Segmentación automática de señales de voz mediante el análisis de cambios en la entropía multirresolución continua

M. E. Torres[1], L. Gamero[1], H. Rufiner[2], C. Martinez[2], D. Milone[2], G. Schlotthauer[1]

[1] *Laboratorio de Señales y Dinámicas No Lineales* – [2] *Laboratorio de Cibernética*
Facultad de Ingeniería, Universidad Nacional de Entre Ríos,
CC47, Suc. 3 (CP 3100), Paraná, Entre Ríos, Argentina
Correspondencia: metorres@ceride.gov.ar

En el campo del Reconocimiento Automático del Habla, la segmentación de voz consiste en dividir una emisión en diferentes trozos, generalmente en fonemas aunque también suele ser de interés la segmentación según sílabas o unidades de nivel superior, como la palabra. Una emisión puede tener muchos segmentos, y así la ubicación correcta de todos sus límites puede ser un problema complejo, más aún si se consideran todas las variaciones asociadas con los distintos lenguajes.

La tarea de etiquetado de la señal de voz es bastante ardua, y el hecho de contar con una presegmentación mediante algoritmos automáticos permite disminuir considerablemente el tiempo empleado para completarla. En este trabajo se introduce la entropía multirresolución continua (EMC) como alternativa a los métodos clásicos de segmentación, la cual combina las ventajas provenientes de la entropía de Shannon y del análisis ondita. Una ventaja adicional frente a los métodos tradicionales, como los basados en Modelos Ocultos de Markov, es que no necesita de un entrenamiento previo sobre grandes cantidades de datos.

Ante la presencia de cambios dinámicos de complejidad en la señal de voz, la EMC presenta variaciones abruptas en su valor, coincidentes con la localización temporal de dicho cambio. Teniendo en cuenta este hecho, se implementó un segmentador automático que combina la EMC con las técnicas de análisis de Componentes Principales y de detección de cambios abruptos de tipo estadístico.

Se realizaron experimentos con diferentes frases de la base de datos TIMIT, donde se muestra que la entropía de Shannon por sí sola no permite distinguir las palabras a causa de sus oscilaciones, y los escalogramas obtenidos con la transformada ondita (Morlet de 5º orden) solamente permiten discriminar las transiciones de alta magnitud. El detector implementado posibilita, sobre las mismas frases, la detección del inicio y finalización entre palabras, y algunas transiciones entre fonemas.

Los resultados obtenidos indican que, con una correcta selección de la ondita y de los distintos parámetros libres de la EMC y del detector (número de voces a analizar en la descomposición ondita, tamaño de las ventanas, desplazamiento, etc.), puede desarrollarse una herramienta que permita realizar una segmentación automática de señales de voz. Este desarrollo puede posteriormente incorporarse como un módulo adicional a sistemas de análisis o reconocimiento del habla, siendo útil en la asistencia al etiquetado tanto de archivos de voz individuales como de bases de datos completas.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
M. E. Torres, L. G. Gamero, H. L. Rufiner, C. Martinez, D. H. Milone & G. Schlotthauer; "Segmentación automática de señales de voz mediante el análisis de cambios en la entropía multiresolución continua"
Memorias del XIV Congreso Argentino de Bioingeniería. pp. 125, 2003.