

Muestreo y procesamiento digital de señales

Guía de trabajos prácticos: Unidad IX

Procesamiento digital de la voz

1. Objetivos

- Aplicar técnicas de procesamiento de señales para obtener parámetros característicos de la señal de voz.
- Integrar conocimientos de unidades anteriores en un caso de frecuente aplicación práctica.
- Comparar diferentes métodos para la estimación de la frecuencia fundamental.
- Conocer algunos rasgos particulares de las señales de voz.

2. Trabajos prácticos

Ejercicio 1: Es de gran utilidad obtener una aproximación a la respuesta del tracto vocal ya que ésta permite estimar a partir de los sucesivos máximos de este espectro, los valores de las frecuencias formantes de cada fonema. Por otro lado, la componente de excitación es útil, por ejemplo, para la estimación del *pitch* (entonación) de la voz.

En el archivo `aeiou.txt` se encuentra una señal de voz registrada durante la emisión de las vocales /a/, /e/, /i/, /o/ y /u/, con frecuencia de muestreo 8000 Hz. Se desea aplicar un *liftrado* para obtener aproximaciones tanto para el espectro de magnitud de la señal de excitación como para la respuesta en frecuencia del tracto vocal, para cada una de las vocales. Realice el *liftrado* suponiendo que la información de la respuesta del tracto vocal se encuentra en los primeros 30 coeficientes en el dominio de las *cuefrecias*, y luego repítalo suponiendo que se encuentra en los primeros 50 coeficientes. Utilice la estimación de las respuestas del tracto vocal a partir del *liftro* con 30 coeficientes para determinar las dos primeras formantes ($F1$ y $F2$) de las cinco vocales

y compare los resultados con la respuesta en frecuencia de un sistema obtenido mediante la técnica de predicción lineal. Para todos cálculos se recomienda utilizar 1024 muestras de la parte central (estable) de cada vocal.

Ejercicio 2: En el archivo `sent.txt` se encuentra el registro a 8KHz de una frase completa sobre la que se desea comparar diferentes técnicas para la estimación de la frecuencia fundamental (*pitch* o F_0). A partir de esta frase genere otras tres que contengan ruido blanco aditivo con relaciones señal ruido de 0, 20 y 50 dB. Finalmente compare las estimaciones de F_0 por tramos, mediante las técnicas de autocorrelación temporal y coeficientes cepstrales.