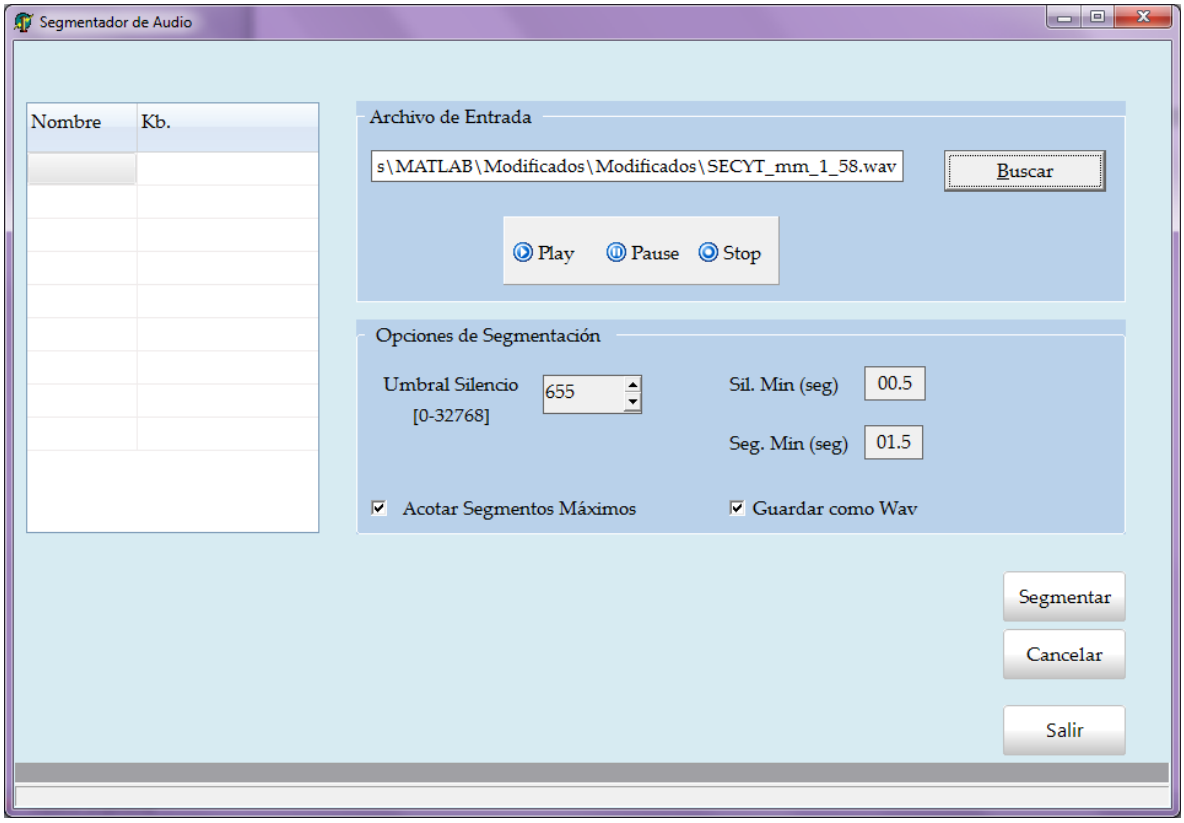


Figura 1: Aplicación para la segmentación de archivos de audio

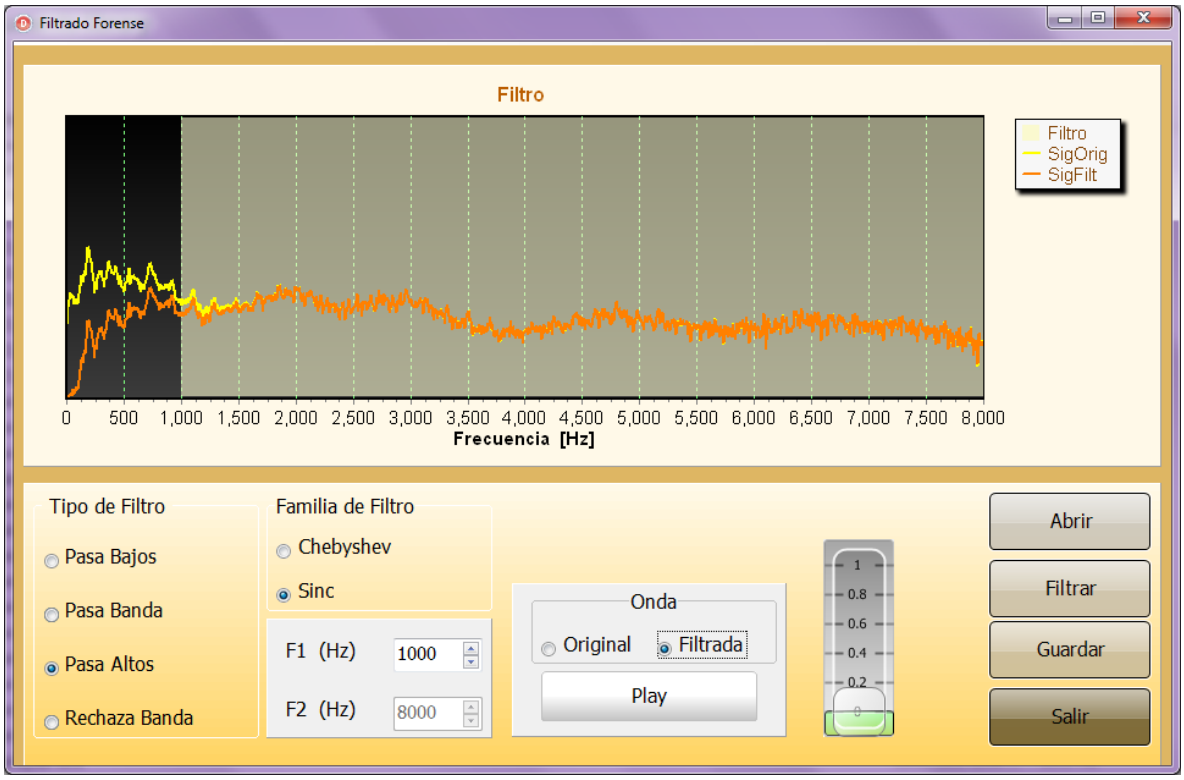


Figura 2: Aplicación para el filtrado acústico de archivos de audio

Filtrado Acústico

Una vez segmentados los archivos de audio, o antes de efectuar dicha segmentación, en algunos casos suele ser necesaria una etapa de acondicionamiento de las señales de audio. En esta etapa se intenta remover componentes de audio que faciliten el análisis o la transcripción de las mismas. La figura 2 muestra una captura de esta aplicación.

Este programa permite aplicar filtrado pasa-bajos, pasa-altos, pasa-banda o rechaza banda con intervalos de frecuencias especificados por el usuario, así como la posibilidad de emplear dos familias de filtrado diferentes: Chebyshev o Sinc. Una vez aplicado el filtrado es posible visualizar el espectro de Fourier de la señal original y la filtrada, así como reproducir tanto la señal original como filtradas. Es importante tener en cuenta que este proceso puede eliminar parte de la información acústica que podría ser útil para la caracterización de los locutores, por lo que el usuario debe evaluar el tipo de filtrado aplicado para saber si una vez transcripto el registro no es conveniente volver a trabajar sobre la señal sin filtrar.

Editor de Etiquetas Léxicas

Una vez segmentados y filtrados los archivos de audio, generalmente es necesario contar con la transcripción léxica de su contenido, por ejemplo para facilitar su análisis a nivel fonético. Esta es una tarea que demanda relativamente mucho tiempo, por lo cual se diseñó una aplicación centrada en la agilidad y sencillez de uso para los transcriptores. La figura 3 muestra una captura de esta aplicación.

El usuario debe seleccionar el directorio de trabajo, y automáticamente se muestra en la

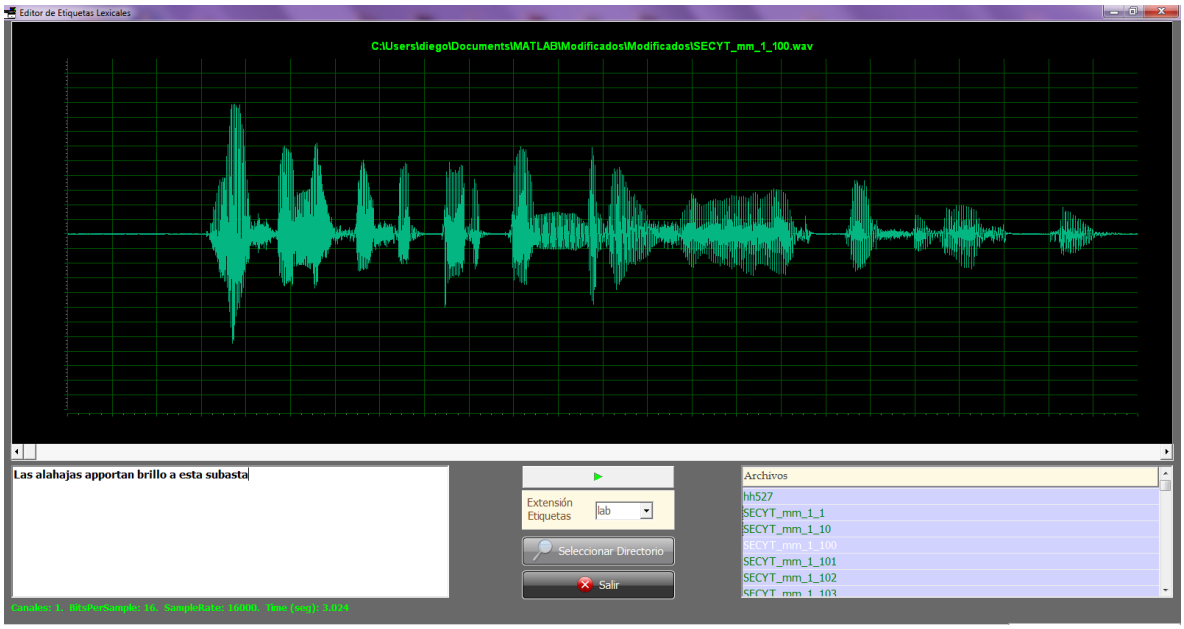


Figura 3: Aplicación para la transcripción léxica de archivos de audio

tabla de la derecha una lista con los archivos wave presentes en esa ubicación. Al hacer click sobre uno de los nombres de archivo, en el área central de la pantalla se muestra la forma de onda de la señal, y en la ventana inferior izquierda, en caso de existir el archivo de transcripción léxica correspondiente, o un cuadro en blanco para efectuar las respectivas anotaciones. El usuario puede seleccionar la extensión que tienen los archivos de transcripciones asociados. Además empleando el mouse sobre la forma de onda, puede expandir y reproducir regiones específicas de la señal, lo que resulta útil en caso que se deba repetir varias veces un fragmento que resulte difícil de transcribir. Una vez efectuada la transcripción, al seleccionar un nuevo archivo el programa consulta al usuario si desea guardar los cambios efectuados. En caso de recibir una respuesta positiva, el programa automáticamente genera un archivo de segmentación léxica inicial en formato Anagraf. En este archivo se divide la duración total del archivo por el número de palabras, y se ubica el instante de inicio y fin de cada una de acuerda a este valor. Si bien esta es una aproximación grosera a la segmentación léxica, acelera el proceso de segmentación en Anagraf, ya que el usuario solamente debe desplazar las etiquetas para obtener los archivos definitivos. Esta aplicación está pensada también para generar conjuntos de datos de audio y transcripción léxica que resulten adecuados para entrenar un reconocedor automático del habla. Una vez obtenido un número adecuado de muestras, el contar con un sistema de reconocimiento automático del habla permitirá agilizar el proceso de segmentación tanto léxica como fonética.

Programa para la Comparación Biométrica de Registros de Voz

Una vez que el usuario obtuvo la segmentación y etiquetado fonético empleando Anagraf, el proceso de análisis biométrico consiste en efectuar un análisis estadístico sobre las características que permitan identificar locutores. En particular se trabajó empleando como atributos para este análisis estadístico el valor de los primeros 4 formantes de regiones vocálicas.

La prueba de identificación consiste en determinar para una vocal determinada si las características de un locutor son compatibles con algún otro locutor del conjunto de datos disponible. Para esta determinación se efectúa una prueba de Chi cuadrado ó T de Hotelling. En las figuras 4 y 5 se muestra la captura de la aplicación desarrollada. La figura 4 muestra la interfaz para la selección de datos a estudiar, y la figura 5 el resultado que ofrece el programa.

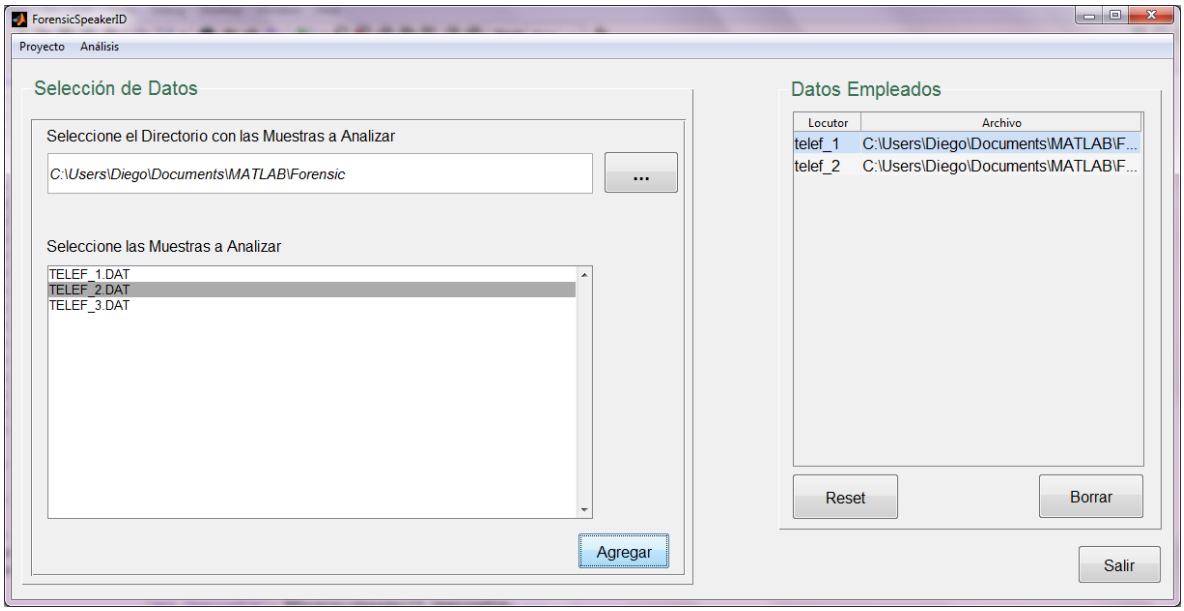


Figura 4: Ingreso de datos para el análisis biométrico de locutores

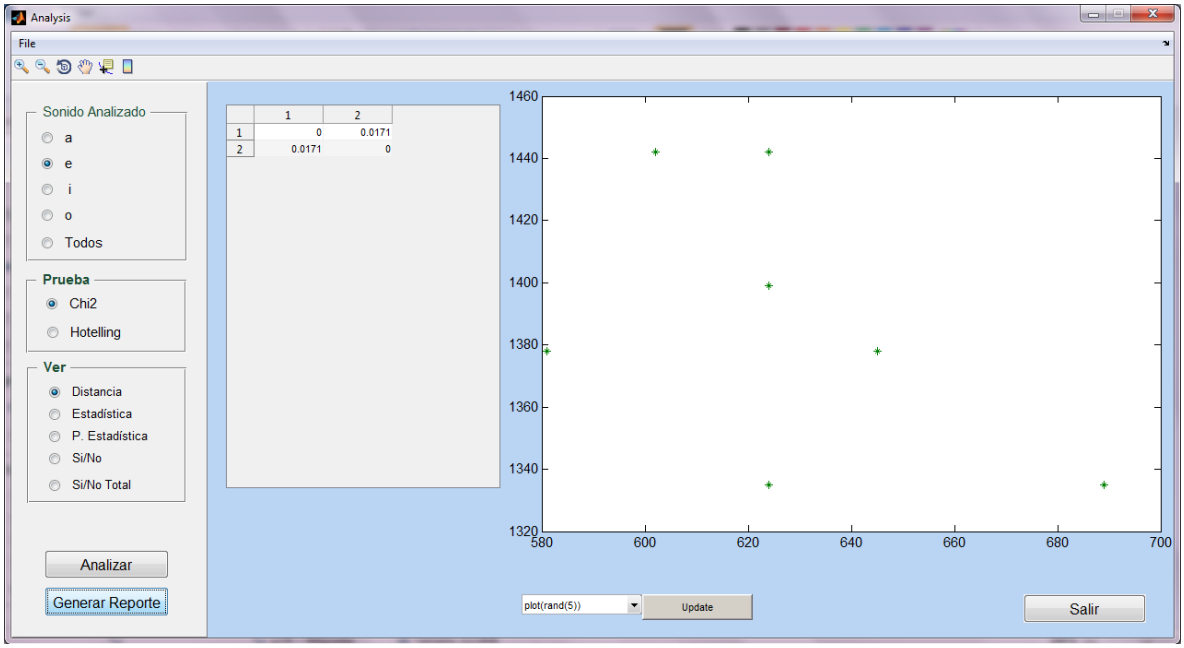


Figura 5: Cálculo estadístico para el análisis biométrico de locutores

Una vez seleccionado el directorio de trabajo, el programa carga automáticamente todos los datos disponibles para cada instancia vocálica de todos los locutores encontrados. El operador posteriormente selecciona el subconjunto de locutores que desea comparar y pasa a la ventana donde se efectúa el estudio estadístico.

C.2. Sistema de Identificación de Hablantes Basado en Estadísticas Sobre Formantes. *Martinez-Soler, M.*

Propósito

Este documento tiene como propósito volcar la información técnica relacionada al desarrollo de un sistema de identificación de hablantes basado en estadísticas sobre formantes, en el marco del proyecto PID 35891 de FonCyT. Mediante el mismo, se pretende construir un repositorio unificado de la información técnica del proyecto, que resulte útil a quienes a partir de ahora participen en su desarrollo y sirva como punto de partida para posibles proyectos futuros.

Introducción

El proyecto PID 35891, "Desarrollo de técnicas para el reconocimiento del hablante" tiene como objetivo introducir el desarrollo de las técnicas de reconocimiento del hablante para su aplicación a nivel forense. El reconocimiento automático del habla y del hablante, es un campo multidisciplinario con especial vinculación con las ciencias de la computación, el reconocimiento de patrones, la inteligencia artificial y la fonética acústica.
Con el objeto de comparar los resultados de las nuevas técnicas se desarrollen en el proyecto, es preciso desarrollar un sistema que implemente la funcionalidad del software IDEM que utiliza acutalmente la institución adoptante (Gendarmería Nacional). Esto es posible porque las técnicas que utiliza el mencionado sistema de identificación de hablantes fueron publicadas en congresos referidos al tema [1, 2].
El sistema IDEM está compuesto de módulos que permiten realizar en secuencia tareas específicas necesarias para resolver el problema del identificación de hablantes. De todos los módulos implementados, el que interesa a los fines de este trabajo es el módulo *SPREAD* que implementa las rutinas necesarias para hacer los análisis estadisticos, con el fin de identificar voces.

Análisis estadísticos implementados

A continuación se detallan los análisis estadísticos del módulo *SPREAD* que fueron implementados. Todos ellos tienen como punto de partida las mediciones de valores de formantes y frecuencia fundamental (F0) de las vocales /a/,/e/,/i/,/o/. La vocal /u/ no se considera por ser poco frecuente. Mediante los análisis del módulo *SPREAD* es posible cual es la probabilidad de que dos conjuntos de datos de formantes y F0 corresponden al mismo hablante. Para ello se modela la variabilidad inter-hablante e intra-hablante mediante matrices de covarianza de las variables en estudio.
Es posible realizar dos tipos de test:

- Test de χ^2: En este test la matriz de covarianza es la misma para todos los hablantes, difiriendo únicamente en la media de los formantes y F0 de cada uno de ellos. Dados n1 y n2, el numero de observaciones de un hablante y otro, respectivamente, se calcula la distancia de Mahalanobis que luego se múltiplica por *(n1*n2)/(n1+n2)*. Esto último asegura que el estadístico seleccionado sigue una distribución χ^2 de con un número de grados de libertad dado por el número de formantes en consideración. Luego es posible realizar un test de hipotesis definiendo una región de aceptación. Este test se ejecuta para cada una de las vocales en consideración que luego son promediados, calculando la probabilidad de falsa identificación, que debe ser lo más pequeña posible. SPREAD define un α=2%.
- T^2 de Hotelling: En este test, la matriz de covarianza es la misma para ambos hablantes y

viene estimada por internamente con los datos disponibles. En este caso, el resultado del cálculo de la distancia de Mahalanobis se multiplica por *(n1+n2)*(n-m-1)/(m*n*(n-2))*, siendo *m* la cantidad de formantes en consideración. Esta última operación garantiza que el estadístico seleccionado sigue una distribución *F* de Fisher. La conclusión del test es análoga al test de χ^2.

Instalación

Para que programa funcione correctamente es necesario instalar previamente MATLAB Compiler runtime. Esto se consigue haciendo doble clic sobre el archivo MCRInstaller.exe incluido con el programa y siguiendo las instrucciones del asistente.

Test de comparación de voces

Precondiciones

Antes de poder ejecutar un test de reconocimiento se debe contar con los archivos DAT que contienen la información de frecuencia fundamental y formantes de las voces que se van comparar.

Procedimiento

Al abrir el programa se presenta la pantalla descripta por la figura 1. Desde ella se pueden seleccionar los archivos DAT que se desean analizar y asignar cada uno a una identidad particular.

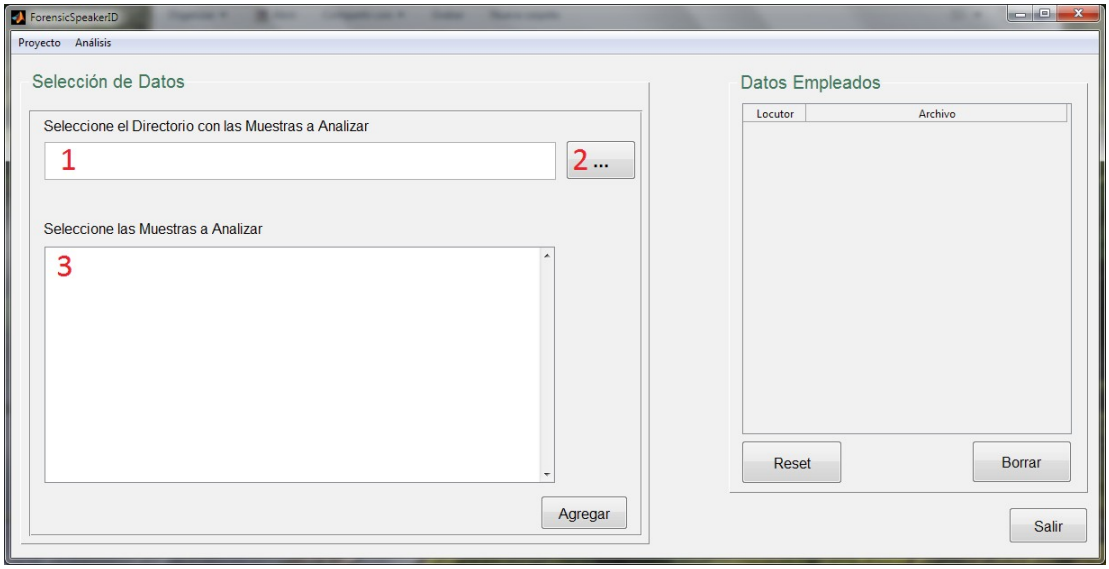
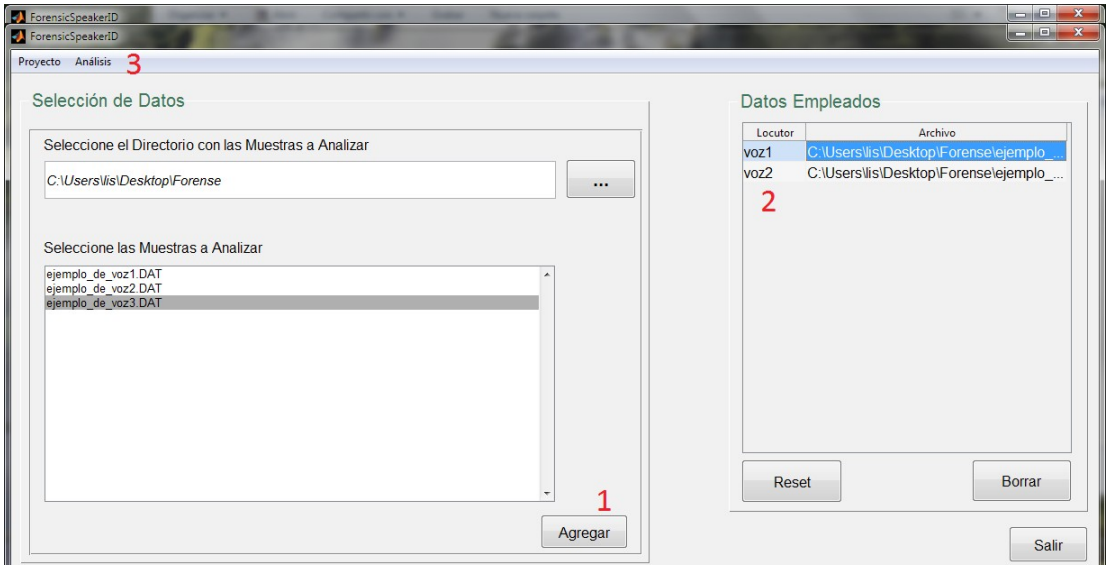
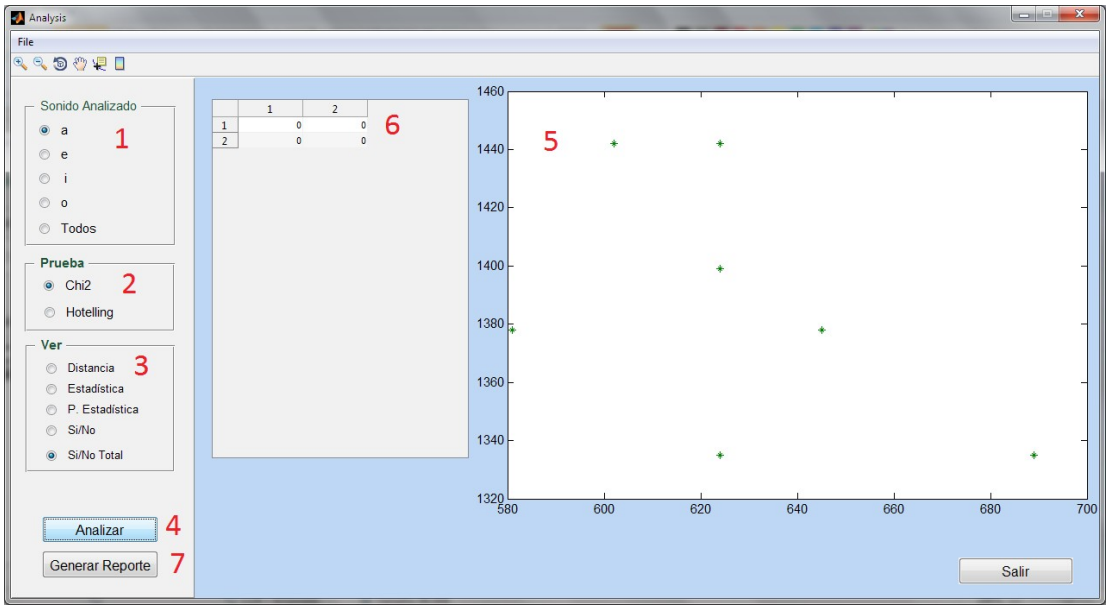


Figura 1

Para cargar los archivos DAT, en primer lugar debemos especificar el directorio en el cual estos se encuentran. Podemos poner la ruta completa en el cuadro de diálogo (1 en la figura) o abrir un cuadro de diálogo para navegar por la estructura de carpetas (presionando en 2 en la figura).
Una vez seleccionado el directorio, los archivos DAT presentes en el aparecerán en el cuadro de lista especificada en la figura con el número 3.
La figura 2 presenta la misma ventana, donde ahora se ha seleccionado un directorio que contenía tres archivos DAT. Utilizando el botón “Agregar” (2) se selecciona aquellos archivos que se desea incluir en el proceso. Luego, es posible especificar las identidades correspondientes a cada voz en la columna “Locutor” de la lista de la derecha (2). Finalmente, pasa a la etapa de análisis presionando en el menú “Análisis” (3).



Seguidamente, se presenta la ventana de análisis (figura 3), donde se puede especificar las vocales que serán tenidas en cuenta (1), el tipo de análisis que se desea realizar (2), y la información que se desea visualizar en la matriz de confusión (se selecciona en 3 y se visualiza en 6).



Una vez seleccionados estos parámetros, se presiona el botón analizar (4). Cada vez que se modifican los parámetros es necesario volver a presionar el mismo botón para actualizar la información visualizada.
El gráfico de la derecha, presenta cada uno de los casos de las voses seleccionadas en coordenadas de formante 1 vs formante 2. La matríz de confusión (6) presenta los datos seleccionados en (3) comparando los locutores especificados en la pantalla anterior.

Vistas posibles:

- **Distancia**: Expresa la distancia de Mahalanobis entre las muestras de las dos voces.
- **Estadística**: Presenta el número correspondiente al estadístico que se utilizará en el test seleccionado.
- **P. Estadística**: Presenta la probabilidad asociada al test estadístico seleccionado. Un valor menor a 0.02 se interpreta como una correspondencia de voces. (0,2 es el alpha o nivel de significación de la prueba estadística).
- **Si/No**: Presenta la misma información que la vista anterior, pero poniendo un 1 cuando hay correspondencia y un 0 cuando no la hay.
- **Si/No Total**: Presenta la misma información que la vista anterior, pero forzando el análisis sobre todas las vocales.

Una vez realizado el análisis es posible generar un reporte escrito presionando el botón “Generar Reporte” (7).

Referencias

[1] Falcone, M., De Sario, N.: A PC Speaker Identification System for Forensic Use: IDEM, In: ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp. 169-172, Martigny, Switzerland (1994)
[2] Falcone, Paoloni, A., M., De Sario, N., Saverione, V.: IDEM: un sistema per l'analisi e la rappresentazione del segnale vocale, In: Ani XX Convegno Nazionale dell'AIA, pp. 417, Roma, Italia (1992)