

VOCODER

Introducción

La producción de la señal de voz se realiza mediante un flujo de aire expelido por los pulmones que pasa a través de las cuerdas vocales, el tracto vocal y, en su caso, el tracto nasal en paralelo. Si durante el paso del aire las cuerdas vocales vibran producen una oclusión intermitente del paso del aire y la voz resultante es aproximadamente periódica y se denomina 'sonora'. Si por el contrario las cuerdas vocales no vibran, el sonido resultante no es periódico y se denomina 'sordo'. Por poner algunos ejemplos: fonemas sonoros son las vocales y algunas consonantes (b, d, g,...), fonemas sordos (f, s, p, t, k,...), fonemas que utilizan el tracto nasal y son sonoras (n, m,...).

Un modelo matemático muy simplificado para representar este modelo de producción de voz consiste en representar el aire expelido por los pulmones como un tren de pulsos si la señal es sonora, o ruido si la señal es sorda. Por su parte, el tracto vocal se puede modelar por un filtro autorregresivo de orden P (P+1 coeficientes).

Para representar adecuadamente los fonemas, la excitación y los filtros deben variar con el tiempo. Para ello la señal se analiza en fragmentos denominados tramas. Una variación adecuada se puede realizar cada 15ms y un fonema puede durar típicamente entre 40ms y 100ms.

Para ver cómo este modelo puede producir voz, la señal en una trama se genera de la siguiente forma:

$$y[n] = bx[n] - \sum_{i=1}^P a_i y[n-i] \quad n = 0, \dots, N_T - 1$$

Donde b es la ganancia del filtro, x[n] es la secuencia de excitación de la trama que será un tren de pulsos si la trama a generar es sonora y ruido si la trama a generar es sorda, a_i son los coeficientes del filtro con $a_0=1$ y N_T es la longitud de la trama a sintetizar

Problema

En este problema se propone:

- A: Generar una señal de voz sintética a partir de una excitación y unos filtros previamente diseñados.
- B: Generar la señal de excitación para utilizarla en unos filtros previamente diseñados.
- C: Realizar modificaciones sobre la señal sintética.

Datos:

Una señal de voz concreta f1.wav ha sido grabada y muestreada a una frecuencia de muestreo $f_m=8\text{KHz}$ obteniéndose un flujo de muestras separadas $T_m= 1/8$ ms. Estas muestras han sido analizadas en fragmentos de 30ms (240 muestras) solapados 15 ms (120 muestras) para estimar la excitación x[n] y el filtro (parámetros b y a_i ($i=0, \dots, P$)) correspondiente a ese segmento.

El resultado es una sucesión de parámetros que modelan la señal de voz cada 15ms. El fichero `sisntesis.mat` contiene:

- La *ganancia* b a aplicar en cada trama.
- Los coeficientes del filtro autorregresivo *coef* a_i ($i=0, P$) con $P=10$ y $a_0=1$
- La excitación compuesta por un vector *excit* de $N_T=120$ muestras que contiene pulsos (deltas) separados la periodicidad de la trama si ésta es periódica o ruido si la trama a generar es sorda.

Parte A Filtrado

La señal se genera de la siguiente forma:

$$y[n] = bx[n] - \sum_{i=1}^P a_i y[n-i] \quad n = 0, N-1$$

Donde en este caso b =ganancia y $x[n]=excit[n]$ y $coef[i]=a[i]$ $i=0, P$ con $a[0]=1$

Y se van concatenando las sucesivas tramas de $y[n]$ hasta lograr tener la señal sintética total.

En Python puede usar directamente la función

`scipy.signal.lfilter(b, a, x, axis=-1, zi=None)`

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.lfilter.html>

Se pide:

- Represente la respuesta impulsional del filtro correspondiente a la trama 100. Para ello deberá aplicar a la entrada del filtro la secuencia $x[n]=\{1, 0, 0, 0, 0, \dots\}$. ¿Qué duración tiene la secuencia de entrada que ha aplicado?
- Represente la salida de la misma trama cuando a su entrada se aplica su correspondiente excitación y ganancia que se encuentran en el fichero *sisntesis.mat*.
- Realice la síntesis completa filtrando todas las tramas con sus excitaciones correspondientes y concatenando las salidas.
- Compruebe que en algunos segmentos, la señal sintetizada cambia bruscamente a cero o tiene segmentos cortos de valor cero.

Los sistemas caracterizados por ecuaciones en diferencias se dicen que están en reposo cuando las condiciones iniciales $y[-1] \dots y[-P]$ son nulas. En este caso, al pasar de una trama a otra, si ponemos las condiciones iniciales nulas vemos que algunos segmentos contienen muestras nulas que no deberían estar ahí, los segmentos deberían seguir la forma de onda de la trama anterior.

- Sintetice ahora las tramas pero al cambiar de trama mantenga las condiciones iniciales del filtro, es decir, el vector *zi* debe tener los valores de las últimas muestras de la salida de la trama anterior que `lfilter` devuelve en *zf*. Verifique que no quedan trozos de señal sintética con ceros.
- Escuche y compare las señales generadas en c) y en e)

Parte B. estimación de la excitación

La señal de excitación está formada por deltas en aquellas tramas en que la señal es periódica y por ruido si la señal de voz no es periódica. En este apartado va a generar una señal de excitación que aplicará a los filtros anteriores. Para ello es preciso determinar si la trama bajo análisis es sonora o sorda y si es sonora su periodicidad. En este problema buscaremos la periodicidad por el método de autocorrelación

a) La correlación de una señal busca el parecido entre ella y ella misma desplazada. Suponga un pulso $f[n]$. Su autocorrelación es $R_{ff}[n]$. Suponga ahora la señal $s[n] = f[n] + f[n-N] + f[n-2N]$. Demuestre que su autocorrelación es $R_{ss}[m] = 3R_{ff}[m] + 2R_{ff}[m-N] + 2R_{ff}[m+N] + R_{ff}[m-2N] + R_{ff}[m+2N]$

b) La autocorrelación tiene un máximo en el origen. Compruebe que si $R_{ff}[n]$ decrece rápidamente, $R_{ss}[n]$ presenta máximos en $n = \pm N$ y $n = \pm 2N$

c) Una trama de una señal de voz sonora contiene unos cuantos periodos de la señal. Según lo visto en el apartado anterior, su autocorrelación presentará un máximo en el origen y a múltiplos del periodo N . Para determinar el periodo de la trama, basta calcular la autocorrelación de la señal en esa trama y buscar la distancia entre el origen y el primer máximo absoluto después del del origen. Desarrolle esta idea para buscar la periodicidad de cada trama de análisis. Para calcular la autocorrelación de la trama puede utilizar

```
mumpi.correlate('array1','array2','full').
```

Tenga en cuenta, en este caso, el vector resultante de calcular la correlación está entre 0 y $2N_T-1$ siendo el valor correspondiente a la **correlación en el origen** el valor que se obtiene en el **centro del vector**.

d) Si la señal es sorda, no encontrará esta periodicidad; si la relación entre $R_{ss}[N]/R_{ss}[0] < 0.7$ considere la señal sorda.

d) Con los periodos N_i calculados para cada trama i genere una secuencia de excitación con deltas separadas N_i muestras. Si la señal es sorda utilice un generador de ruido (por ejemplo, gaussiano blanco). Una dificultad que puede encontrar al generar la excitación es que la primera delta de la trama n -ésima no tiene que empezar en $n=0$ sino que debe empezar en la posición N_i -número de muestras entre la última delta y N_T .

e) Pruebe el resultado de su problema sustituyendo las excitaciones generadas en este apartado con los filtros definidos en `sisntesis.mat`

Parte C. Modificación de la señal sintética

f) Genere la excitación de las tramas periódicas modificando el valor del periodo encontrado en cada trama N_i . Por ejemplo, podría modificar el periodo de cada trama incrementándolo o decrementándolo un 20% (y pasándolo a entero) o ser muy creativo. Sintetice la señal y escuche el resultado.