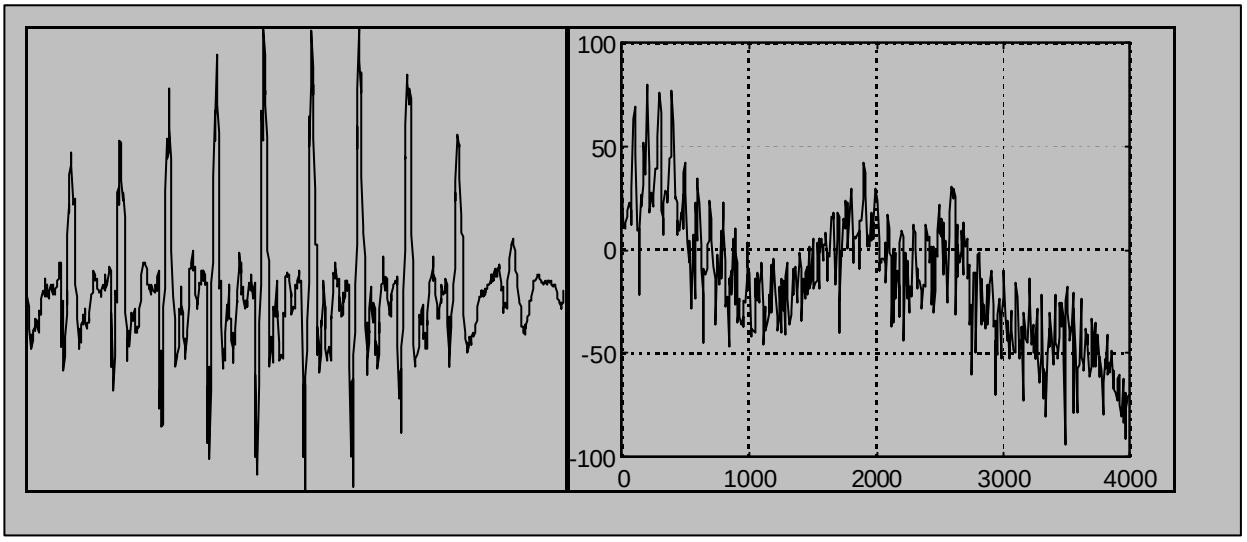
DEPARTAMENTO DE INGENIERÍA
ELECTRÓNICA Y COMUNICACIONES

CENTRO POLITÉCNICO SUPERIOR

UNIVERSIDAD DE ZARAGOZA

TECNOLOGIAS DE LA VOZ



PRÁCTICA I

LA SEÑAL DE VOZ
Características acústico-fonéticas: Generación

LA SEÑAL DE VOZ
Características acústico-fonéticas
Generación

I. Introducción

El procesado digital de señales de voz es una de las áreas mas fructíferas dentro del campo de aplicaciones del procesado digital de señal. Para poder comprender y solventar las dificultades con las que nos encontramos a la hora de desarrollar tecnologías de la voz, esta primera práctica tiene como objetivo básico el estudio de la señal de voz desde el punto de vista acústico-fonético. Para ello, estudiaremos características temporales y frecuenciales de sonidos básicos, relacionando estas características con el proceso de generación de la señal de voz.

Esta práctica es un complemento al capítulo II (Generación de la señal de voz: Producción) del temario. Al finalizar esta práctica, el alumno debe consolidar:

1. Identificación de sonidos sonoros/sordos
2. Noción de pitch y formantes
3. Noción de coarticulación
4. Fonemas y alófonos
5. Modelos de pulso glotal y tracto vocal
6. Síntesis de vocales

II. Estudio previo

Nota importante

El estudio previo consiste en una serie de ejercicios que hay que realizar **antes** de iniciar la práctica. Una copia del mismo hay que entregarla al iniciar la práctica en el laboratorio. Los conceptos teóricos necesarios para realizar la práctica han sido desarrollados en el tema II de las clases teóricas.

1.

La figura 1 muestra la forma de onda de una realización por parte de un locutor masculino de la frase ***"Los bárbaros invadieron el imperio romano"***.

a. Escribir la representación fonética de la frase utilizando el sistema SAMPA (apéndice A y B).

b. Realizar, de la forma más precisa posible, la segmentación fonética de la forma de onda, indicando el inicio y final de cada fonema sobre la forma de onda. Hacer una tabla indicando las muestras de inicio y final (aproximadamente) de cada fonema.

2.

La figura 2 muestra cuatro formas de onda y sus correspondientes espectros estimados mediante una FFT de cuatro alófonos.

a. Para cada uno de ellos, indicar:

- tipo de excitación
- para los sonoros, frecuencia de pitch
- Posición aproximada de los formantes

b. Dad una posible lista de fonemas que podrían corresponderse con estos alófonos. Justificar la asignación

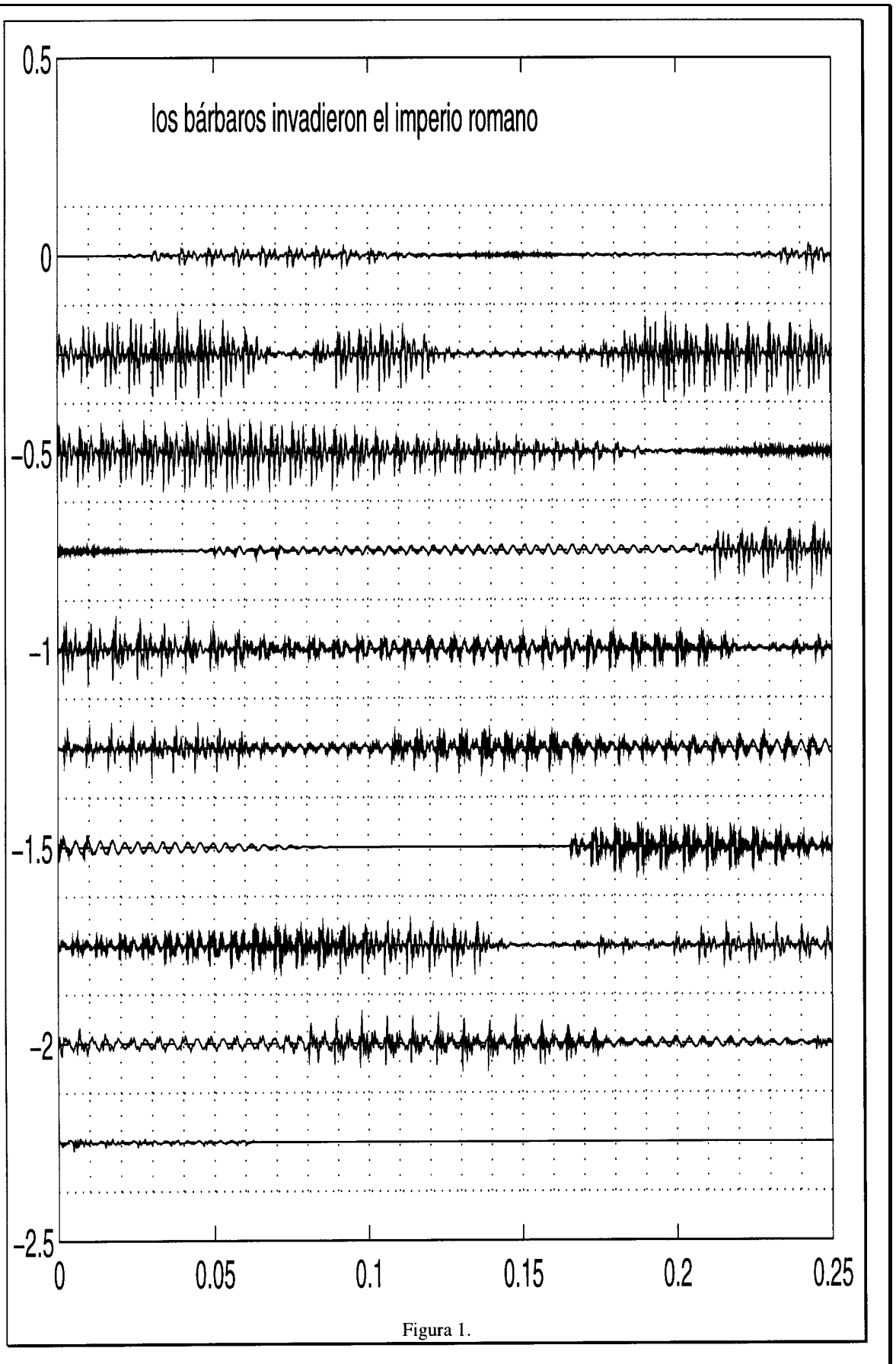
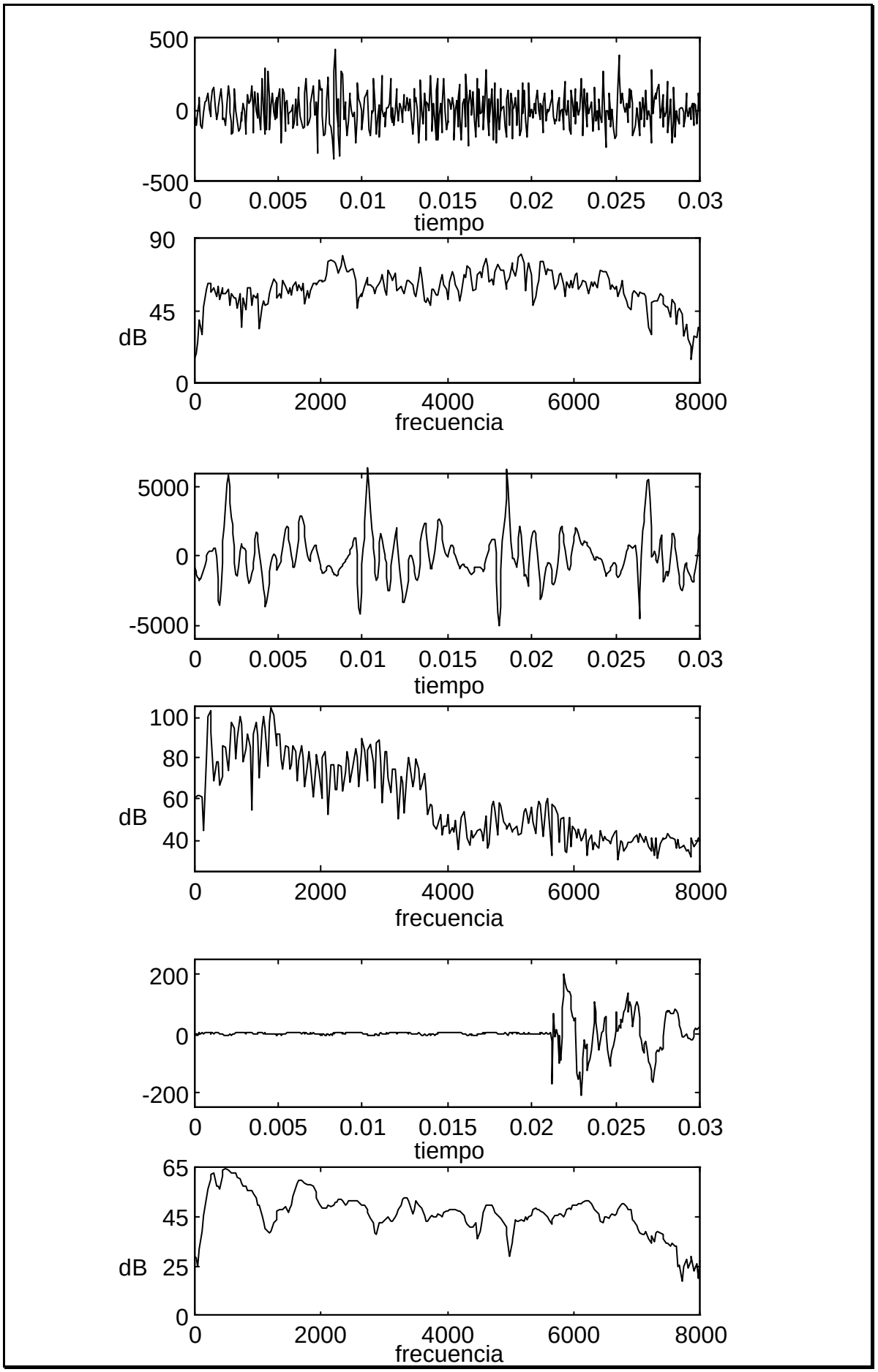

los bárbaros invadieron el imperio romano
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.05
0.1
0.15
0.2
0.25

Figura 1.


500
0
-500
0
0.005
0.01
0.015
0.02
0.025
0.03
tiempo
90
45
dB
0
0
2000
4000
6000
8000
frecuencia
5000
0
-5000
0
0.005
0.01
0.015
0.02
0.025
0.03
tiempo
100
80
60
dB
40
0
2000
4000
6000
8000
frecuencia
200
0
-200
0
0.005
0.01
0.015
0.02
0.025
0.03
tiempo
65
45
dB
25
0
0
2000
4000
6000
8000
frecuencia

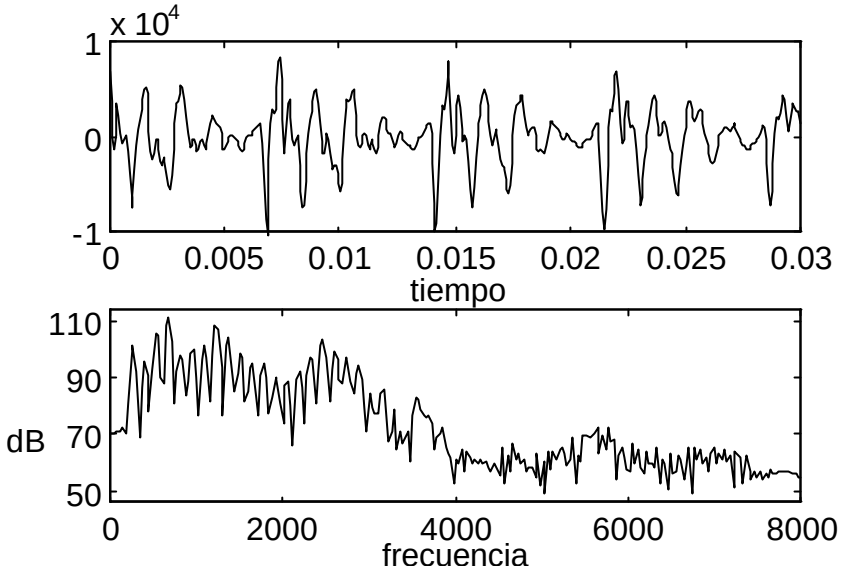


Figura 2.

3.

La figura 3 muestra el modelo básico de producción digital de la señal de voz. Basado en el filtrado inverso de la señal de voz, Rosenberg (1971) definió el modelo de pulso glotal representado por la ecuación

$$g_R[n] = \begin{cases} \frac{1}{2}\left[1 - \cos\left(pn / N_1\right)\right] & 0 \le n \le N_1 \\ \cos\left[p\left(n - N_1\right) / \left(2N_2\right)\right] & N_1 \le n \le N_1 + N_2 \\ 0 & \textit{otros valores} \end{cases}$$

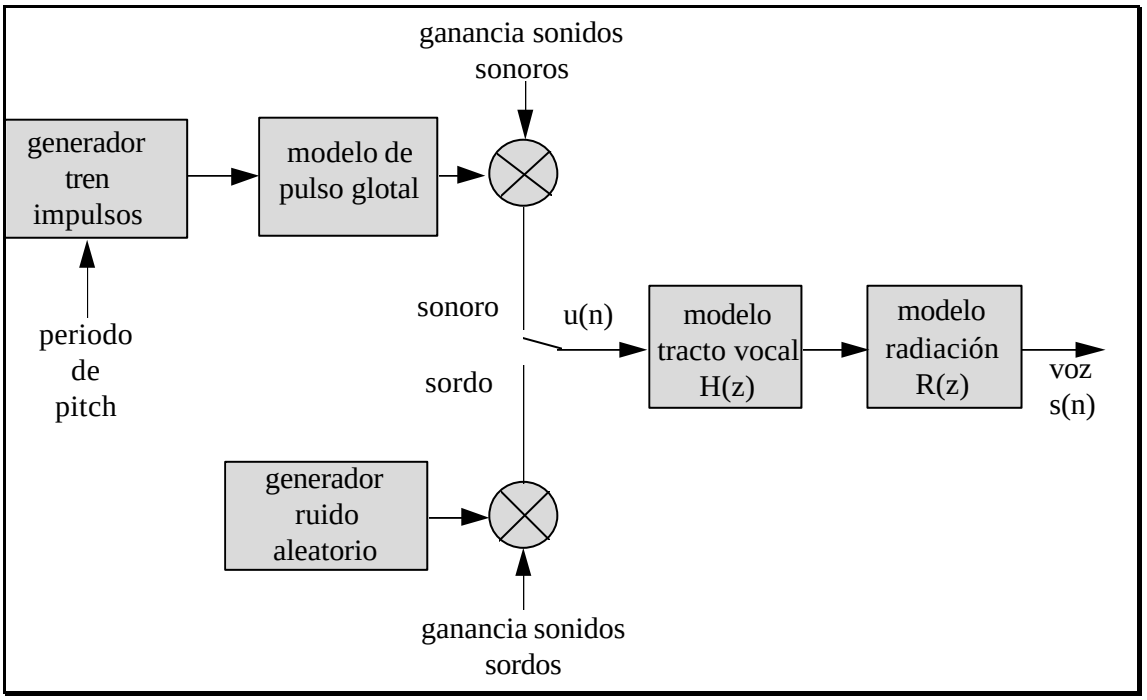


Figura 3.

Escribir una función de Matlab que calcule los $N_1 + N_2 + 1$ puntos del pulso glotal de Rosenberg y calcule la respuesta frecuencial del pulso glotal. La función debe llamarse

[g,G,W]=glottalR(N1,N2,Nfreq)

donde g es la forma de onda del pulso glotal de longitud N1+N2+1, G es la respuesta frecuencial del pulso glotal en las Nfreq frecuencias W entre 0 y π radianes.

4.

Una aproximación utilizada para modelar la transmisión del sonido por el tracto vocal es la suposición de que el tracto vocal puede ser modelado por una concatenación de tubos acústicos de distinta sección e igual longitud y sin pérdidas, como el presentado en la figura 4.

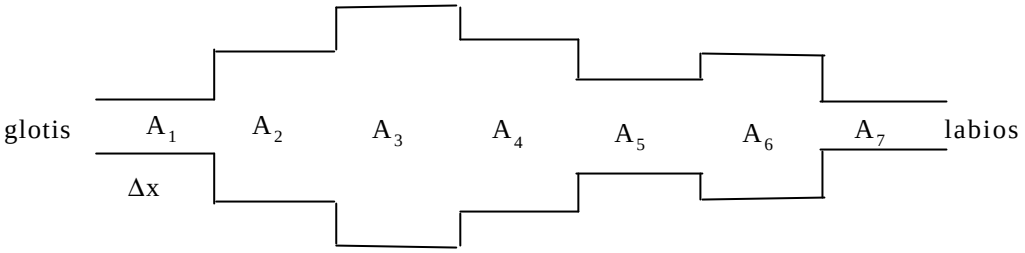


Figura 4.

En tema II hemos visto que para señales muestreadas con un periodo T = 2 τ, donde τ=Δx/c, este modelo se corresponde con un filtro digital tipo celosia (lattice filter) cuya función de red para r_G=1 (no hay pérdidas en la glottis) se puede expresar como

$$V(z) = \frac{\prod_{k=1}^{N}(1+r_k)z^{-N/2}}{D(z)}$$

El polinomio del denominador D(z) cumple la siguiente recursión

$$D_0(z) = 1$$
$$D_k(z) = D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}) \quad k = 1, 2, \ldots, N$$
$$D(z) = D_N(z)$$

donde los r_k son los coeficientes de reflexión en la conexión de los tubos,

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k}$$

En este modelo se supone que todas las pérdidas se producen en la unión con los labios a través del coeficiente de reflexión $r_L = r_N$ y A_{N+1} es el área de una tubo de impedancia adaptada (no hay reflexión en su final) que puede ser elegida para introducir las pérdidas en el sistema.

Despreciando el retardo de N/2 en V(z), podemos reescribir la función de red como,

$$V(z) = \frac{G}{D(z)} = \frac{G}{1 - \sum_{k=1}^{N} a_k z^{-k}}$$

Se dispone de una función en Matlab que nos permite calcular los coeficientes de reflexión y el polinomio D(z) dado un array de áreas de tubos y el coeficiente de reflexión en los labios. Sea la llamada a dicha función de la siguiente forma

[r,D,G]=AtoV(A,rN)

donde rN es el coeficiente de reflexión en los labios (abs(rN)<1), A un array de áreas, D un array con los coeficientes del denominador, G el numerador de la función de red y r los coeficientes de reflexión correspondientes.

Escribir una función en Matlab que nos permita dibujar la secuencia de tubos, la respuesta frecuencial del sistema y dé como salida la respuesta impulsional del sistema. La llamada a la función será

[h,H,W]=tracvo(A,rN,Nfreq)

donde h es la respuesta impulsional, H es la respuesta frecuencial en Nfreq frecuencias W de 0 a π radianes, rN es el coeficiente de reflexión en los labios A es el array de áreas.

5.

Para sonidos sonoros, la señal de excitación es un tren de impulsos cuasi-periodicos que excita al modelo de pulso glotal, al tracto vocal y al modelo de radiación. El modelo de radiación se corresponde con una función de sistema paso-alto definida como

$$R(z) = 1 - z^{-1}$$

Escribir una función en Matlab que permita sintetizar un sonido sonoro a partir de un array de áreas y el coeficiente de reflexión en los labios. Para ello hacer uso de las funciones anteriores y las funciones **filter()** y **conv()** de Matlab. La llamada a la función será

s=sinte(A,rN,fpitch,N1,N2,Nperiod,fmues)

donde fmues es la frecuencia de muestreo, Nperiod es el número de periodos de señal a sintetizar, N1 y N2 son los parámetros que definen el pulso glotal de Rosenberg, fpitch es la frecuencia de pitch de la señal a sintetizar, rN es el coeficiente de reflexión en los labios, A es el array de áreas y s es la señal sintetizada.

III. Realización de la práctica

Nota importante

Para la realización de la práctica en el laboratorio es imprescindible haber realizado y presentado al inicio de la misma una copia del estudio previo.

Parte 0. Conversión A/D y D/A de la señal de voz

Los ordenadores utilizados en la realización de las prácticas disponen de una placa de adquisición tipo Sound Blaster que permite realizar la conversión A/D y D/A de señales de audio a partir de programas propios de Sound Blaster o a traves de Matlab (Conversión D/A, de momento). Como vamos a trabajar con Matlab, veamos el proceso a seguir para la conversión A/D y D/A desde el entorno Matlab.

Conversión D/A

Para la conversión D/A, Matlab dispone de una función que nos permite realizar la conversión D/A de una señal almacenada en un array. Dicha función se llama de la siguiente forma

sound(x,fs)

donde x es el array de muestras y fs es la frecuencia de muestreo. Para mas información **help sound** en el entorno Matlab.

Conversión A/D

De momento no disponemos de una función sobre PC para la conversión A/D directamente desde Matlab. Sin embargo, podemos hacer el proceso de conversión A/D a través del programa **Grabadora de Sonidos** del sistema Windows. Este programa permite convertir A/D y almacenar la señal en ficheros tipo **.wav** que podemos leer desde Matlab con la función **wavread()** (para mas información sobre la función **help wavread** en el entorno Matlab). Antes de realizar la conversión A/D, debemos fijar el nivel de grabación y los valores de grabación (número de bits, frecuencia de muestreo, etc). Para ello disponemos del menu **opciones** donde podemos activar el mezclador para fijar los niveles de grabación y fijar los parámetros de grabación. Como valores fijos utilizaremos **16 bits** y grabación **mono**. La grabación se activa presionando con el ratón en el botón con el círculo rojo (botón de la derecha). La grabación se puede escuchar pulsando en el boton de la izquierda (play). Una vez validada la grabación, en el menu **fichero** se puede grabar la señal adquirida en un fichero con extensión **.wav**.

1. Utilizando el programa **Grabadora de Sonidos** convertir y escuchar a distintas frecuencias de muestreo desde 44 kHz a 8 kHz la frase

las seis casas se sostienen fuera por si solas

¿Cómo afecta la frecuencia de muestreo en la intelegibilidad de los distintos sonidos de la frase?

Parte 1. Segmentación y etiquetado fonético

En este apartado vamos a verificar y precisar mejor la segmentación realizada en el estudio previo. Para ello leer el fichero **nino.mat** mediante el comando **load**. El resultado es el array **sam** que contiene las muestras de la frase utilizada en el estudio previo con una frecuencia de muestreo a 16 kHz.

2. Repetir el apartado 1 del estudio previo para precisar mejor la segmentación realizada utilizando las posibilidades tanto gráficas como de procesado de señal de Matlab.

3. Leer el fichero **roma.mat** (variable **samples**) y comentar las diferencias mas notables con relación a la señal anterior.

4. Verificar que en ambas señales, la transcripción fonética es la adecuada.

5. ***(Ópcional)*** Grabar la frase "***Los bárbaros invadieron el imperio romano***" con vuestra voz y comentar las diferencias más notables en relación a los anteriores ficheros.

Parte 2. Estudio del modelo de pulso glotal

Generar un pulso glotal, g_R, utilizando la función **glottalR** con N1=40 y N2=10. Crear un pulso glotal causal, g_{RC}, de la forma

$$g_{RC}[n] = g_R\left[-\left(n - N_1 - N_2\right)\right]$$

para ello utilizar las funciones Matlab **fliplr()** o **flipup()** según se trabaje con vectores línea o columna.

6. ¿Cúal es la relación analítica entre las transformadas de Fourier de ambas señales?

7. Dibujar la forma de onda de ambos pulsos, así como el módulo de la transformada de Fourier. Experimentar con los parámetros del modelo, viendo como afecta la forma del pulso temporal en su respuesta en frecuencia.

8. Utilizando la función **roots()** de Matlab, encontrad los zeros de la transformada z de los pulsos glotales g_R y g_{RC}. Dibujar su posición utilizando la función **zplane()**. Comentar las diferencias y resultados.

Parte 3. Estudio del modelo del tracto vocal

Para la realización de esta parte de la práctica, utilizaremos los siguientes valores de las áreas del modelo de tubos.

Sección	1	2	3	4	5	6	7	8	9	10
Vocal a	1.6	2.6	0.65	1.6	2.6	4	6.5	8	7	5
Vocal i	2.6	8	10.5	10.5	8	4	0.65	0.65	1.3	3.2

9. Utilizando la función definida en el estudio previo **AtoV()**, obtener el denominador D(z) de la función de red del sistema del tracto vocal para ambas vocales con rN=0.71 y rN=1. Factorizar el polinomio D(z) y dibujar los polos en el plano z utilizando **zplane()**. ¿Como afecta rN en la posición de los polos?. Convertir los ángulos de las raizes a frecuencias analógicas, suponiendo una frecuencia de muestreo de 8000 Hz.

10. Utilizando la función creada en el estudio previo **tracvo()**, dibujar el modelo de tubos y su respuesta frecuencial para los casos del apartado anterior.

Parte 4. Síntesis de vocales

Para la realización de esta parte de la práctica, supondremos que la **frecuencia de muestreo** es 8000 Hz.

11. Utilizando la función creada en el estudio previo **sinte()**, sintetizar las dos vocales utilizadas en la parte 3 de la práctica con un pitch de 100 Hz y 250 Hz, N1=40, N2=10 y rN=0.7 en ambos casos. Almacenar las señales sintetizadas para un apartado posterior. ¿Qué ocurriría si utilizamos rN=1?

12. Utilizando la función **sound()**, escuchar las señales sintetizadas.

13. Suponed que la excitación del tracto vocal es de tipo susurro (turbulencia producida en la glotis). Este tipo de excitación se puede modelar mediante un ruido aleatorio de tipo gaussiano de media cero (función **randn()**). Repetir el apartado 11 con este tipo de excitación. Escuchar las señales sintetizadas y compararlas con el caso de excitación periodica. Seleccionar un segmento de 256 muestras y utilizando la función **fft()** dibujar la magnitud en dB del espectro de la señal sintética para el caso de excitación periodica y aleatoria.

Apendice A: Notación fonética SAMPA

SAMPA		Ejemplo	Transcripción
p	explosiva bilabial sorda	**p**ala	**p**ala
b	explosiva bilabial sonora	**b**ala	**b**ala
t	explosiva dental sorda	**t**ala	**t**ala
d	explosiva dental sonora	**d**ar	**d**ar
k	explosiva velar sorda	**c**ala	**k**ala
g	explosiva velar sonora	**g**ala	**g**ala
m	nasal bilabial sonora	**m**ala	**m**ala
n	nasal alveolar sonora	**n**ada	**n**aDa
N	nasal velar sonora (precede a una consonante velar)	ho**n**go	o**N**go
J	nasal palatal sonora	ca**ñ**a	ka**J**a
tS	africada palatal sorda	**ch**ico	**tS**iko
B	aproximante bilabial sonora	la**v**a	la**B**a
f	fricativa labiodental sorda	**f**also	**f**also
T	fricativa interdental sorda	**z**ona	**T**ona
D	aproximante dental sonora	ca**d**a	ka**D**a
s	fricativa alveolar sorda	**s**ala	**s**ala
z	fricativa alveolar sonora (precede a una consonante sonora)	de**s**de	de**z**De
jj	fricativa palatal sonora	a**y**er	a**jj**er
x	fricativa velar sorda	**j**amón	**x**amon
G	aproximante velar sonora	la**g**o	la**G**o
l	lateral alveolar sonora	**l**a	**l**a
L	lateral palatal sonora	**ll**ana	**L**ana
rr	vibrante múltiple alveolar sonora	ca**rr**o	ka**rr**o
r	vibrante simple alveolar sonora	ca**r**o	ka**r**o
i	vocal anterior cerrada	t**i**la	t**i**la
j	semivocal palatal (aproximante palatal sonora)	lab**i**o	laB**j**o
e	vocal anterior media	t**e**la	t**e**la
a	vocal central abierta	t**a**l	t**a**l
o	vocal posterior media redondeada	t**o**do	t**o**D**o**
u	vocal posterior cerrada redondeada	t**u**l	t**u**l
w	semivocal labiodental (aproximante labio-velar sonora)	ag**u**a	aG**w**a

Apendice B: Reglas de transcripción SAMPA

grafema	reglas para la transcripción SAMPA	ejemplos
<a>	**a**	
<b>	después de pausa: **b** tras <m> o <n>: **b** otros casos: **B**	 c'om-**b**a: k'om-**b**a l'a-**b**io: l'a-**B**jo
<c>	seguida por <e> or <i>: **T** al final de palabra y seguida por <l> or <r>: **G** seguida por <b>, <d>, <g> (ante <a>, <o>, <u>), <m>, <n>, <ñ> o <v>: **G** otros casos : **k**	**c**'e-lo: **T'**e-lo a**c**-n'e: a**G**-n'e t'a**c**-to: t'a**k**-to **c'**oro: **k'**o-ro t'e-**c**la: t'e-**k**la
<ch>	**tS**	**ch'**e-lo: **tS'**e-lo

<d>	después de pausa: **d** tras <l>, <m> o <n>: **d** otros casos: **D**	 c'al-**d**o: k'al-**d**o c'o-**d**o: k'o-**D**o
<e>	**e**	
<f>	**f**	c'o-**f**ia: k'o-**f**ja
<g>	después de pausa y seguida por <r>, <l>, <a>, <o> o <u>: **g** tras <m> or <n> y seguida por <a>, <o> o <u>: **g** seguida por <i> o <e>: **x** otros casos: **G**	 t'on-**g**o: t'on-**g**o **g'**e-nio: **x'**e-njo t'i-**g**re: t'i-**G**re l'a-**g**o: l'a-**G**o
<h>	al inicio de sílaba y seguida por <ie>: **jj** al inicio de sílaba y seguida por <ue>: **G** otros casos: muda	**h**i'er-ba: **jj'**er-Ba a'**h**ue-c'ar: a-**G**ue-c'ar **h'**a-lo: 'a-lo
<i>	en posición silábica nuclear: **i** en posición silábica no nuclear: **j**	t'**i**-po: t'**i**-po c**i'**e-lo: T**j'**e-lo ar-m'a-r**i**o: ar-m'a-r**j**o
<j>	**x**	**j**a-r'a-na: **x**a-r'a-na
<k>	**k**	**k**i'os-**k**o: **k**jos-**k**o
<l>	**l**	**l'**ote: **l'**o-te
<ll>	**L**	t'a-**ll**o: t'a-**L**o
<m>	**m**	'ar-**m**a: 'ar-**m**a
<n>	seguida por <p>, <b>, <v>, <m> o <f>: **m** seguida por <c> (ante <a>,<o>,<u>,<r> o <l>), <g>,<j> o <qu>: **N** otros casos: **n**	'a**n**-fo-ra: 'a**m**-fo-ra 'a**n**-ca: 'a**N**-ka c'o-**n**o: 'ko-**n**o
<ñ>	**J**	'u-**ñ**a: 'u-**J**a
<o>	**o**	
<p>	**p**	**p'**e-rro: **p'**e-rro
<q>	always followed by <u>: **k**	**q**u'e-so: **k'**e-so
<r>	al inicio de palabra: **rr** tras <l>, <n> o <s>: **rr** otros casos: **r**	**r'**a-ma: **rr'**a-ma h'on-**r**a: 'on-**rr**a 'a**r**-pa: 'a**r**-pa t**r'**am-pa: t**r'**am-pa p'e-**r**a: p'e-**r**a a-m'o**r**: a-m'o**r**
<rr>	**rr**	c'a-**rr**o: k-a-**rr**o
<s>	seguida por <b>, <d>, <g> (ante <a>, <o>, <u>, <r> o <l>), <l>, <m>, <n>, <r> o <v>: **z** otros casos: **s**	 r'a**s**-go: rr'a**z**-Go c'a-**s**a: k'a-**s**a tr'a**s**-to: tr'a**s**-to
<t>	al final de sílaba: **D** otros casos: **t**	a**t**-l'eta: a**D**-l'eta **t'**o-ro: **t'**o-ro
<u>	sin diéresis tras <g> o <q> : muda en posición silábica nuclear: **u** conjunción: **w** en posición silábica no nuclear: **w**	q**u'**e-so: k'e-so l'**u**-jo: l'**u**-xo **u** 'o-tro: **w** 'o-tro ci-g**ü'**e-ña: Ti-G**w'**e-Ja
<v>	después de pausa: **b** tras <m> o <n>: **b** otros casos: **B**	 con **v'**e-lo: kom **b'**e-lo c'al-**v**o: k'al-**B**o
<w>	en palabras extranjeras: **Gu, gü** o como <v>	**w**h'is-ky: **gw'**is-ki **w'**a-ter: **B**'a-ter
<x>	en posición no inicial de palabra y ante una vocal: **Gs** otros casos: **s**	e-**x'**a-men: e**G-s'**a-men e**x**-t'er-no: e**s**-t'er-no

<y>	al inicio de sílaba y seguida de vocal: **jj** conjunción en contacto con una vocal distinta de <y>: **j** conjunción entre dos vocales distintas de <y>: **jj** conjunción en otras situaciones: **i** otros casos: se trata como <i>	**y'**un-que: **jj'**un-ke c'on-**y**u-ge: k'on-**jj**u-Ge r'a-so **y** t'ul: rr'a-so **j** t'ul se-s'en-ta **y** 'u-no: se-s'en-ta **jj** 'u-no dos **y** dos: dos **i** Dos mu'**y**: mw'**i**
<z>	**T**	**z'**ar-**z**a: **T'**ar-**T**a t'i**z**-ne: t'i**T**-ne