

Sistema de segmentación SAUTRELA

October 27, 2006

1 Condiciones de evaluación

Para el entrenamiento de modelos de locutor, se ha hecho uso exclusivo de los ficheros de audio de cada uno de los 5 locutores:

- LOC1 (17.35sg)
- LOC2 (12.51sg)
- LOC3 (13.12sg)
- LOC4 (11.17sg)
- LOC5 (10.04sg)

El ajuste de los parámetros de segmentación e identificación se ha llevado a cabo tomando como referencia el único fichero de audio inicialmente destinado a tal efecto:

- fragmento09 (265.65sg)

2 Aproximación algorítmica

2.1 Parametrización

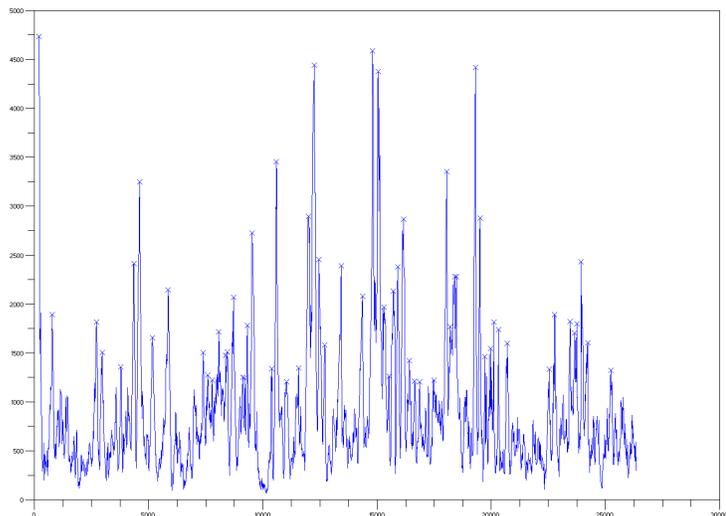
Las señales de audio han sido parametrizadas, obteniendo un vector de 12 cepstrales que excluye la energía.

2.2 Segmentación

Se toma una ventana temporal que se divide en dos partes iguales, siendo cada mitad utilizada para generar un modelo de una única gaussiana ($N(izq)$ y $N(der)$). A partir de ambos modelos se calcula lo que denominaremos *disimilitud* entre ambos segmentos de señal:

$$d(izq, der) = prob(izq)_{N(izq)} - prob(izq)_{N(der)} + prob(der)_{N(der)} - prob(der)_{N(izq)}$$

Desplazando dicha ventana, se obtiene una función cuyos valores máximos representan puntos de máxima disimilitud:



Para localizar dichos máximos se utiliza un simple criterio: es máximo todo aquel punto que supere un umbral y no tenga un valor superior en su entorno. El tamaño de ventana, desplazamiento, umbral y entorno han sido escogidos tratando de optimizar la segmentación de la señal de referencia *fragmento09.pcm*.

2.3 Identificación

Tomando las señales de cada locutor, se ha generado para cada uno de ellos un modelo de una sola gaussiana, $N(LOC_i)$. Igualmente, se genera un modelo de una sola gaussiana con el propio segmento a identificar, $N(seg)$. Sea $LOC_{max} = \arg \max_{LOC_i} (prob(seg)_{N(LOC_{max})})$ el locutor de máxima probabilidad, entonces el algoritmo de identificación utilizado es el siguiente:

$$Etiqueta(seg) = \begin{cases} LOC_{max} & , \quad prob(seg)_{N(LOC_{max})} - prob(seg)_{N(seg)} \geq umbral \\ OTROS & , \quad prob(seg)_{N(LOC_{max})} - prob(seg)_{N(seg)} < umbral \end{cases}$$

El umbral utilizado ha sido obtenido mediante la maximización de la tasa de identificación de la señal de referencia *fragmento09.pcm* (tasa obtenida mediante el software de evaluación).