

PRÁCTICA FINAL DE TRATAMIENTO DIGITAL DE LA SEÑAL DE VOZ

(Curso 2010-2011)

OBJETIVO:

Realizar un pequeño sistema de tratamiento de señal que abarque los diferentes conceptos presentados durante el curso:

- Detección de actividad de voz.
- Filtrado de señal.
- Detección de frecuencia fundamental.
- Síntesis de voz.
- Reconocimiento de habla.

Valores para toda la práctica:

Frecuencia de muestreo: 11025 Hz.

Tipo de fichero de sonido: mono, 16 bits.

ETAPAS:

1. Detección de actividad de voz.

Este módulo deberá, dada una señal de voz, detectar las zonas de silencio y de actividad de voz. Para ello, habrá que definir un umbral de energía a partir del cual se determine que estamos ante presencia de señal. La energía se calculará como la suma de las muestras al cuadrado dentro de una ventana.

Longitud de la ventana: 512 muestras.

Solapamiento entre ventanas: 75% (nos desplazamos 128 muestras).

2. Detección de frecuencia fundamental f_0 .

Aprovechando el enventanado de la etapa anterior, se realizará un detector de frecuencia fundamental basado en cepstrum, tal y como se vio en las sesiones de laboratorio. Los valores de frecuencia se darán sólo si el detector de actividad de voz decide que existe presencia de señal. Los posibles valores para la frecuencia fundamental estarán comprendidos entre 75 y 400 Hz.

3. Reconocimiento de vocales.

Se completará el parametrizador MEL cepstrum visto en las sesiones de laboratorio. Para poder reconocer vocales habrá que seguir las siguientes etapas:

Entrenamiento:

- Se pronunciarán de forma separada las cinco vocales del español: /a/, /e/, /i/, /o/, /u/.
- Se calcularán los parámetro MEL cepstrum para cada vocal, tal y como se explica en la práctica de reconocimiento de vocales.
- Para cada vocal se calculará un vector “promedio” de parámetros cepstrum que la caracterice.
- Estos vectores característicos de cada vocal se guardarán para que sirvan de modelo a las realizaciones de vocales que posteriormente se quieran reconocer.

Reconocimiento:

- Se pronunciarán nuevas realizaciones de las cinco vocales españolas.
- Cada una de esas realizaciones se detectará (módulo de detección de actividad de voz), se calculará su frecuencia fundamental (módulo de detección de f_0), se calcularán sus parámetros MEL cepstrum para cada trozo inventanado y se le asignará como reconocida la vocal correspondiente al vector característico más próximo entre los hallados en la etapa de entrenamiento.

4. Síntesis de vocales.

Una vez detectada la actividad de voz, la frecuencia fundamental correspondiente y reconocida la vocal pronunciada, se sintetizará la vocal reconocida mediante el esquema visto en las sesiones de laboratorio que tenga las características de frecuencia fundamental de la vocal reconocida.