

# Práctica 2. Análisis y síntesis de la señal de voz con el modelo LPC de producción de voz

## 1 El modelo LPC de producción de voz

El modelo que se usa más frecuentemente para representar de forma paramétrica la señal de voz es el modelo todo polos. En el modelo todo polos la señal  $s(n)$  viene dada como una combinación lineal de  $p$  valores anteriores y de una entrada  $u(n)$

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (1)$$

donde  $G$  es un factor de ganancia y  $a_k$  los denominados coeficientes de predicción. Este modelo también recibe el nombre de modelo de predicción lineal o modelo LPC de producción de voz. También podemos expresar la ecuación (1) en el dominio  $z$  haciendo la transformada  $z$  a ambos lados de (1). Si llamamos  $H(z)$  a la función de transferencia del sistema entonces de (1) tenemos para  $H(z)$  la expresión

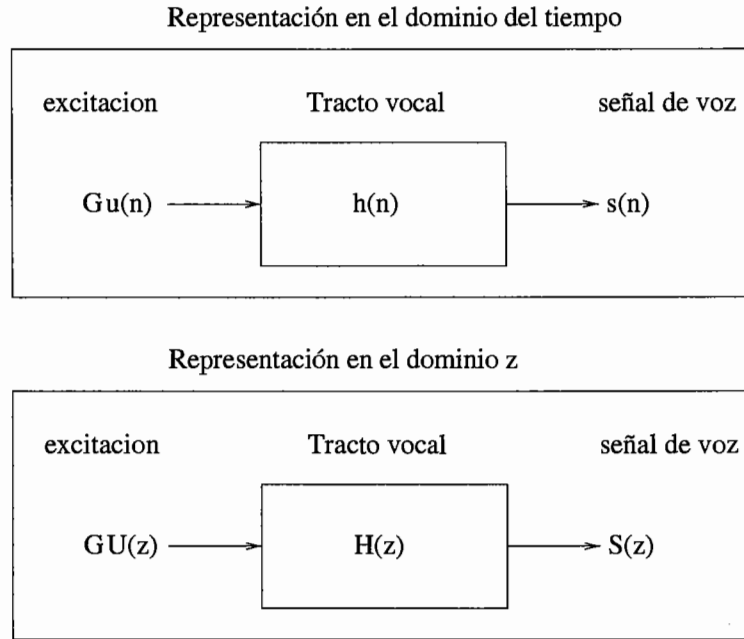
$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (2)$$

donde  $S(z)$  es la transformada  $z$  de  $s(n)$  y  $U(z)$  es la transformada  $z$  de  $u(n)$ .

Analizándolo podemos ver que con este modelo se está representando el tracto vocal con un filtro lineal  $H(z)$  y además se separa la contribución al espectro de la señal por parte de la excitación  $GU(z)$  y por parte del filtro lineal  $H(z)$ . El filtro determina la envolvente espectral mientras que la excitación crea la estructura fina. Se considera que el filtro lineal permanece constante durante cierto intervalo de tiempo, que suele durar de 10 a 20 ms, modelándose la evolución del tracto vocal como una sucesión de filtros lineales estacionarios. En la figura 1 se muestra un esquema del modelo LPC de producción de voz tanto en el dominio del tiempo como en el dominio  $z$  ( $h(n)$  es la respuesta impulso del filtro lineal  $H(z)$ ).

## 2 Análisis y representación paramétrica de la voz usando el modelo LPC

El modelo LPC nos suministra una forma alternativa de representación de la señal de voz. Esta queda representada por los coeficientes de predicción lineal  $a_k$  del filtro que caracteriza al tracto vocal  $H(z)$  y por la excitación  $Gu(n)$ . Por consiguiente, para obtener una representación paramétrica

Figure 1: *Modelo LPC de producción de voz*

de la señal de voz de esta forma hemos de tomar la señal de voz, dividirla en segmentos de tamaño entre 10 y 20 ms (se considera que durante ese tiempo el filtro que caracteriza al tracto vocal no cambia) y determinar para cada segmento de voz el valor de los coeficientes  $a_k$  y de la excitación  $Gu(n)$ . A continuación describiremos en detalle como calcular cada uno de esos parámetros. En esta práctica utilizaremos un tamaño de segmento de 15 ms.

- 1. Coeficientes  $a_k$ . Estos se determinan para cada segmento a partir de las muestras de voz de dicho segmento y el procedimiento consiste básicamente en resolver un sistema lineal de ecuaciones, tal y como se ha estudiado en clase. Existen dos métodos para determinar los  $a_k$ , que son el método de autocorrelación y el método de covarianza. En esta práctica se va a utilizar el método de autocorrelación, utilizándose para ello la función `lpc` de Matlab. El número  $p$  de parámetros  $a_k$  mínimos para caracterizar lo suficientemente bien la señal de voz viene dado por  $p = F_s + 2$  donde  $F_s$  es la frecuencia de muestreo en kHz.
- 2. Excitación  $Gu(n)$ . Una vez conocidos los coeficientes  $a_k$  para un segmento, se puede determinar fácilmente la excitación  $Gu(n)$  para dicho segmento, simplemente calculando

$$Gu(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (3)$$

donde se ha de tener en cuenta que para calcular la excitación del segmento actual se han de considerar también las  $p$  muestras de voz anteriores del segmento precedente.

## 2.1 Realización práctica

- a) Determine la señal de excitación  $Gu(n)$  y los parámetros  $a_k$  de las secuencias de filtros que modelan las señales de voz de los ficheros de voz del directorio **fichvoz**. Para comprobar que

ha realizado bien el cálculo de la señal de excitación, vuelva a obtener la señal de voz original, pasando dicha excitación por el filtro del tracto vocal  $H(z)$ , es decir, realizando la operación

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (4)$$

- b) Tome varios segmentos sonoros de voz y determine los espectros de las señales correspondientes (Matlab dispone de la función `fft` que realiza la transformada discreta de Fourier). Obtenga para cada segmento el espectro de la respuesta impulso  $h(n)$  del filtro  $H(z)$ , que se determina  $h(n) = \sum_{k=1}^p a_k h(n-k) + Gu(n)$ , teniendo en cuenta que en este caso  $Gu(n)$  es la señal impulso unidad y que la respuesta impulso para muestras anteriores a la primera muestra del segmento vale 0. Calcule también el espectro de la excitación  $Gu(n)$  de cada segmento. Realice lo mismo para varios segmentos sordos.

## 2.2 Preguntas

- 1.¿ Qué relación encuentra entre el espectro de la señal y los espectros de la respuesta impulso y la excitación?
- 2.¿ Qué diferencias encuentra entre los espectros de la señal, de la respuesta impulso y de la excitación para señales sonoras y sordas?
- 3. Determine cuáles son los formantes para cada uno de los segmentos elegidos.

## 3 Modelo simplificado de producción de voz

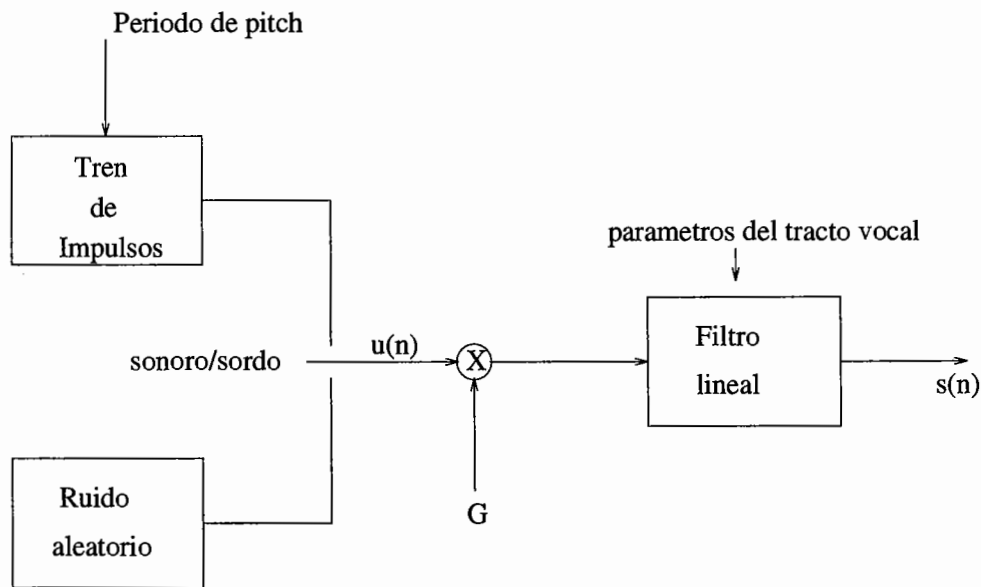
En los apartados anteriores se ha determinado de forma exacta el valor de los parámetros (coeficientes del filtro y excitación) que constituyen el modelo LPC de producción de voz. Ahora vamos a presentar un modelo LPC simplificado, en el cual lo que se simplifica es la excitación, que pasa a estar representada por un tren de impulsos con una periodicidad igual al periodo de pitch para segmentos sonoros y por ruido aleatorio para segmentos sordos, tal y como se muestra en la figura 2.

Este modelo simplificado permite representar la señal de voz con un número muy bajo de bits/s y también manipular la señal de voz de forma más fácil, aunque la calidad de la señal de voz generada es más baja.

### 3.1 Realización práctica

En este apartado se trata de sintetizar y manipular señal de voz con el modelo LPC simplificado. Para ello:

- 1. Tome un fichero de voz del directorio **fichvoz** y aplicando el algoritmo desarrollado en la práctica 1 determine para cada segmento de voz si es sordo o sonoro y en este último caso el valor del periodo de pitch.
- 2. Determine el valor de los parámetros  $a_k$  para cada segmento de voz tal y como se hizo en la sección anterior.
- 3. Determine la excitación para cada segmento de voz, tal y como se hizo en la sección anterior, y a continuación calcule la energía  $E_e$  de la excitación.

Figure 2: *Modelo LPC simplificado de producción de voz*

- 4. Genere la excitación simplificada para el filtro de cada segmento de la siguiente forma
  - a) Si el segmento es sonoro la excitación  $Gu(n)$  tendrá la siguiente forma:  $u(n)$  es un tren de impulsos unidad separados un número de muestras  $T$  igual al periodo de pitch del segmento y  $G$  se calcula  $G = \sqrt{E_e/L}$  donde  $L$  es el número de impulsos unidad que caen en el segmento.
  - b) Si el segmento es sordo la excitación  $Gu(n)$  tendrá la siguiente forma:  $u(n)$  es una señal aleatoria de media cero con varianza unidad (Matlab dispone de una función **randn** para generar una señal de ese tipo) y  $G$  se calcula  $G = \sqrt{E_e/E_u}$  donde  $E_u$  es la energía de la señal  $u(n)$  generada por la función **randn** de Matlab.
- 5. Una vez generada la excitación simplificada, obtenga la señal de voz correspondiente pasando dicha excitación por el filtro  $H(z)$ .

Estudie la calidad de la señal de voz obtenida.

### 3.1.1 Manipulación de la señal de voz

El modelo simplificado de producción de voz nos permite manipular la señal de voz facilmente. Tome un fichero de voz correspondiente a un locutor femenino, doble el periodo de pitch de los segmentos sonoros en el modelo de excitación simplificado y genere una nueva señal de voz. Haga lo mismo con un fichero de voz correspondiente a un locutor masculino, pero esta vez dividiendo por la mitad el valor del periodo de pitch de los segmentos sonoros. Tome un fichero de voz y ponga el mismo valor del periodo de pitch para todos los segmentos sonoros. ¿Qué ocurre en cada uno de los casos?.