

DISEÑO DE UN DETECTOR DE VOZ EN CONDICIONES ADVERSAS*

Mónica de Celis, Francisco Cabrera, Eduardo Hernández, Juan L. Navarro

Departamento de Señales y Comunicaciones
Universidad de Las Palmas de Gran Canaria
navarro@dsc.ulpgc.es

ABSTRACT

We present a speech detection algorithm in adverse conditions. It is intended to be the baseline for complex speech systems. Experiments confirm previous results about the benefits of our algorithm as well as the improvement of the bispectral-based detection function over the energy-based one.

1. INTRODUCCIÓN

En el bloque de parametrización de muchas aplicaciones del tratamiento de voz es necesario un primer elemento de detección de señal. El funcionamiento de este elemento depende de muchos factores. Desde un punto de vista externo al sistema, quizás, las peores condiciones las impone el ruido de fondo. Éste puede ser tan adverso que el funcionamiento de las etapas posteriores queda influenciado negativamente por sus efectos sobre la detección. Desde un punto de vista interno nos planteamos desarrollar una aplicación capaz de lidiar con las condiciones que impone el ruido de forma satisfactoria. Además presentamos un desarrollo modular que permite ampliar a cualquier aplicación de tratamiento de voz.

Los requerimientos que se pueden pedir a un detector de voz son varios. Aquí abordamos los de mayor interés práctico: fiabilidad en la detección de los instantes de principio y fin, exactitud en las detecciones, robustez frente al ruido, adaptatividad a las condiciones que impone, simplicidad en su concepción y baja complejidad computacional.

2. APLICACIÓN DEVEOS

DeVEOS está basado en un Detector de Voz mediante Estadísticas de Orden Superior. En su concepción está pensado para abordar en primera instancia el diseño de un detector capaz de cumplir los requerimientos planteados en el apartado anterior. La idea básica [1] es aprovechar las propiedades de las estadísticas de orden superior en su versión frecuencial para eliminar el efecto del ruido sobre la señal y así facilitar su detección aun cuando la SNR sea baja. Para esto aprovechamos el concepto de bispectro integrado [2] que es un versión de una sola variable frecuencial del bispectro clásico. Conseguimos así

una función de detección con menor complejidad matemática que el bispectro a la vez que retenemos las propiedades de interés de las EOS. A efectos de comparación dispone también de una función de detección clásica basada en la energía. En ambos casos el análisis se realiza trama a trama. Una vez obtenida la función de detección se obtiene de ésta un umbral que se adapta durante los intervalos de silencio (con ruido). De la comparación entre función de detección y umbrales se obtienen las marcas automáticas de principio y fin.

Nuestra aplicación ha sido desarrollada mediante el lenguaje de programación C++ [3] bajo el entorno gráfico Windows. Se han implementado diversas clases para la realización de esta aplicación de tal manera que dichas clases C++, por a su modularidad, puedan ser utilizadas dentro de un sistema más complejo. Las principales funciones que puede hacer esta aplicación son:

- Grabación y representación de señales (.wav). En la figura 1 podemos apreciar un ejemplo de pantalla en la que se pueden ver varias de las características de la aplicación: edición de señal y su correspondiente espectrograma, etc.

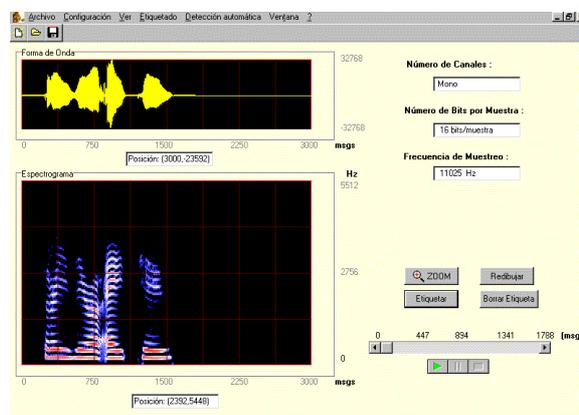


Figura 1. Pantalla principal de DeVEOS.

- Edición de marcas de principio y fin. Esta es una opción fundamental porque nos ha permitido obtener las marcas que sirvieron de referencia para evaluar el funcionamiento como detector. Una vez generadas las marcas quedan guardadas para su uso posterior.

* Este trabajo ha sido parcialmente financiado por los Proyectos Nacionales de I+D 1FD97-1183-C02-02 y TIC99-1172-C02-02.

- Experimentos de detección automática. Es aquí donde se definen los parámetros sobre los que haremos la detección (figura 2). A saber; función de detección (energía o biespectro integrado), directorio en que se guardan los resultados, tipo de ruido y SNR. Además se ha dejado a la elección del usuario tamaño y desplazamiento de las tramas y subtramas si bien por defecto aparecerán los valores que mejores resultados de detección han dado en cada caso. Para facilitar la detección de múltiples señales se puede definir el nombre de los mismos en un fichero de texto para el que se dispone de un simple editor en la aplicación.

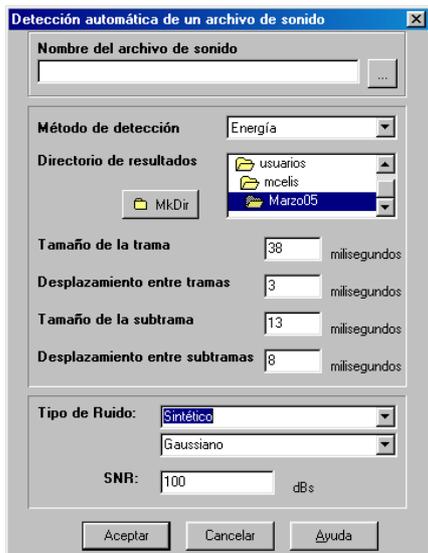


Figura 2. Experimento Detección Automática.

3. EXPERIMENTOS Y RESULTADOS

La finalidad de los experimentos que hemos realizado es doble. Por un lado, ajustar la aplicación en su totalidad: cálculo de las funciones de detección, algoritmo de adaptación de umbrales, tamaño de tramas y comparación entre funciones de detección. Por otro, pretendemos verificar hasta qué punto alcanzamos los requerimientos planteados.

Las señales utilizadas para nuestros experimentos son dígitos concatenados. La frecuencia de muestreo es de 11025 Hz y han sido cuantificados con 16 bits/muestra. En total hay 250 señales, 125 hombres y 125 mujeres. El número de parejas de marcas de principio y fines de 255. Una detección se da por buena si el error en la misma es inferior a 200 ms. Las condiciones de ruido varían desde 25 hasta 0 dB. Los tipos de ruido usados son: gaussiano, ventilador de PC y murmullo de gente.

Los resultados que presentamos son de porcentajes de acierto en principios (figura 3) y en finales (figura 4) y se obtienen del promedio de detección de todos los ruidos. En cuanto a principios hay que destacar dos cosas. La robustez es alta hasta 10 dB de SNR pues por debajo de este valor empeora claramente la detección. La detección de finales es peor que la de principios y menos robusta frente la SNR. Por otro lado, la función de

detección basada en el biespectro integrado es mejor que el basado en la energía, especialmente en los finales.

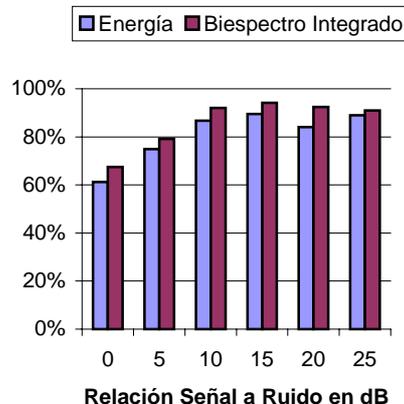


Figura 3. Principios detectados.

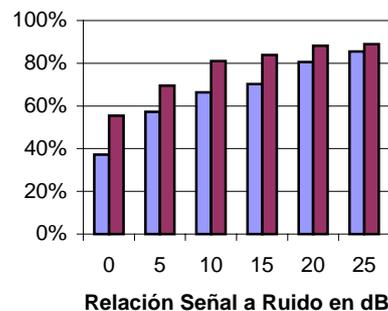


Figura 4. Finales detectados.

4. CONCLUSIONES

DeVEOS es una aplicación que permite diseñar experimentos de detección de voz. Además, posee una estructura de clases que le hace útil para el desarrollo de aplicaciones más complejas. Sus cualidades de detección cumplen los requerimientos planteados y la función de detección basada en el biespectro integrado se confirma como una alternativa plausible a la de energía.

5. REFERENCIAS

- [1] J. L. Navarro Mesa, E. Lleida Solano, "Representación Tiempo-Frecuencia no paramétricas de voz y aplicaciones". Tesis Doctoral. Departamento de Teoría de la Señal y Comunicaciones. Universidad Politécnica de Catalunya. Diciembre 1997.
- [2] J. K. Tugnait. "Detection of Non-Gaussian Signals Using the Integrated Polyspectrum". IEEE Transactions on SP, Vol. 42, nº 11, pag. 3137-3149, Nov. 1994.
- [3] Antonio Mata. "Turbo C/C++". Editorial Paraninfo. 1995.