

Reporte Técnico

-Diciembre de 2010-

Desarrollo de Herramientas para el Análisis Forense de la Voz

Diego A. Evin, Jorge A. Gurlekian

Resumen

En este reporte se detallan las actividades llevadas a cabo durante el desarrollo de herramientas para el soporte del análisis forense de la voz, cuyo destinatario es el Ministerio Público de la Provincia de Buenos Aires. Asimismo se describen las principales características de los programas implementados. Específicamente estas herramientas de soporte comprenden: 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.

Desarrollo

El estudio forense de la voz involucra una serie de actividades de procesamiento que se deben aplicar sobre las señales acústicas antes de efectuar el análisis propiamente dicho, e implica además diferentes metodologías para efectuar el examen biométrico de la voz. El objetivo de las tareas de desarrollo expuestas en el reporte es construir herramientas que brinden soporte al profesional facilitando y agilizando su labor. A continuación se describen las características de cada uno de los programas implementados.

1. SEGMENTADOR DE ARCHIVOS DE AUDIO

En el ámbito forense es habitual que los registros de voz disponibles para efectuar un análisis biométrico lleguen al profesional dentro de grabaciones de considerable duración, como puede ser el caso de grabaciones correspondientes a escuchas telefónicas. En muchos casos trabajar directamente con estas grabaciones tan extensas dificulta la utilización de programas de análisis. Por lo cual se hace 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

segmentados. 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 al formato *wave*, que es el más adecuado para las fases de análisis posteriores. El programa fue desarrollado en lenguaje ***Delphi*** (*Embarcadero Technologies, Inc., San Francisco, CA, USA*).

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 duraciones mínima y máxima para cada fragmento. El usuario debe seleccionar el umbral de amplitud máximo que permite 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.

A continuación, la figura 1 muestra una captura de pantalla correspondiente a esta aplicación.

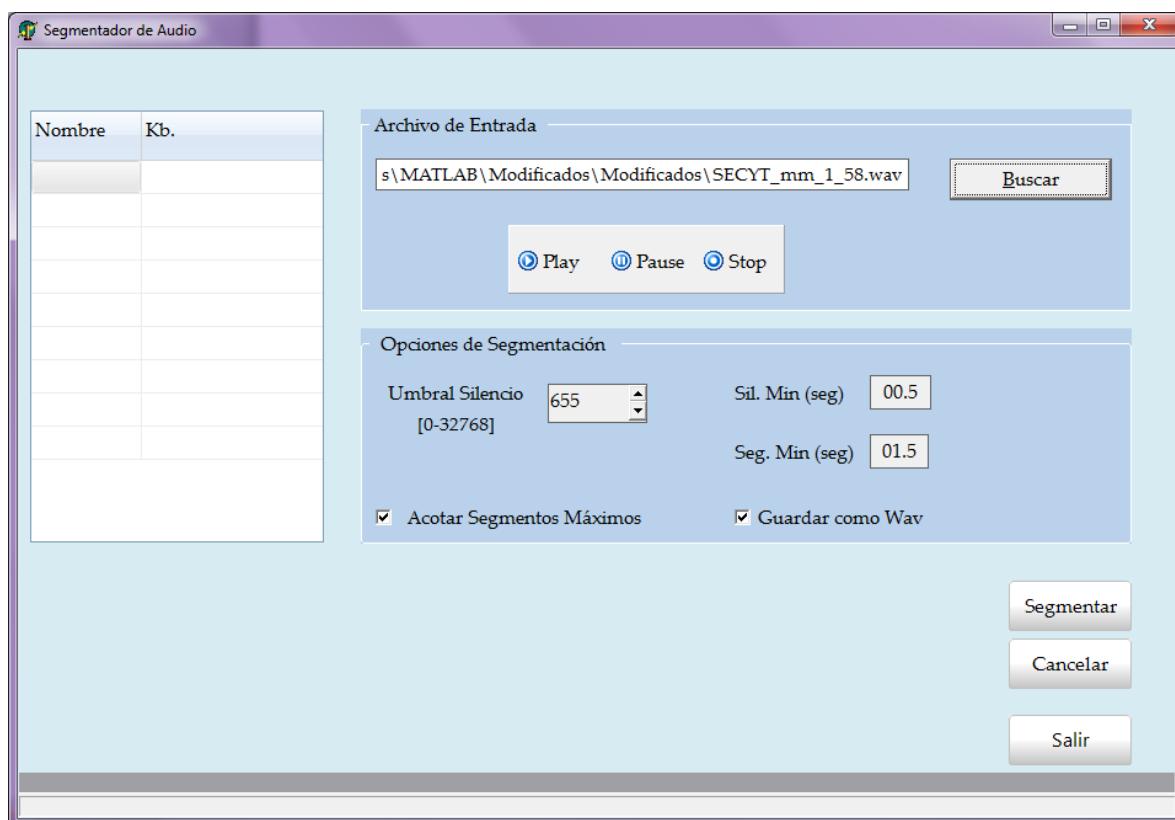


Figura 1. Aplicación para la segmentación de archivos de audio

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.

2. FILTRADO ACÚSTICO

Ya sea antes de efectuar la segmentación de los archivos de audio, o después de la misma, en algunos casos suele ser necesaria una etapa de acondicionamiento de las señales de audio que permitan reducir el ruido presente. Si bien para el análisis forense se busca no alterar las características de las señales de voz, cosa que sucede al efectuar un proceso de filtrado, como parte de las actividades de estos profesionales en muchos casos es necesario efectuar una transcripción de su contenido. El módulo de filtrado acústico busca remover componentes en la señal de audio que faciliten la comprensión durante el proceso de transcripción de las mismas.

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.

La figura 2 muestra una captura de esta aplicación, que también se implementó utilizando el lenguaje *Delphi* [1].

Al igual que en el módulo anterior, al guardar la señal procesada el programa utiliza el mismo nombre que la señal de entrada al cual agrega un sufijo, y se lo guarda en la misma carpeta que aquella. De esta forma se reduce el número de pasos de interacción con el usuario y se acelera el proceso.

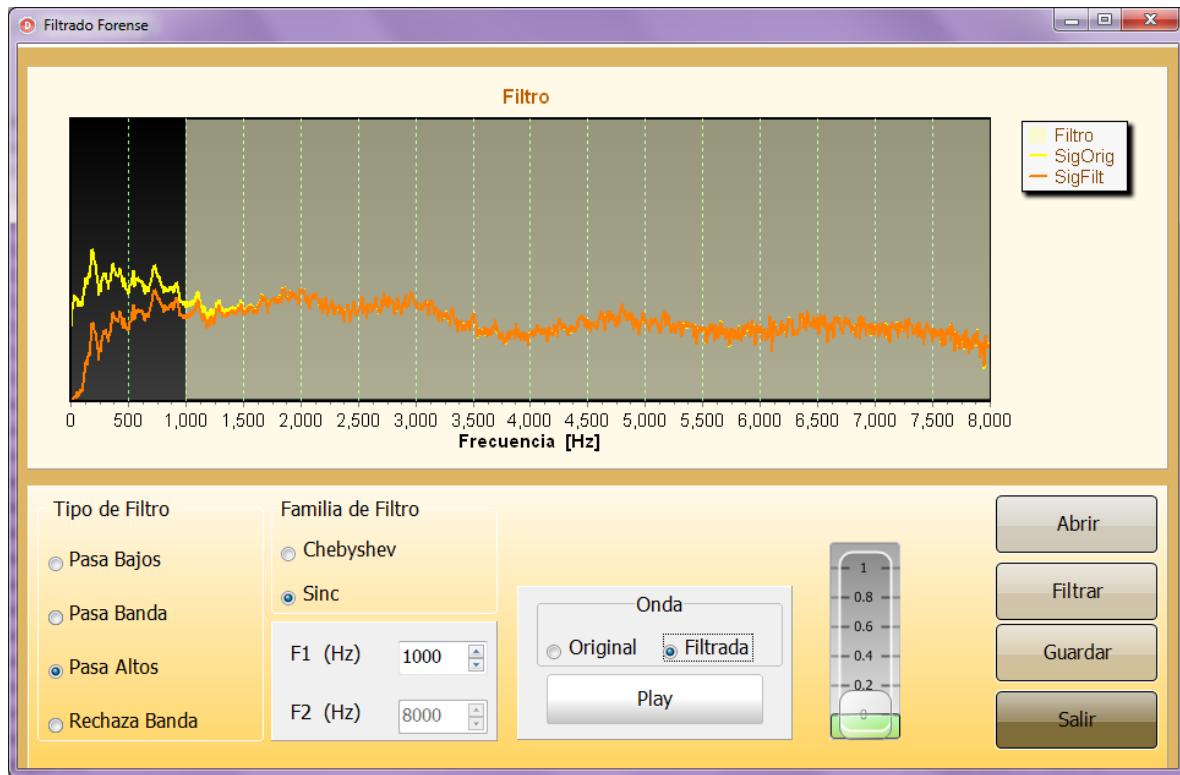


Figura 2. Aplicación para el filtrado acústico de archivos de audio

Por otra parte este programa permite que el usuario escuche la señal filtrada antes de guardarla, lo que permite experimentar con distintos parámetros y obtener la salida final más adecuada antes de efectuar la grabación.

3. 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. La figura 3 muestra una captura de esta aplicación. El usuario debe seleccionar el directorio de trabajo, y automáticamente se muestra en la tabla ubicada en la derecha, una lista con los archivos de formato *wave* presentes en esa ubicación. Al hacer click sobre uno de los nombres de archivo, se reproduce automáticamente la señal de audio, en el área central de la pantalla se muestra la forma de onda de esa 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. Para cargar automáticamente esos archivos de transcripciones se utiliza el mismo nombre que el archivo de audio elegido, pero con una extensión diferente. El usuario puede seleccionar esa extensión. Este programa permite reproducir segmentos específicos de la señal simplemente seleccionando esa región con el mouse. Presionando el click izquierdo, arrastrando liberando dicho botón se selecciona un subsegmento de la onda. Al oprimir el botón derecho del mouse se reproduce el segmento seleccionado. De esta forma se 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 [2]**. 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 acuerdo 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 no necesita escribir y posicionar las etiquetas sino solamente desplazarlas a sus posiciones definitivas.

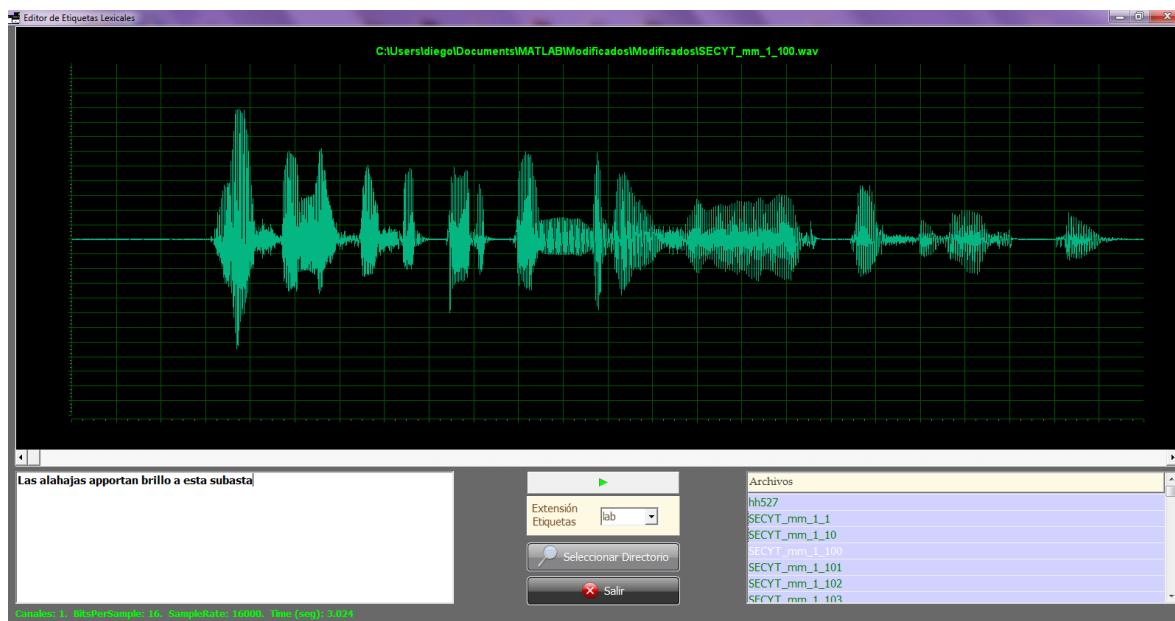


Figura 3. Aplicación para la transcripción léxica de archivos de audio

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.

Este módulo también fue implementado en lenguaje **Delphi** [1].

4. 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.

Una vez seleccionado el directorio de trabajo, el programa carga automáticamente todos los datos disponibles para cada instancia vocal 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.

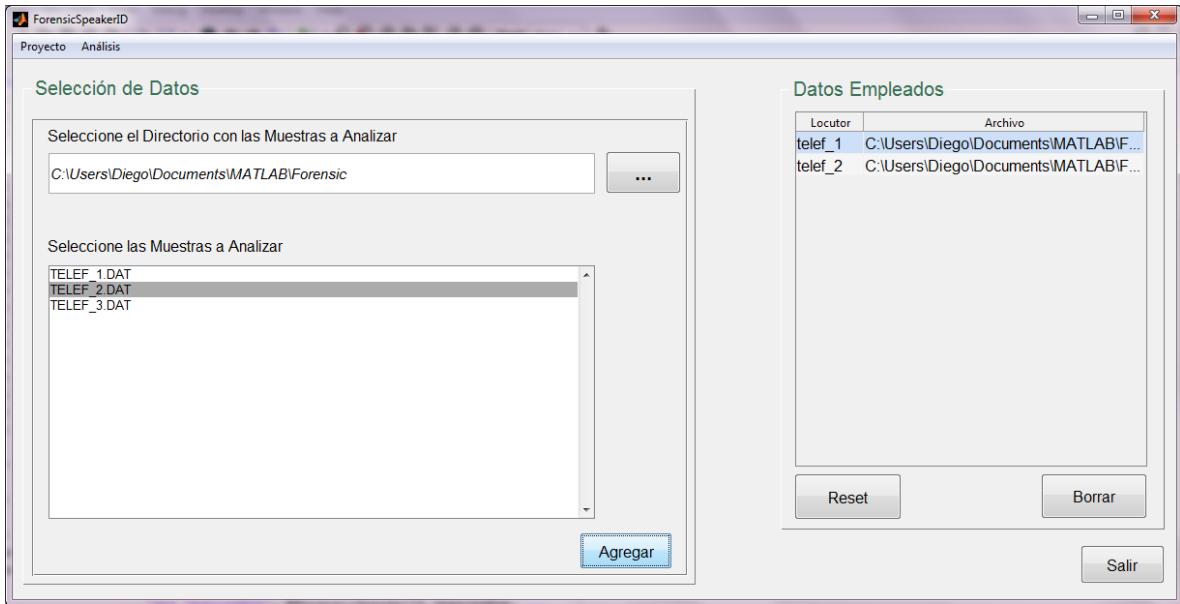


Figura 4. Ingreso de datos para el análisis biométrico de locutores

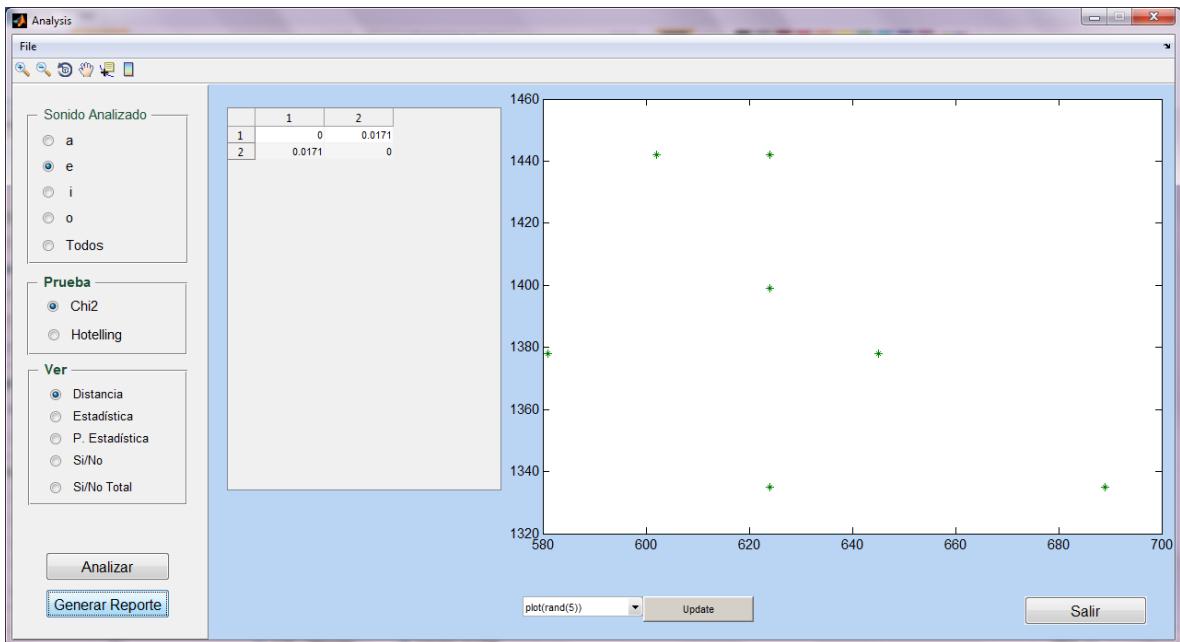


Figura 5. Cálculo estadístico para el análisis biométrico de locutores

Este módulo fue implementado en **Matlab** (*The MathWorks, Inc, Natick, MA*), y se generaron ejecutables distribuibles.

Conclusiones

A través de las actividades de diseño y desarrollo llevadas adelante se pudo implementar un conjunto de herramientas computacionales centradas en la sencillez y agilidad de uso, que brindan soporte al analista de voces forense en las distintas etapas del proceso de estudio.

Bibliografía

- [1] M. Cantù, “Delphi 2010 Handbook”. Wintech, Roma, 2010.
- [2] L. R. Rabiner, R. W. Schafer, “Digital Processing of Speech Signals”. Prentice Hall, Englewood Cliffs, New Jersey, 1978.
- [3] D. G. Childers, “Speech Processing and Synthesis Toolboxes”. John Wiley & Sons, New York, NY, 2000.
- [4] M. R. Schroeder, “Computer Speech”, 2Ed. Springer-Verlag, Berlin Heildeberg, 2004.