Proyecto Especial: Sintetizador de Voz

Sebastián Sampayo

1° de Junio de 2013

Índice

1. Objetivos **3**

2. Desarrollo **3**

2.1. Mediante el espectrograma de la señal "Habla.wav" identificar los distintos fonemas y comprobar la diferencia mencionada entre las distintas clases. . . . 3

2.2. Realizar un análisis espectral de la señal "Vocales.wav" mostrando por un lado la frecuencia glótica y por otro los formantes. ¿Qué consideraciones especiales deben tomarse en cada caso? Determinar el valor de los formantes y comparar los resultados con los del programa wavesurfer. . . . 3

2.3. Segmentar el archivo separando cada una de las vocales y graficar su espectro para distinguir mejor los formantes. . . . 5

2.4. Para una frecuencia de resonancia de 1000Hz graficar en un mismo gráfico el módulo de la transferencia en función del factor de amortiguamiento. Tomar 5 valores en el rango $0 < \xi < 0{,}6$. Determinar el valor de frecuencia angular ω_m para el cual la transferencia presenta un máximo. Comprobar que la relación entre la frecuencia máxima ω_m y ω_r está dada por $\omega_m = \omega_r\sqrt{1-2\xi^2}$. . . . 8

2.5. Determinar B_w, el ancho de banda a -3dB con respecto al valor máximo y graficarlo en función de ξ. 9

2.6. Utilizando los valores de las tablas anteriores y teniendo en cuenta lo hallado en los puntos 4 y 5, encontrar mediante aproximaciones sucesivas los parámetros de las 25 transferencias. . . . 9

2.7. Generar y graficar la señal del tren de pulsos glóticos para una frecuencia glótica de 150 Hz. . . . 10

2.8. A partir de los 5 pasabajos analógicos que modelan el tracto vocal obtener los respectivos pasabajos discretos para una frecuencia de muestreo de 16KHz. Justificar el método de conversión elegido. Graficar para una de las vocales la respuesta en frecuencia de los 5 resonadores por separado y la respuesta del sistema conectado en cascada. En caso de haber utilizado el método de transformación bilineal, en el gráfico de cada resonador dibujar simultáneamente la respuesta en frecuencia con y sin prewarping. . . . 11

Teorema de Muestreo . . . 11

Transformación Bilineal . . . 12

2.9. Obtener el filtro discreto que simula el tren de pulsos glóticos para una frecuencia de muestreo de 16 KHz. Justificar el método de conversión utilizado. Generar y graficar la secuencia de pulsos glóticos y compararla con la obtenida analógicamente. . . . 18

2.10. Sintetizar 300ms de cada una de las vocales, utilizando la cascada de los 6 sistemas excitados por el tren de pulsos generado en el punto 9. . . . 19

2.11. Realizar el análisis espectral de las vocales sintéticas concatenadas entre sí, de modo de formar la misma secuencia "a-e-i-o-u". ¿Qué diferencias encuentra entre esta señal y la del punto 2? ¿Qué diferencia encuentra al escucharlas? ¿Coinciden los picos con los de la Tabla I? ¿Coinciden con los de las señales reales? . . . 20

2.12. Grabar una vocal con su propia voz y calcular aproximadamente la frecuencia máxima y el ancho de banda de cada formante. Utilizando los valores obtenidos sintetizar la vocal mediante el modelo desarrollado en este proyecto pero utilizando como frecuencia glótica la de su propia voz. Discutir los resultados. . . . 21

1. Objetivos

2. Desarrollo

2.1. Mediante el espectrograma de la señal "Habla.wav" identificar los distintos fonemas y comprobar la diferencia mencionada entre las distintas clases.

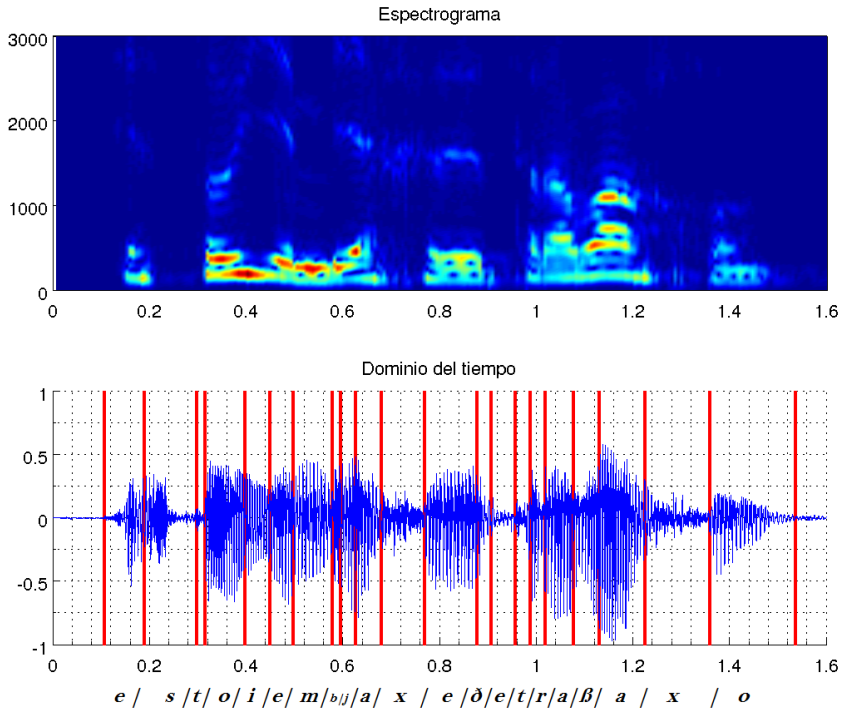

Figura 1: "Estoy en viaje de trabajo"

En el espectrograma se puede ver como los fonemas modelados por ruido (como la /s/) presentan uniformidad en frecuencia, es decir igual cantidad de energía en todo el espectro. Además, se puede ver que las modeladas con un tren de pulsos (como la /e/ o la /o/) son señales periódicas en el dominio del tiempo.

2.2. Realizar un análisis espectral de la señal "Vocales.wav" mostrando por un lado la frecuencia glótica y por otro los formantes. ¿Qué consideraciones especiales deben tomarse en cada caso? Determinar el valor de los formantes y comparar los resultados con los del programa wavesurfer.

El objetivo de las siguientes figuras es determinar por un lado la frecuencia glótica fuente del sistema, y por otro el valor de los formantes. Para esto se realizaron varios espectrogramas jugando con los valores del largo de la ventana, teniendo en cuenta que lo que se ve al "ventanear" es el espectro original convolucionado con la transformada de Fourier de la ventana. Por ejemplo, si se utiliza una ventana cuadrada, a la salida se verá el espectro original convolucionado con una *sinc*[1], y cuanto más ancha sea la ventana en tiempo, más angosta será la *sinc*. De esta manera, con una ventana larga, (y una *sinc* angosta) se distinguirá mejor la frecuencia del tren de pulsos (dado que sabemos que la fuente del sistema es un tren de pulsos periódico). Por el contrario, al disminuir el largo de la ventana (y evidenciar más la *sinc*), se resaltarán los picos del espectro, i.e., los formantes.

Por otro lado, la cantidad de puntos de la DFT, indicará la cantidad de muestras tomadas del espectro, es decir, la resolución. En este caso se eligieron 2048 puntos.

La ventana utilizada para este ejercicio fue una ventana de *Hamming*, ya que la transformada de esta es menos destructiva que la *sinc*.

En este caso en particular, en el gráfico, se puede ver que cada vocal dura aproximadamente 0.3 segundos, que corresponde a 6000 puntos con una frecuencia de muestreo de 20KHz. Esto significa que la ventana más ancha que se puede usar es de 6000 puntos.

Para distinguir la frecuencia glótica se realizó el espectrograma de la señal utilizando una ventana grande de 2048 puntos. De esta manera, se pueden observar los picos del espectro, que al ser la transformada de un tren de pulsos de período N, estarán a una distancia de $\frac{2\pi}{N}$. Si se mide esta distancia en el gráfico se obtiene una frecuencia glótica de 160Hz aproximadamente.

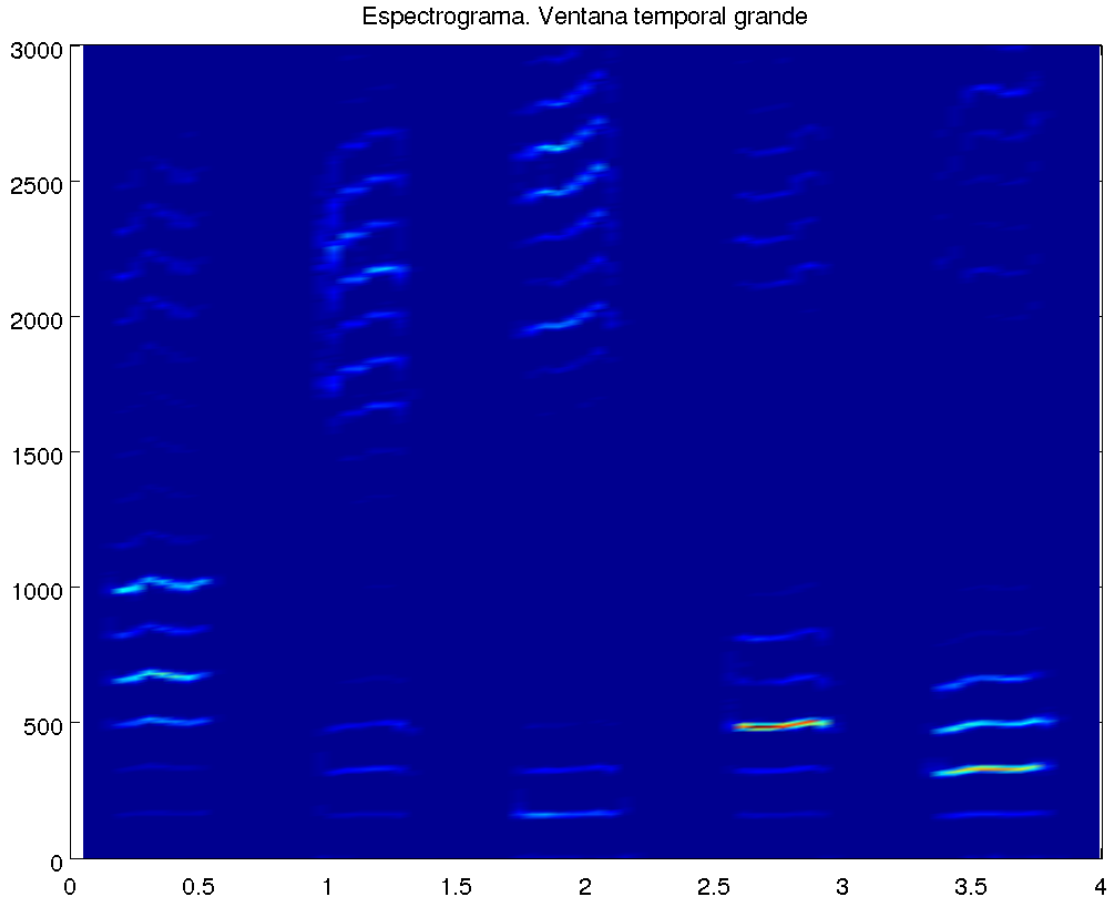

Figura 2: Vocales. DFT de 2048 puntos. Ventana de igual tamaño.

Aquí se puede ver que la fuente del sistema no es exactamente un tren de pulsos periódico constante en el tiempo. Por el contrario, se puede ver como a medida que avanza el tiempo las líneas que representan los pulsos se deforman, posiblemente por una variación en la frecuencia de dicho tren.

Por otro lado, para focalizarse en la envolvente del espectro y ver donde se encuentran los formantes, se eligió una ventana más angosta, de 256 puntos, obteniéndose el siguiente gráfico.

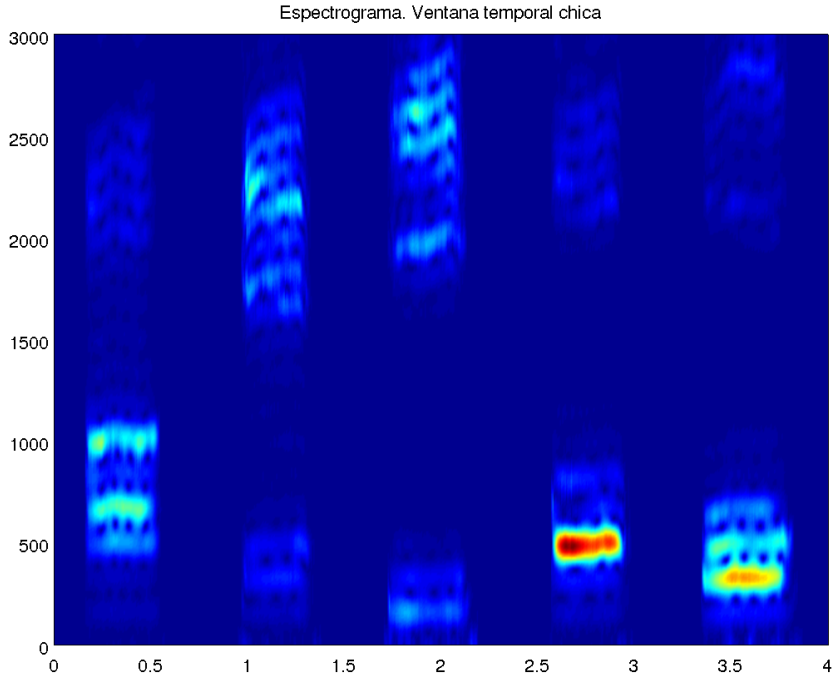

Figura 3: Vocales. DFT de 2048 puntos. Ventana de tamaño 256.

Utilizando el programa *WaveSurfer* para detectar mediante un algoritmo los formantes de cada vocal y comparando el resultado con la figura anterior se obtuvo la siguiente tabla:

Vocal	Espectrograma			*WaveSurfer*		
	F1	F2	F3	F1	F2	F3
/a/	700	1000	2200	620	1017	2097
/e/	450	1800	2300	402	1783	2204
/i/	200	2000	2700	235	1960	2500
/o/	480	870	2400	490	2200	2587
/u/	330	600	2200	356	2200	2800

Cuadro 1: Formantes de las vocales según el espectrograma realizado en MATLAB. (en Hz)

Se puede apreciar como los fonemas /o/ y /u/ presentan más dificultades

2.3. Segmentar el archivo separando cada una de las vocales y graficar su espectro para distinguir mejor los formantes.

En este caso se separó cada una de las vocales en archivos distintos. Luego se graficó la DFT de cada señal multiplicada por una ventana cuadrada. El objetivo de esto es conocer la envolvente del espectro para cada vocal (sabiendo que el espectro real es un tren de pulsos modulado), con lo cual se jugó con los valores del largo de la ventana y de la cantidad de puntos de DFT para visualizarla mejor. Además, se buscó, a través de un *offset*, en que instante ventanear para evidenciar mejor los formantes. Al disminuir el largo temporal de la ventana aplicada, el ancho de la *sinc* que convoluciona en frecuencia con el espectro es mayor, prácticamente interpolando de esta manera los picos del tren de pulsos, es decir, permitiendo que el resultado se enfoque en la envolvente. Por otro

lado, la cantidad de puntos de la DFT, determina la resolución en frecuencia, i.e. la cantidad de muestras tomadas del espectro convolucionado.

Adicionalmente, se agregó un algoritmo muy breve con el cual detectar los picos más altos de cada espectro, que en principio serían los mayores formantes. Esta rutina en primer lugar detecta el máximo absoluto del espectro (variable M, ver código "ejercicio03.m"). Luego decide si el pico es formante o no según el valor del pico sea mayor o menor que un umbral establecido en base al máximo absoluto M. En este caso se eligió un umbral de 0.22M.

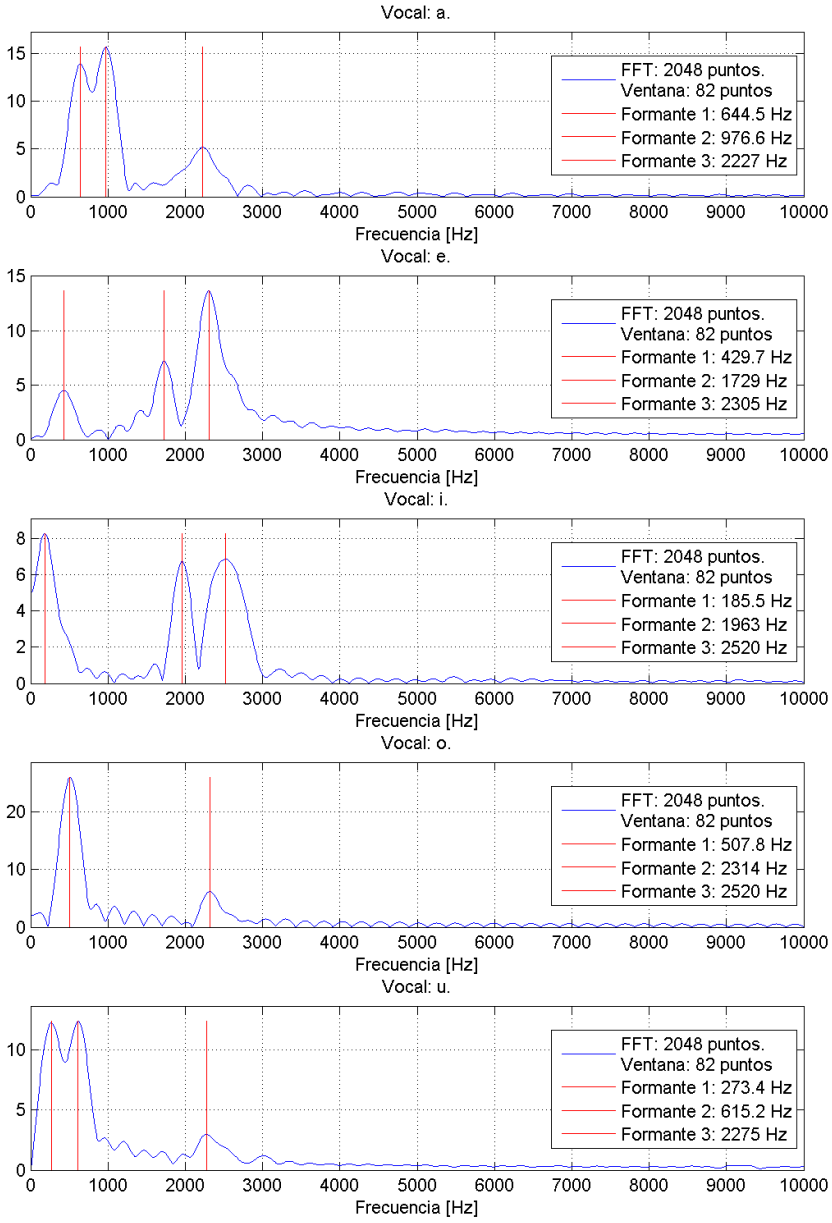

Figura 4: Vocales separadas.

2.4. Para una frecuencia de resonancia de 1000Hz graficar en un mismo gráfico el módulo de la transferencia en función del factor de amortiguamiento. Tomar 5 valores en el rango $0 < \xi < 0{,}6$. Determinar el valor de frecuencia angular ω_m para el cual la transferencia presenta un máximo. Comprobar que la relación entre la frecuencia máxima ω_m y ω_r está dada por $\omega_m = \omega_r\sqrt{1-2\xi^2}$.

La función de transferencia de los sistemas pasabajos de segundo orden está dada por:

$$H(s) = \frac{\omega_r^2}{s^2 + 2\xi\omega_r s + \omega_r^2}$$

Aplicando logaritmo en base 10 y multiplicando por 20 a $|H(s = j\omega)|$, se obtuvo el siguiente gráfico:

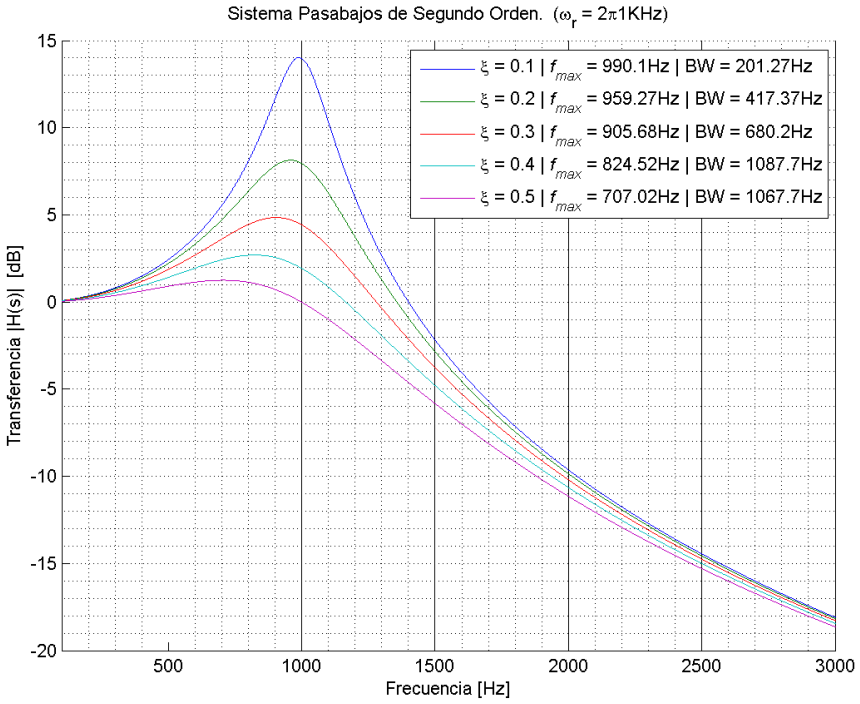

Figura 5: Módulo de la transferencia en función del factor de amortiguamiento ξ.

Teóricamente, los máximos de las transferencias para cada ξ según la formula indicada para calcular ω_m dan:

ξ	f_m teórico	f_m del gráfico
0.1	989.95	990.1
0.2	959.17	959.27
0.3	905.54	905.68
0.4	824.62	824.52
0.5	707.11	707.02

Cuadro 2: Frecuencia máxima teórica, en Hz.

2.5. Determinar B_w, el ancho de banda a -3dB con respecto al valor máximo y graficarlo en función de ξ.

La respuesta al presente ejercicio se encuentra en el gráfico anterior. El ancho de banda es el largo del conjunto de frecuencias cuya atenuación es menor o igual a 3dB respecto del máximo de amplificación. Por lo tanto, para hallar este valor en primer lugar se buscó el máximo de la función (M) y luego toda la banda de frecuencias que se encuentran atenuadas a menos de 3dB respecto a este (osea amplificadas en más de M-3dB). El ancho de banda será la diferencia absoluta entre los límites de dicho conjunto.

Cabe destacar que en los casos en que el pico de la función es menor a 3dB de amplificación (que corresponden a $\xi = 0{,}4\, y\, 0{,}5$ del gráfico anterior) la definición de ancho de banda no está correctamente especificada. En estos casos se decidió utilizar el mismo algoritmo, donde el ancho de banda quedará determinado por la única frecuencia, a derecha del máximo, que atenúa en 3dB respecto a este. No obstante, para el trabajo de formantes los valores de ξ son siempre menores a 0.2 evitando este problema.

2.6. Utilizando los valores de las tablas anteriores y teniendo en cuenta lo hallado en los puntos 4 y 5, encontrar mediante aproximaciones sucesivas los parámetros de las 25 transferencias.

Para obtener los parámetros se utilizó la función del *script* "get_filter_parameters.m". En este código lo que se hace es armar un vector de frecuencia centrado en la frecuencia máxima pedida y probar con diferentes valores de ξ hasta encontrar uno que se aproxime a los parámetros pedidos según cierta tolerancia (en este caso se utilizó una tolerancia del 1 %). El valor de la frecuencia de resonancia a probar en cada iteración queda determinado por el ξ actual y el $\omega_m$pedido:

$$\omega_{ri} = \frac{\omega_m}{\sqrt{1 - 2\xi_i^2}}$$

según la ecuación que relaciona $\omega_r$con ω_m vista en el ejercicio 3.

Los parámetros obtenidos se encuentran detallados en las siguientes tablas:

Vocales	f_{r1}	f_{r2}	f_{r3}	f_{r4}	f_{r5}
/a/	705.83	1221	2602.4	3304.7	3752.6
/e/	455.32	1800.5	2501.2	3304.7	3752.6
/i/	311.59	2024.9	2973.2	3304.7	3752.6
/o/	473.3	1101.1	2400.5	3304.7	3752.6
/u/	302.02	870.86	2241.1	3304.7	3752.6

Cuadro 3: Frecuencias de resonancia de cada filtro para las vocales, en Hz.

Vocales	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
/a/	0.090698	0.028482	0.030603	0.037532	0.026479
/e/	0.10773	0.016547	0.021875	0.037532	0.026479
/i/	0.071339	0.048941	0.066483	0.037532	0.026479
/o/	0.08337	0.031565	0.014535	0.037532	0.026479
/u/	0.081633	0.03141	0.022181	0.037532	0.026479

Cuadro 4: Factor de amortiguamiento para cada filtro.

2.7. Generar y graficar la señal del tren de pulsos glóticos para una frecuencia glótica de 150 Hz.

Para modelar el tren de pulsos glóticos se va a utilizar el siguiente filtro pasabajos de segundo orden, el cual es excitado por un tren de impulsos cuyo período es la inversa de la frecuencia glótica.

$$H_g(s) = \frac{7{,}14 \cdot 10^6}{s^2 + 4300s + 7{,}14 \cdot 10^6}$$

Para generar la señal del tren de pulsos glóticos, se debe filtrar un tren de pulsos con Hg. En una primera aproximación, para implementar la solución a este problema, se buscó la transformada inversa de Laplace del filtro $H_g(s)$, obteniendo la respuesta impulsiva $h_g(t)$. Luego se convolucionó esta señal con un tren de pulsos con una frecuencia de 150Hz. Para simularlo en MATLAB, se muestreo el tiempo a 40KHz y se modelaron las deltas de *Dirac* con impulsos de amplitud unitaria. Para mantener los valores correctos de amplitud se multiplicó el filtro por Ts, el período de muestreo. Esto se puede ver claramente recordando las propiedades de la transformada de Fourier:

$$x(t) \cdot p(t) \xrightarrow{\mathcal{F}} \frac{1}{2\pi} X(\omega) \cdot P(\omega)$$

siendo $p(t)$ el tren de pulsos de período T_s que muestrea a la señal $x(t)$. Y como $P(\omega)$ contiene un factor igual a $\frac{2\pi}{T_s}$, 2πse anula y queda $\frac{1}{T_s}$ que es la frecuencia de muestreo.

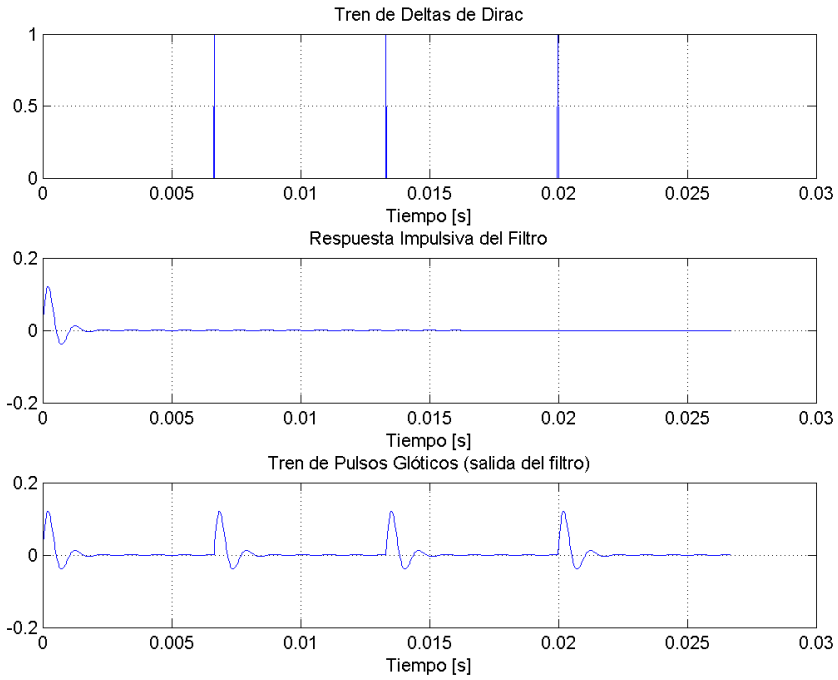

Figura 6: Señal del tren de pulsos glóticos según el modelo de filtro pasabajos H_g.

2.8. A partir de los 5 pasabajos analógicos que modelan el tracto vocal obtener los respectivos pasabajos discretos para una frecuencia de muestreo de 16KHz. Justificar el método de conversión elegido. Graficar para una de las vocales la respuesta en frecuencia de los 5 resonadores por separado y la respuesta del sistema conectado en cascada. En caso de haber utilizado el método de transformación bilineal, en el gráfico de cada resonador dibujar simultáneamente la respuesta en frecuencia con y sin prewarping.

Teorema de Muestreo

En este caso se buscó la solución general para discretizar un filtro pasabajos analógico de 2° orden. Se podría optar por un mapeo continuo-digital basado en el teorema de muestreo, dado que si bien el filtro no es estrictamente de banda limitada, se puede considerar que si lo es dada la gran atenuación en alta frecuencia. Muestreando a una frecuencia de 16KHz, se debería esperar que entre 8KHz y 16KHz la respuesta en frecuencia del filtro original continuo sea despreciable frente a la amplitud de 8KHz para abajo. Dado que la máxima frecuencia de resonancia utilizada en el trabajo práctico será de menos de 4KHz, esto se cumplirá siempre como se ve en la siguiente figura donde se graficó para el caso límite.

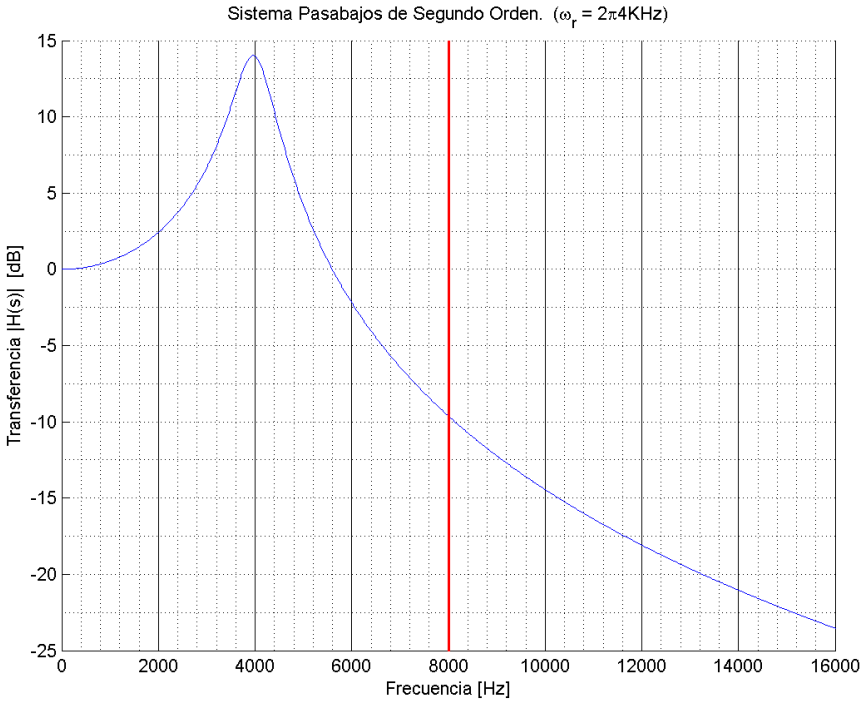

Figura 7: Atenuación entre 8KHz y 16KHz mayor a 10dB.

En primer lugar se buscó la respuesta al impulso del filtro antitransformando en Laplace H(s).

$$H(s) \xrightarrow{\mathcal{L}^{-1}} h(t) = \alpha e^{-\beta t} \sin(\gamma t)$$

con $\alpha = \frac{\omega_r}{\sqrt{1-\xi^2}}$, $\beta = \omega_r \xi$ y $\gamma = \omega_r \sqrt{1-\xi^2}$

Para obtener $h(n)$ se muestreó $h(t)$ como se explicó anteriormente. Luego se calculó la transformada Z de dicha función, la cual es muy común y se encuentra en tablas:

$$h(n) = \alpha e^{-\beta n T_s} \sin(\gamma n T_s) T_s$$

$$h(n) \xrightarrow{\mathcal{Z}} H(Z) = \frac{A \cdot Z^{-1}}{1 + B \cdot Z^{-1} + C \cdot Z^{-2}}$$

con $A = \alpha e^{-\beta T_s} \cdot \sin(\gamma T_s) T_s$, $B = -2e^{-\beta T_s} \cdot \cos(\gamma T_s)$ y $C = e^{-2\beta T_s}$
Finalmente teniendo en cuenta que

$$H(Z) = \frac{Y(Z)}{X(Z)}$$

Se obtiene:

$$Y(Z) = -B \cdot Z^{-1} Y(Z) - C \cdot Z^{-2} Y(Z) + A \cdot Z^{-1} X(Z)$$

Luego, antitransformando en Z se llega a la ecuación en diferencias que será sumamente útil para implementar el filtro:

$$y(n) = -B \cdot y(n-1) - C \cdot y(n-2) + A \cdot x(n-1)$$

No obstante, este método siempre tendrá algo de *aliasing*, a pesar de lo explicado en el primer párrafo.

Transformación Bilineal

Otro método para obtener un filtro discreto a partir de uno continuo es el de la Transformación Bilineal. Sabiendo que

$$s = 2F_s \left(\frac{1 - Z^{-1}}{1 + Z^{-1}} \right)$$

se evaluó $H(s)$ con ese reemplazo obteniéndose:

$$H_d(Z) = H_c \left(s = 2F_s \left(\frac{1 - Z^{-1}}{1 + Z^{-1}} \right) \right)$$

$$H_d(Z) = \frac{\omega_r^2}{\left(2F_s \left(\frac{1-Z^{-1}}{1+Z^{-1}} \right) \right)^2 + 2\xi\omega_r \left(2F_s \left(\frac{1-Z^{-1}}{1+Z^{-1}} \right) \right) + \omega_r^2}$$

$$H_d(Z) = \frac{B_0 + B_1 Z^{-1} + B_2 Z^{-2} + B_3 Z^{-3}}{A_0 + A_1 Z^{-1} + A_2 Z^{-2} + A_3 Z^{-3}}$$

siendo

$$\begin{array}{ll} A_0 = 4F_s^2 + 4F_s\xi\omega_r + \omega_r^2 & B_0 = \omega_r^2 \\ A_1 = -4F_s^2 + 4F_s\xi\omega_r + 3\omega_r^2 & B_1 = 3\omega_r^2 \\ A_2 = -4F_s^2 - 4F_s\xi\omega_r + 3\omega_r^2 & B_2 = 3\omega_r^2 \\ A_3 = 4F_s^2 - 4F_s\xi\omega_r + \omega_r^2 & B_3 = \omega_r^2 \end{array}$$

Finalmente teniendo en cuenta que

$$H(Z) = \frac{Y(Z)}{X(Z)}$$

Se obtiene:

$$A_0 Y(Z) + A_1 \cdot Z^{-1} Y(Z) + A_2 \cdot Z^{-2} Y(Z) + A_3 \cdot Z^{-3} Y(Z) = B_0 X(Z) + B_1 \cdot Z^{-1} X(Z) + B_2 \cdot Z^{-2} X(Z) + B_3 \cdot Z^{-3} X(Z)$$

Luego, antitransformando en Z se llega a la ecuación en diferencias que se utilizará para implementar el filtro:

$$\sum_{i=0}^{3} A_i y(n-i) = \sum_{i=0}^{3} B_i x(n-i)$$

Por otro lado, esta transformación implica una relación no lineal entre la frecuencia en tiempo continuo (ω) y la frecuencia en tiempo discreto (Ω) dada por

$$\omega = 2F_s \tan\left(\frac{\Omega}{2}\right)$$

$$\Omega = 2 \arctan\left(\frac{\omega}{2F_s}\right)$$

en consecuencia, es de esperar que en altas frecuencias el espectro tienda a comprimirse (fenómeno conocido como *warping*), en cambio con valores de bajas frecuencias se obtiene una relación prácticamente lineal: $\Omega \approx \frac{\omega}{F_s}$.

Para evitar dicho problema es posible utilizar una técnica de *pre-warping* que consiste en anticiparse a dicha compresión cambiando los parámetros del filtro continuo original para conservar el cumplimiento de la especificación en tiempo discreto. En nuestro sistema pasabajos de segundo orden, esto significa desplazar el máximo y modificar el ancho de banda ligeramente. Como se quiere obtener en tiempo discreto un máximo (Ω_m) tal que

$$\Omega_m = \frac{\omega_m}{F_s}$$

Como

$$\Omega_m = 2 \arctan\left(\frac{\omega'_m}{2F_s}\right)$$

entonces, en tiempo continuo, el máximo (ω'_m) que se le pide al filtro para cumplirlo debe ser

$$\omega'_m = 2F_s \tan\left(\frac{\Omega_m}{2}\right)$$

$$\omega'_m = 2F_s \tan\left(\frac{\omega_m}{2F_s}\right)$$

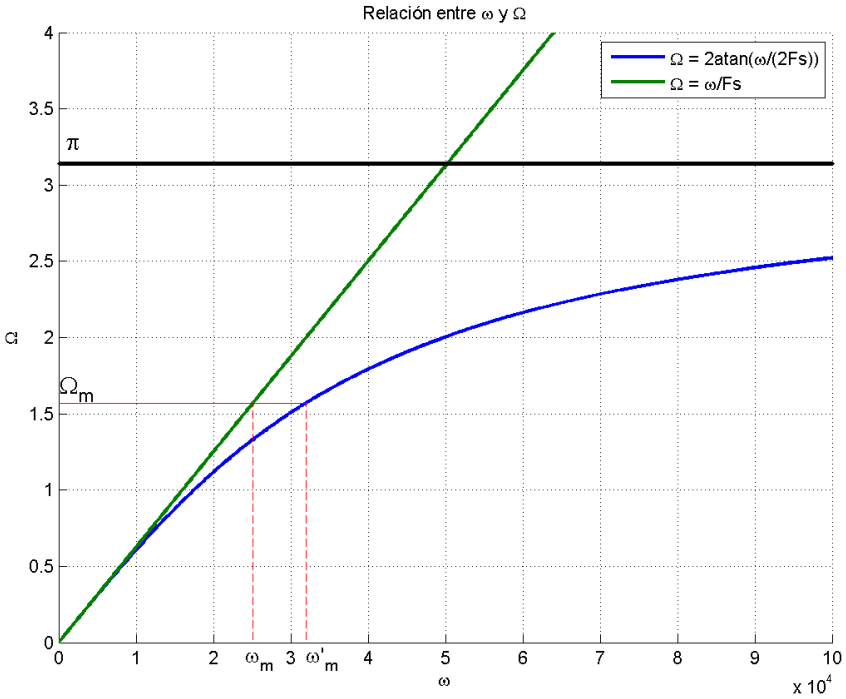

Figura 8: *Warping* en Transformación Bilineal. Frecuencia de muestreo de 16KHz.

Como el parámetro que se le pasa al filtro es en realidad ω_r, para los alcances de este trabajo el *pre-warping* se realizará en realidad en dicha frecuencia. Aumentar esta parámetro modificará también en consecuencia el valor del ancho de banda resultante dado que se mantiene fijo ξ. Lo más correcto sería tener en cuenta estas relaciones y modificar tanto ω_r como ξ antes de entrar al filtro. En una primera aproximación, se ignoró dicho problema creyendo que la diferencia entre ω_m y ω'_m sería pequeña en relación a la variabilidad del ancho de banda por mantener fijo ξ, entonces la diferencia en $B_\omega$sería despreciable. No obstante, al graficar la respuesta en frecuencia de cada filtro por separado, se encontró que la diferencia obtenida en el ancho de banda era notablemente considerable, incluso para pequeños valores de ω_r. Por esta razón se decidió aumentar ξ en proporción a la relación entre ω_r y ω'_r con la intención de opacar el problema, esto es:

$$\xi' = \xi \left(\frac{\omega'_r}{\omega_r} \right)^2$$

De esta manera se obtuvieron los siguientes resultados (se muestra el caso de la vocal “a”):

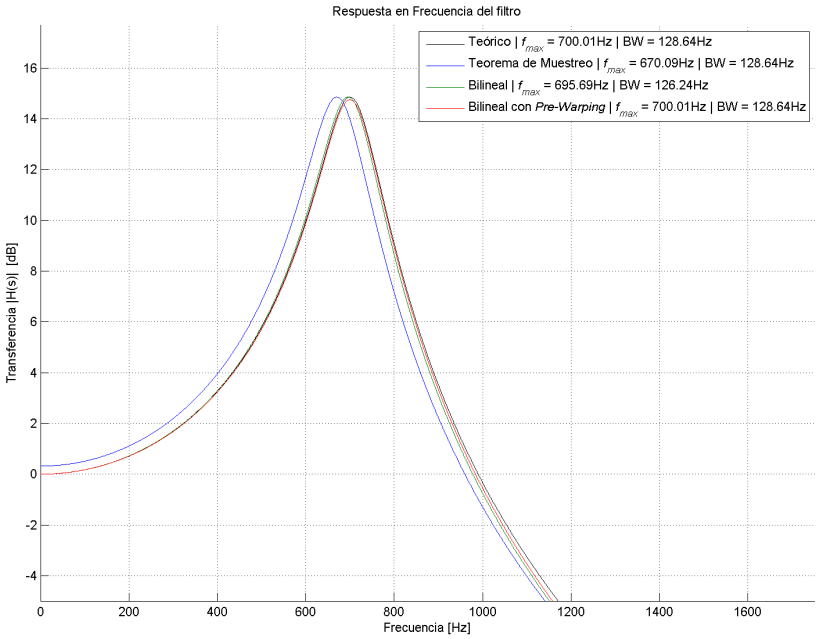

Figura 9: Filtro 1

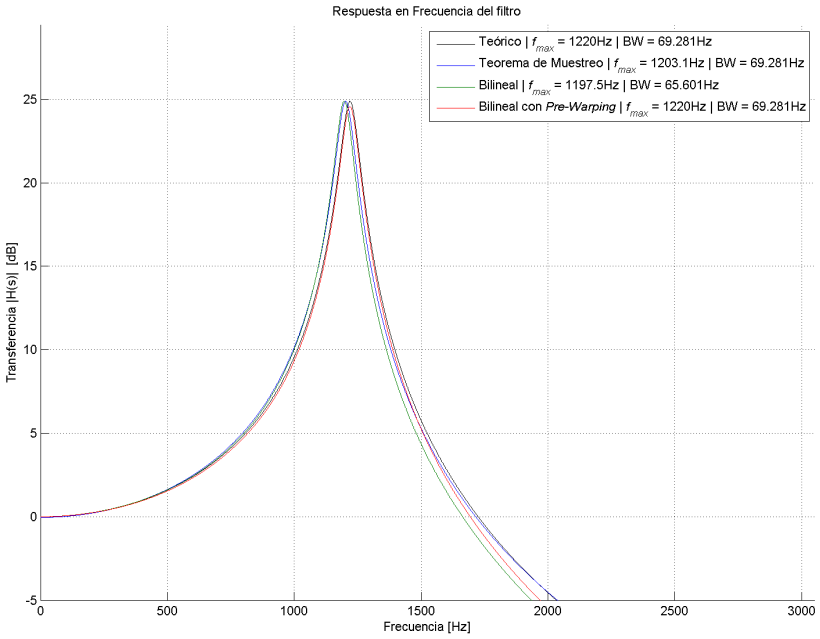

Figura 10: Filtro 2

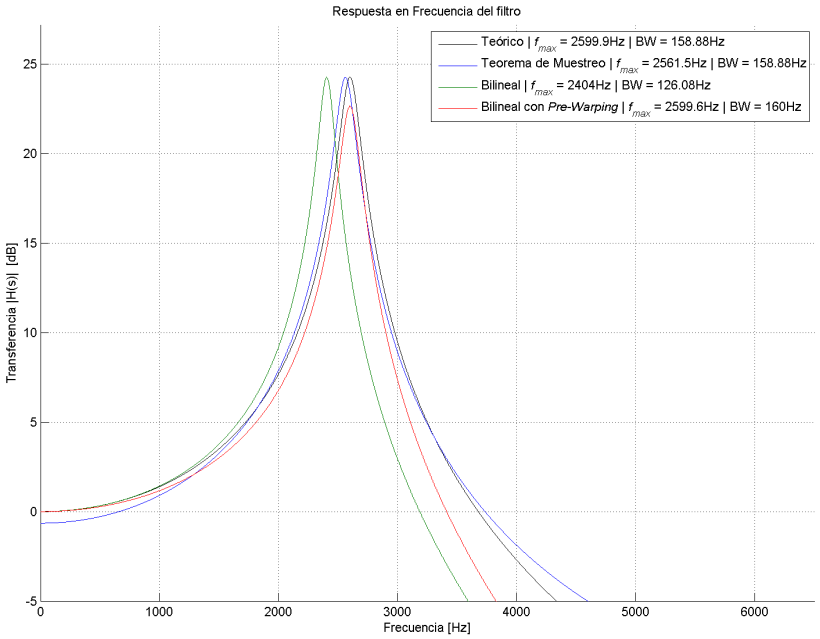

Figura 11: Filtro 3

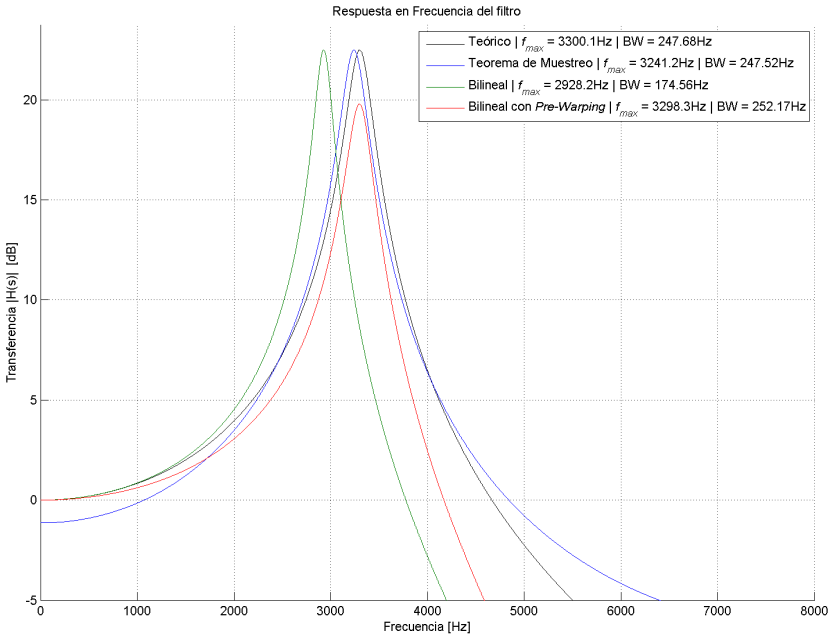

Figura 12: Filtro 4

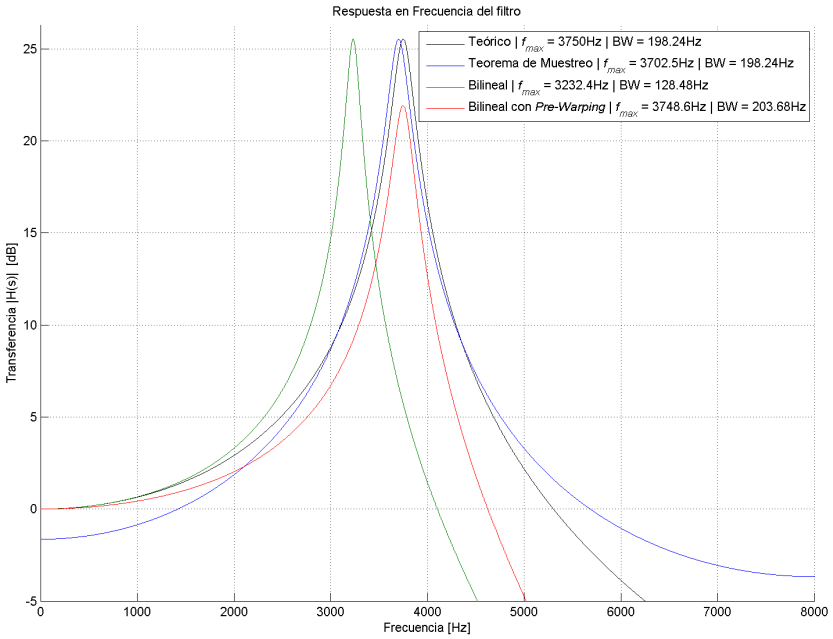

Figura 13: Filtro 5

La comparación de los resultados para los diferentes métodos se puede apreciar en las siguientes tablas:

		F1	F2	F3	F4	F5
Teórico (≈Especificación)[2]		700.01	1220	2599.9	3300.1	3750
Teorema de Muestreo		670.09	1203.1	2561.5	3241.2	3702.5
Transformación Bilineal	Sin *Pre-Warping*	695.69	1197.5	2404	2928.2	3232.4
	Con *Pre-Warping*	700.01	1220	2599.6	3298.3	3748.6

Cuadro 5: Frecuencia máxima, en Hz.

Aquí se puede ver como el peor filtro, en lo que respecta a frecuencia máxima, es el Bilineal sin *pre-warping*. Mientras que el mejor en este sentido es aquél con *pre-warping*.

		Bw1	Bw2	Bw3	Bw4	Bw5
Teórico (≈Especificación)[3]		128.64	69.281	158.88	247.68	198.24
Teorema de Muestreo		128.64	69.281	158.88	247.52	198.24
Transformación Bilineal	Sin *Pre-Warping*	126.24	65.601	126.08	174.56	128.48
	Con *Pre-Warping*	128.64	69.281	160	252.17	203.68

Cuadro 6: Ancho de Banda, en Hz.

Por otro lado, en cuanto al ancho de banda obtenido, se ve que el peor es nuevamente el Bilineal sin *pre-warping*. En este caso, el mejor es el basado en el Teorema de Muestreo. Sin embargo, con la Transformación Bilineal utilizando el *pre-warping* conjuntamente para ω_r y ξ, se obtienen anchos de bandas muy próximos a los requeridos por la especificación.

A continuación se muestra la respuesta en frecuencia al sistema completo con los 5 filtros en cascada.

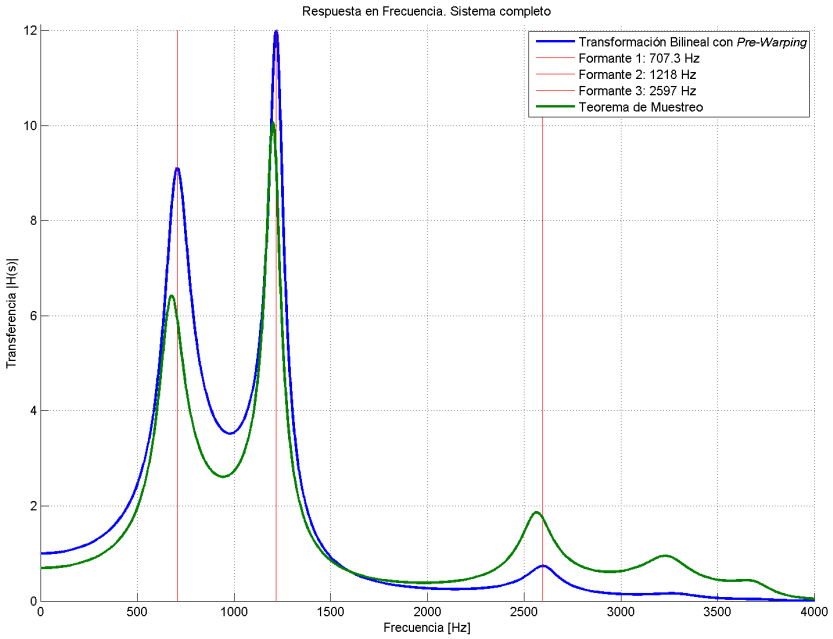

Figura 14: Sistema completo. 5 filtros en cascada.

Se puede ver que las respuestas son similares. Por un lado, con el método del Muestreo se obtiene algo mas de ganancia en los formantes de frecuencia más alta pero menos en los primeros formantes. Lo contrario sucede con la Transformación Bilineal, donde apenas se distinguen los últimos formantes, dada la alta atenuación de los otros filtros, acentuada por la compresión inherente. En conclusión, dado que la consigna sólo pide la ubicación y el ancho de banda y no especifica amplitudes para cada formante, la Transformación Bilineal con *Pre-Warping* se adecua mejor a este trabajo.

2.9. Obtener el filtro discreto que simula el tren de pulsos glóticos para una frecuencia de muestreo de 16 KHz. Justificar el método de conversión utilizado. Generar y graficar la secuencia de pulsos glóticos y compararla con la obtenida analógicamente.

Para este caso se utilizó nuevamente la formula genérica y los filtros discretos hallados en el punto 8 con los parámetros de $H_g(s)$ y una frecuencia de muestreo de 16 KHz.

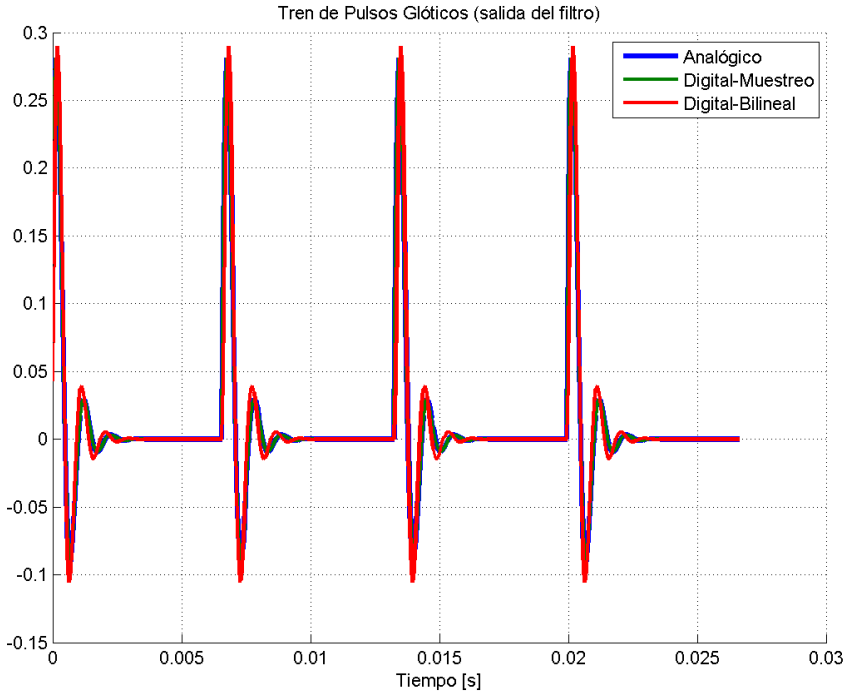

Figura 15: Salida del filtro $h_g(t)$ y $h_g(n)$.

Se puede ver que en el dominio del tiempo el filtro por Muestreo es el más fiel.

2.10. Sintetizar 300ms de cada una de las vocales, utilizando la cascada de los 6 sistemas excitados por el tren de pulsos generado en el punto 9.

Para imitar la modulación de frecuencia del tren de pulsos fuente del sistema, se varió el período de esta señal según un seno con la siguiente fórmula:

$$F_g = F_g + 3\sin\left(\frac{2\pi}{4T}t\right)$$

donde F_g es la frecuencia glótica y T es la duración total del sonido. Al utilizar un período 4 veces mayor a la duración total, lo único que se tendrá es el primer cuarto del seno, equivalente a una leve entonación ascendente de la voz.

Escuchar:
"./Sounds/test_vocales_bi_pre_fsweep_labios.wav".
"./Sounds/test_vocales_bi_pre_fsweep.wav". (sin filtro labial)
"./Sounds/test_vocales_muestreo_fsweep_labios.wav".
"./Sounds/test_vocales_muestreo_fsweep.wav". (sin filtro labial)

2.11. Realizar el análisis espectral de las vocales sintéticas concatenadas entre sí, de modo de formar la misma secuencia "a-e-i-o-u". ¿Qué diferencias encuentra entre esta señal y la del punto 2? ¿Qué diferencia encuentra al escucharlas? ¿Coinciden los picos con los de la Tabla I? ¿Coinciden con los de las señales reales?

Utilizando el mismo código que en el ejercicio 2 se obtuvieron los siguientes gráficos para visualizar frecuencia glótica y envolvente respectivamente:

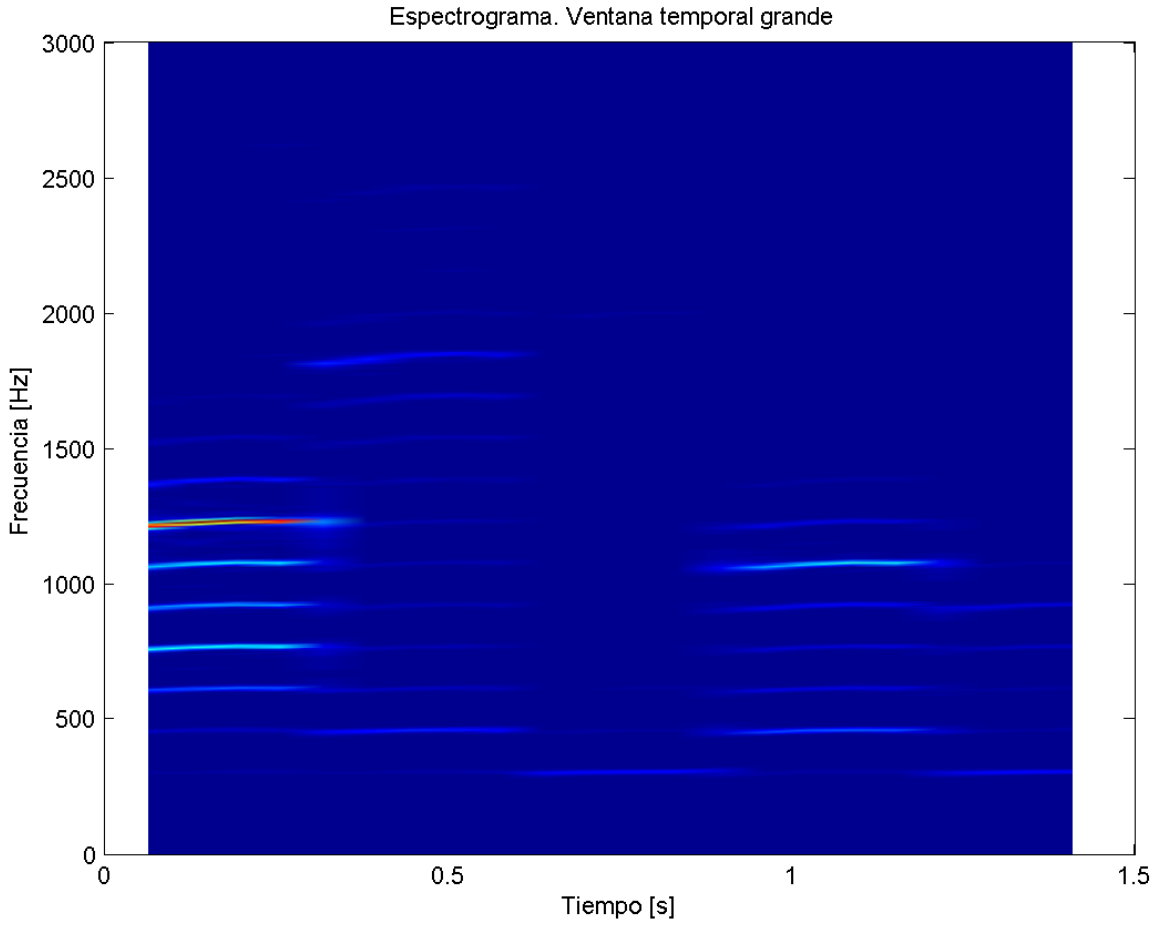

Figura 16: Vocales. DFT de 2048 puntos. Ventana de igual tamaño.

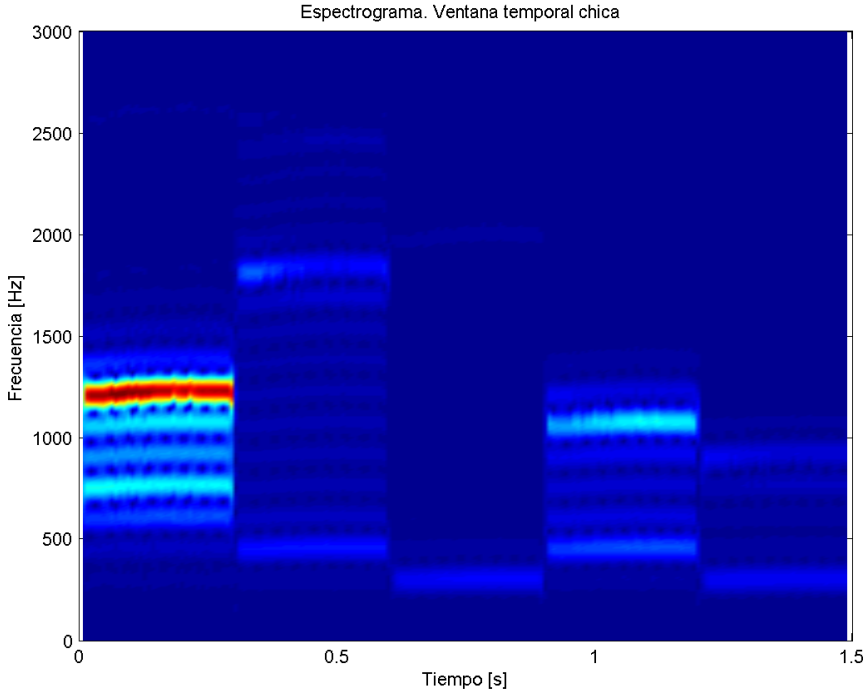

Figura 17: Vocales. DFT de 2048 puntos. Ventana de tamaño 256.

En principio, se puede notar la entonación provocada por el seno, explicada anteriormente.

Por otro lado, también se encuentran diferencias en las amplitudes de cada formante, dado que en ningún momento eso se tuvo en consideración al poner los filtros en cascada y en principio no se sabe si hace a la distinción del fonema.

Al escuchar las distintas versiones, la sintetizada en este trabajo se siente más robótica y antinatural, posiblemente debido a la simplificación del modelo.

Respecto a la tabla de frecuencias máximas y anchos de banda presentada en el modelo, se puede ver que las vocales sintetizadas cumplen ampliamente con la especificación. Comparando con las tablas 3 y 4 y la figura 3 se puede ver que los valores de las formantes son ligeramente distintos pero aproximadamente se ubican en la misma zona como se vio en las tablas 5 y 6.

2.12. Grabar una vocal con su propia voz y calcular aproximadamente la frecuencia máxima y el ancho de banda de cada formante. Utilizando los valores obtenidos sintetizar la vocal mediante el modelo desarrollado en este proyecto pero utilizando como frecuencia glótica la de su propia voz. Discutir los resultados.

A partir del archivo "./Sounds/mis_Vocales.wav" se realizaron las mismas rutinas que en los ejercicios 2 y 3 para distinguir los formantes y la frecuencia glótica de los fonemas en la grabación. En el espectrograma que utiliza una ventana pequeña para focalizarse en la envolvente, se aplicó el logaritmo en base 10 al espectro para ver mejor los formantes de más alta frecuencia, ya que de otro modo no se distinguían.

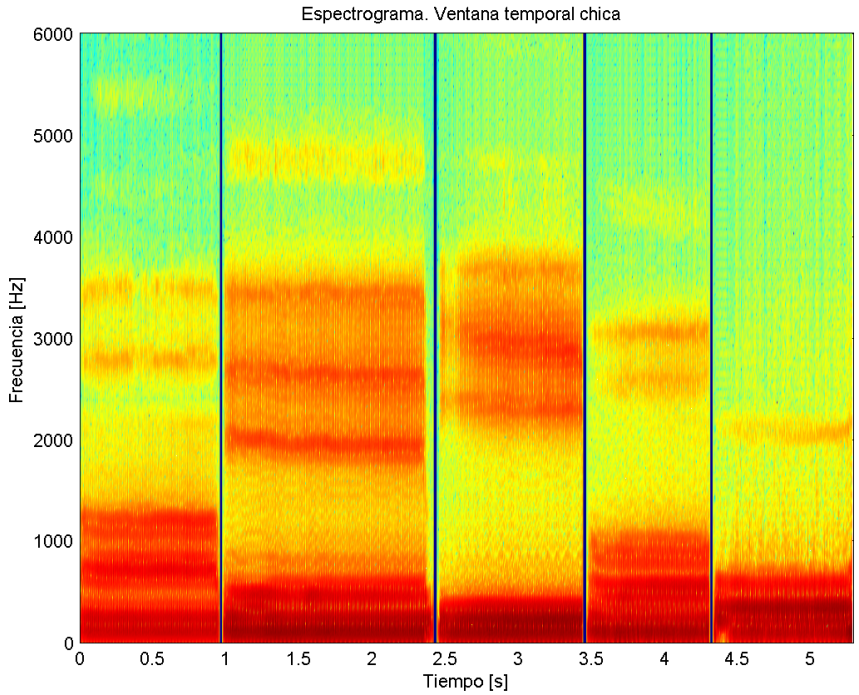

Figura 18: Vocales grabadas. DFT de 2048 puntos. Ventana de tamaño 450.

Luego para apreciar la frecuencia glótica se utilzó una ventana más ancha

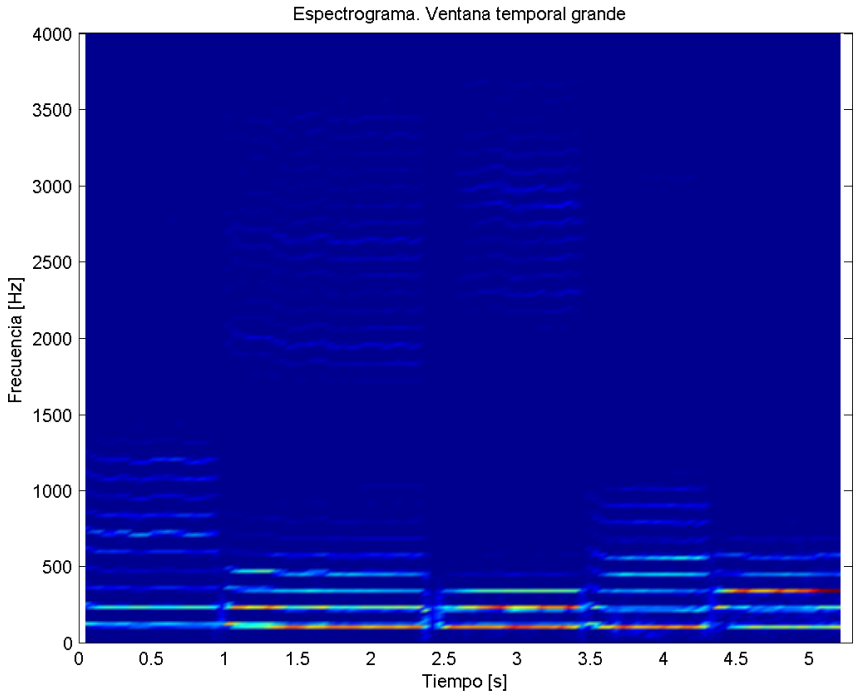

Figura 19: Vocales grabadas. DFT de 2048 puntos. Ventana de tamaño 2048.

Midiendo entre 2 lineas verticales, se ve que la frecuencia del tren de pulsos fuente del sistema es aproximadamente de 120Hz.

También se segmentó el archivo como en el ejercicio 3, y se graficó el espectro aproximando ubicación de formantes y anchos de banda respectivos. En este caso se optó por una escala en dB para apreciar mejor los formantes de alta frecuencia. Además, se buscó, a través de un *offset*, en que instante ventanear para evidenciarlos mejor.

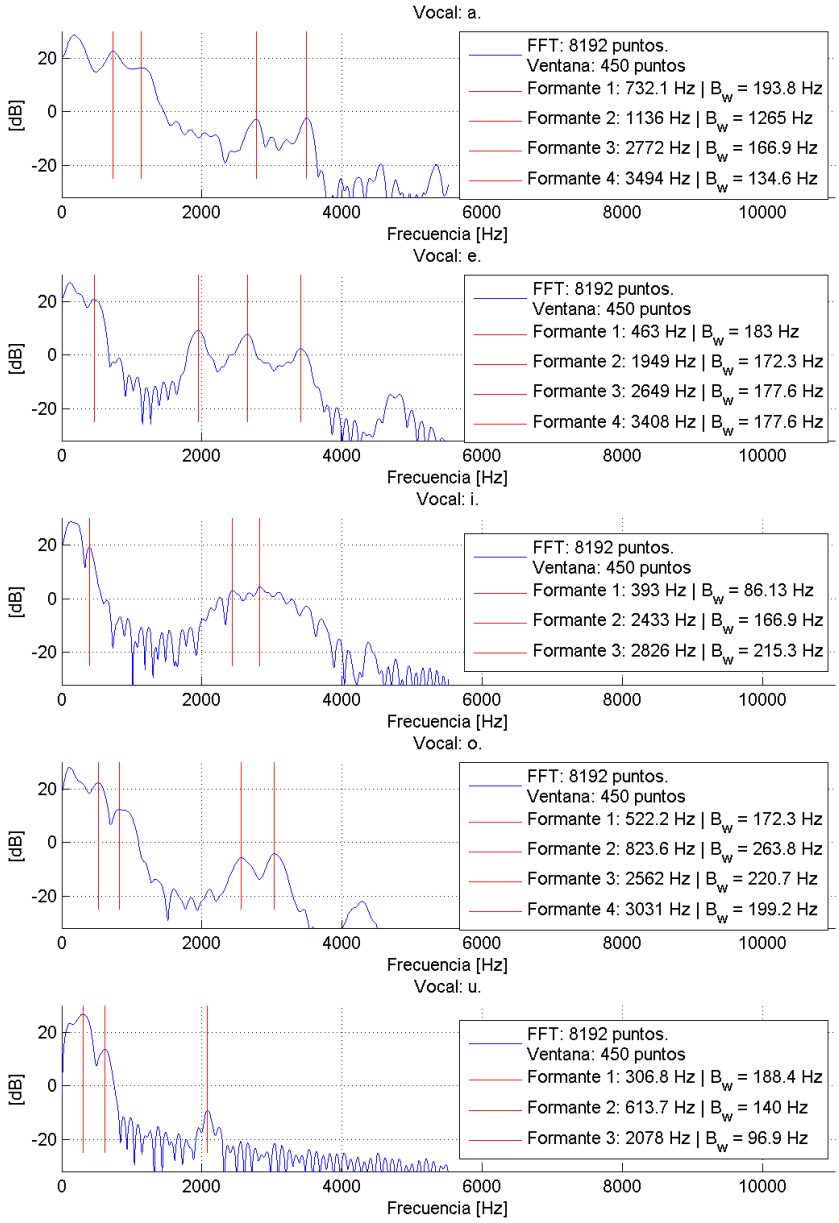

Figura 20: Vocales separadas. Formantes y Ancho de Banda aproximado.

En todos los casos se encontró que el primer pico se hallaba en el mismo lugar, aproximadamente 120Hz. Este fue descartado a la hora de construir los resonadores, dado que se lo consideró producto de la frecuencia glótica y no un formante.

Para hallar el ancho de banda, se buscó, en cada máximo local, el punto más cercano que bajaba 3dB a cada

lado.

Como el algoritmo detector no encontró los formantes de frecuencias más altas se decidió añadir aquellos picos que se pueden ver con ayuda del espectrograma de la figura 18 y de los espectros de la figura 20. Estos valores corresponden al último resonador de cada fonema, al cuarto de la "i" y al segundo de la "a" (este último solo con problemas a la hora de estimar el ancho de banda, el cual se aproximó a 200Hz).

Los resultados se encuentran en la siguiente tabla en la que se puede ver los valores del modelo de la especificación para comparar:

Vocal	Especificación					Grabación				
	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
/a/	700	1220	2600	3300	3750	732.13	1135.9	2772.4	3493.8	5420
/e/	450	1800	2500	3300	3750	462.96	1948.8	2648.6	3407.6	4743
/i/	310	2020	2960	3300	3750	392.98	2433.3	2826.2	3700	4377
/o/	470	1100	2400	3300	3750	522.18	823.65	2562.5	3030.8	4274
/u/	300	870	2240	3300	3750	306.85	613.7	2078	-	-

Cuadro 7: Frecuencia máxima, en Hz.

Vocal	Especificación					Grabación				
	Bw1	Bw2	Bw3	Bw4	Bw5	Bw1	Bw2	Bw3	Bw4	Bw5
/a/	130	70	160	250	200	193.8	200	166.88	134.58	113
/e/	100	60	110	250	200	183.03	172.27	177.65	177.65	242
/i/	45	200	400	250	200	86.133	166.88	215.33	200	91
/o/	80	70	70	250	200	172.27	263.78	220.72	199.18	232
/u/	50	55	100	250	200	188.42	139.97	96.899	-	-

Cuadro 8: Ancho de Banda, en Hz.

En el caso del ancho de banda, la similitud es mucho menor. Es un parámetro más difícil de aproximar debido a la dependencia de su visualización con los parámetros de la ventana utilizada al realizar la DFT.

Finalmente, con estos valores se ejecutaron nuevamente los códigos fuente obteniendo un nuevo sonido que se puede escuchar en "./Sounds/mis_vocales_fg120_bi_pre_fsweep_labios.wav" donde se han utilizado todos los filtros. Adicionalmente, se modificó el último filtro (aquél que simula el paso por los labios) de modo que no se pierda tanto el cuerpo de la voz (es decir las bajas frecuencias), quedando:

$$y(n) = x(n) - 0{,}6x(n-1)$$

El resultado se puede oir en "./Sounds/mis_vocales_fg120_bi_pre_fsweep_labios0.6.wav"

En ambos casos el sonido sintetizado es más robótico y nasal. Esto último se cree que debe estar relacionado con la poca precisión con la que se aproximaron los anchos de banda de cada formante.