

## 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 algunas particularidades 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 20 coeficientes en el dominio de las frecuencias, y luego repítalo suponiendo que se encuentra en los primeros 40 coeficientes. Utilice la estimación de las respuestas del tracto vocal a partir del liftro con 20 coeficientes para determinar las dos primeras formantes ( $F_1$  y  $F_2$ ) de las

cinco vocales y compare los resultados con la respuesta en frecuencia de un sistema obtenido mediante la técnica de predicción lineal, con un sistema de orden 12. Para todos cálculos se recomienda utilizar 1024 muestras de la parte central (estable) de cada vocal.

**Ejercicio 2:** En los archivos `mic_F01_sa2.wav` y `mic_M01_sa1.wav` se encuentran los registros a 48 kHz de dos frases completas emitidas por un hablante femenino y uno masculino, respectivamente. Sobre las mismas se desea comparar diferentes técnicas para la estimación de la frecuencia fundamental (pitch o  $F_0$ ). A partir de estas frases 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. Además, se brinda a los fines de comparación las señales `ref_F01_sa2.f0` y `ref_M01_sa1.f0`, que contienen los valores de frecuencia fundamental obtenidos a partir de una señal de laringografía registrada en simultáneo a la grabación del micrófono. Las mismas se obtuvieron aplicando un método automático a la señal del laringógrafo, utilizando procesamiento por ventanas de 32 milisegundos de duración, con desplazamiento de la ventana de 10 milisegundos entre ventanas sucesivas (además, el procesamiento se comenzó media ventana después del inicio de la señal).