

Sistema de Identificación de Hablantes Basado en Estadísticas Sobre Formantes

Informe Técnico Interno

Miguel Martínez Soler

Laboratorio de Investigaciones Sensoriales - LIS
Facultad de Medicina
Universidad de Buenos Aires

Buenos Aires, mayo de 2011

Control de cambios

Fecha	Comentario	Autor
31/05/2011	Primera versión	Miguel Martínez Soler

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 actualmente 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 reconocimiento 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 estadísticos, 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 cuál 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 n_1 y n_2 , el número de observaciones de un hablante y otro, respectivamente, se calcula la distancia de Mahalanobis que luego se multiplica por $(n_1 \cdot n_2) / (n_1 + n_2)$. 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 hipótesis 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 $\alpha=2\%$.
- 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 $(n_1+n_2)*(n-m-1)/(m*n*(n-2))$, 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 MCRIInstaller.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.

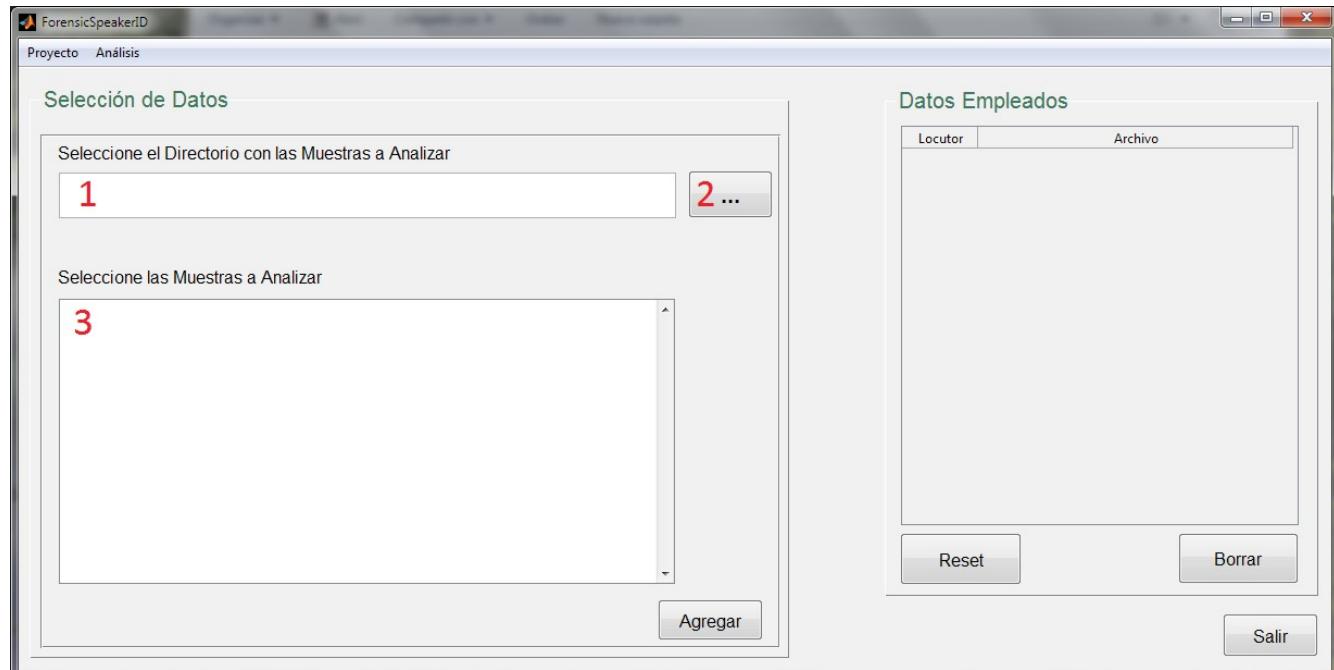
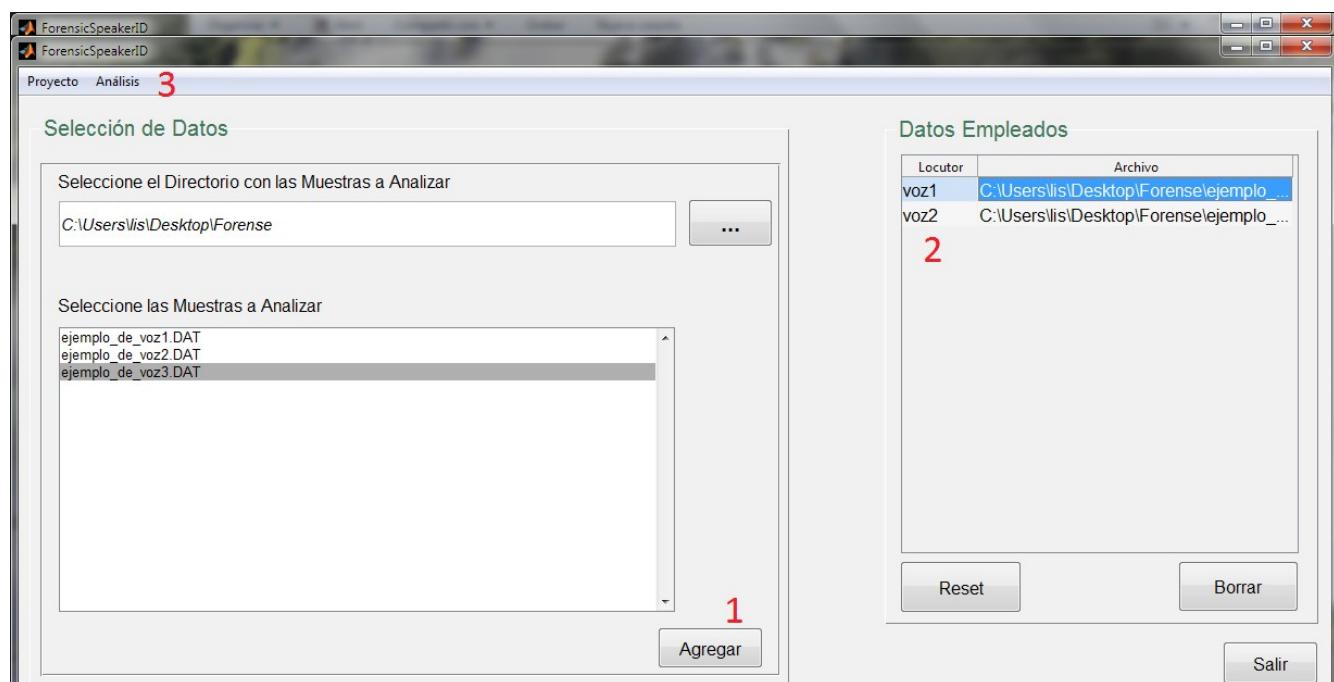


Figura 1

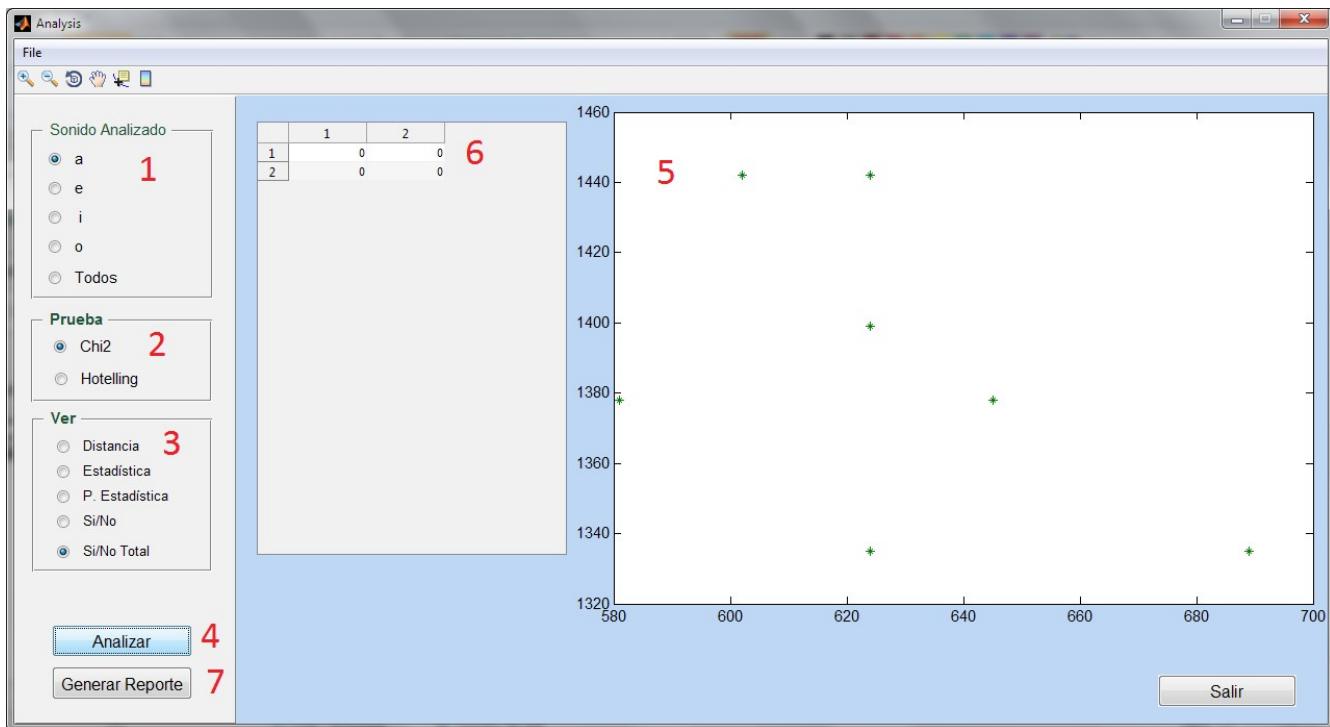
Para cargar los archivos DAT, en primer lugar debemos especificar el directorio en el cual estos se encuentran. Podemos poner la ruta completa en el cuadro de diálogo (1 en la figura) o abrir un cuadro de diálogo para navegar por la estructura de carpetas (presionando en 2 en la figura).

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 “Aregar” (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 voces seleccionadas en coordenadas de formante 1 vs formante 2. La matriz 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)