

Proyecto Especial Señales y Sistemas

Análisis y síntesis de sonidos vocálicos

Objetivo

El objetivo del presente proyecto es la utilización de algunas de las técnicas de análisis y procesamiento de señales y sistemas discretos estudiados en este cuatrimestre para modelizar y sintetizar sonidos vocálicos del habla. La señal de habla es una señal cuyas características de frecuencia varían en función del tiempo, y por lo tanto requiere para su visualización y síntesis herramientas del análisis de Fourier dependiente del tiempo, como por ejemplo la transformada de Fourier de corto tiempo y otras más complejas. En una primera aproximación los sonidos vocálicos aislados pueden ser modelizados por la salida de un sistema LTI alimentado con una entrada adecuada. El objetivo del presente proyecto especial es analizar estas señales y sintetizarlas utilizando los conocimientos adquiridos sobre procesamiento discreto de señales.

La descripción de la síntesis de los sonidos vocálicos está basada en el trabajo original de Klatt, disponible en la hemeroteca de la facultad. En dicho trabajo es posible encontrar detalles, ampliaciones y la metodología para la síntesis de los otros tipos de fonemas no vistos en este TP, así como las implementaciones de las coarticulaciones entre fonemas.

El contenido específico del TP es:

- Modelización de los sonidos vocálicos de la señal de habla mediante cascada de sistemas de segundo orden.
- Implementación de un sintetizador de vocales del castellano.
- Comparación entre las señales generadas y señales de habla reales mediante Transformada de Corto Tiempo

Requisitos de aprobación

El proyecto especial tendrá una fecha límite de vencimiento y su evaluación está establecida en el calendario de la materia, día en el cual el alumno deberá presentarse indefectiblemente con el informe del proyecto en forma impresa. Habrá un rango de fechas anterior a la fecha definitiva de entrega en el cual el alumno podrá hacer una preentrega del proyecto especial en versión electrónica, pdf solamente, en <http://campus.fi.uba.ar/>. Durante ese rango de fechas el docente

puede aconsejar al alumno la revisión de ciertos puntos en el proyecto. Luego del cierre del período de preentrega, se habilitará un período de entrega definitiva, donde el alumno debe depositar su versión pdf del informe y los algoritmos correspondientes. Luego del vencimiento del período de entrega definitivo (anterior a la evaluación del proyecto especial en unos días), no se admitirán más entregas y el alumno que no cumpla este requisito quedará libre. Luego el docente de cada curso evaluará el mismo en tiempo y forma, utilizando la versión electrónica o la impresa que de ninguna manera pueden diferir entre sí. La modalidad de la evaluación se realizará según el docente lo crea conveniente (oral, escrita, el día de la entrega, otro día, etc.), de modo de asegurar el conocimiento del tema desarrollado y la realización individual del trabajo por parte del alumno. El trabajo sólo podrá ser presentado y evaluado en el curso en el cual el alumno se halla inscripto.

La evaluación final puede incluir preguntas sobre:

Ítems particulares sobre los ejercicios de esta guía y su implementación en Matlab/Octave.

Conceptos teóricos necesarios para realizar los ejercicios.

Puede requerirse también al alumno que implemente alguno de los ejercicios similares en la computadora en el momento de la evaluación. Por lo tanto el alumno debe presentarse el día de la evaluación con:

Esta guía.

Las señales utilizadas para el desarrollo del TP en versión electrónica.

Las soluciones a los problemas planteados: Cuando el problema requiera una implementación, la misma debe estar adecuadamente descripta y debidamente justificada. Es decir, si es necesario justificación teórica, ésta debe estar desarrollada. Si se pide una implementación práctica la misma debe estar adecuadamente documentada de modo que el docente pueda constatar que las especificaciones requeridas se cumplen. Esto incluye la presentación del programa de Matlab/Octave utilizado, y los gráficos necesarios para mostrar los resultados obtenidos en formato electrónico e impresos. Se sugiere que el formato electrónico no dependa de que funcione internet para poder verse, para evitar inconvenientes. Todos los gráficos deberán tener título, comentarios en ambos ejes sobre la unidad a representar y el eje de abscisas debe estar en unidades de tiempo o frecuencia según corresponda.

Nota del trabajo práctico especial

El proyecto especial se aprobará con 60 puntos sobre 100. La nota final de cursada se determina ponderando la nota del parcial N_{par} con la nota del proyecto especial N_{pe} del siguiente modo:

$$N_{cur} = 0,6N_{par} + 0,4N_{pe}$$

Como tanto la nota del parcial N_{par} como la nota del proyecto especial N_{pe} deben ser mayores a 60, para la aprobación de la cursada la nota de cursada N_{cur} debe ser mayor que 60 puntos, y se recuerda que la aprobación del Proyecto Especial también es obligatoria para la aprobación de la cursada. Cada ejercicio totalmente bien resuelto tendrá el puntaje que figura en la parte de ejercitación. Cada docente evaluará de acuerdo a esa grilla el trabajo realizado.

Desarrollo del Proyecto Especial

MODELIZACIÓN DE LA SEÑAL DE HABLA

La señal de habla está compuesta de secciones de propiedades cambiantes. Si pensamos a la señal de habla como la salida de un sistema, podemos atribuir dichas variaciones a dos causas:

1. Cambios en la excitación
2. Cambios en la configuración del tracto vocal, es decir en el sistema representado por la transferencia $H(\omega)$. Fig.1

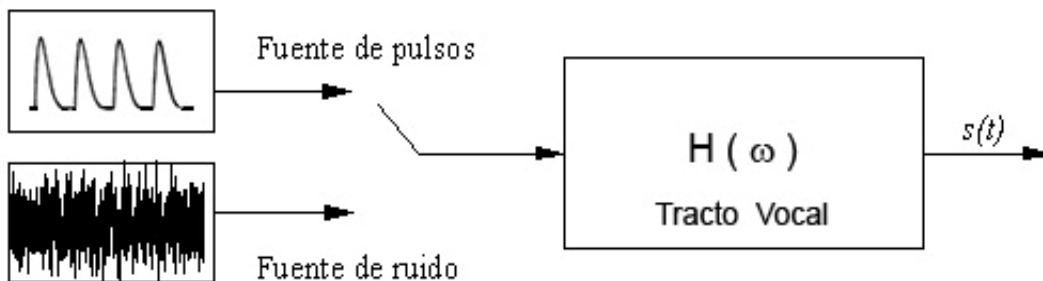


Fig.1 Modelo de generación de la señal de habla

Si la entrada se comporta como un tren de pulsos cuasi-periódicos, la salida será uno de los posibles sonidos vocálicos (/a/,/e/,/i/,/o/,/u/,/m/,/n/,/l/). Si la entrada en cambio es un generador de ruido blanco, el sonido obtenido será un fonema fricativo (/s/,/f/,/sh/). Los sonidos explosivos (/p/,/k/,/t/) en cambio tienen una naturaleza distinta, y son más parecidos a un transitorio que a un sonido estacionario. Se denomina fonema a la mínima unidad sonora que tiene significado en un idioma determinado. Por tal motivo distintas letras del alfabeto pueden estar asociadas al mismo fonema, como en el caso de la “k” y la “q” que tienen el mismo fonema, y a la inversa una letra como la “c” puede tener asociada dos fonemas según opere como fricativa en la palabra “hace” o como explosiva en la palabra “casa”. La distinción entre los fonemas de la misma clase se produce por la forma que va tomando el tracto vocal para cada uno de ellos. La variación de la transferencia del sistema se supone que es suficientemente lenta como para considerar que la señal de habla es la concatenación de porciones de señales que se originan como salida de un sistema LTI. Por esto aparecerán bien representados en un espectrograma.

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

MODELOS LTI PARA SONIDOS VOCÁLICOS

Los sonidos vocálicos se pueden modelizar en primera aproximación como un tren de pulsos cuya característica espectral es modificada por un filtro fijo que caracteriza a cada sonido. La frecuencia del tren de pulsos se conoce como frecuencia glótica. Si trazamos la envolvente del espectro de un sonido vocálico esta tiene varias frecuencias donde su amplitud presenta picos. Estos picos del espectro se denominan formantes. La distinta ubicación en frecuencia de dicho picos o formantes distingue un sonido vocálico de otro.

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

Una buena aproximación usada para síntesis de voz consiste en modelar el tracto vocal por medio de una cascada de cinco resonadores donde los dos últimos son fijos y los tres resonadores primeros varían de acuerdo con el fonema que se quiera emitir. Como resonador en este proyecto usaremos sistemas pasabajos de segundo orden subamortiguados.

SISTEMA PASABAJOS DE SEGUNDO ORDEN

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

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

donde ξ es el factor de amortiguamiento y ω_r la frecuencia angular de resonancia, es decir aquella para la cual la transferencia es máxima cuando el amortiguamiento es nulo. Como se recordará, para valores de $\xi > 1$ el sistema está sobreamortiguado, para $\xi = 1$ el amortiguamiento es crítico y para $0 < \xi < 1$ el sistema es subamortiguado. Este último caso es el que nos ocupa ya que el espectro presenta un sobrepico que es el que va a simular cada uno de los formantes.

- 4. Para una frecuencia de resonancia de 1000 Hz graficar en un mismo gráfico el módulo de la transferencia en función del factor de amortiguamiento. Tomar 5 valores en el rango de $0 < \xi < 0.6$. 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**

$$\omega_m = \omega_r \sqrt{1 - 2\xi^2}$$

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

A continuación se muestran en las tablas siguientes un conjunto de posibles valores de F_m y B_w para cada una de las vocales:

	F1	F2	F3	F4	F5
/a/	700	1220	2600	3300	3750
/e/	450	1800	2500	3300	3750
/i/	310	2020	2960	3300	3750
/o/	470	1100	2400	3300	3750
/u/	300	870	2240	3300	3750

Tabla 1. Frecuencias de los máximos de cada resonador para las vocales, en Hz .

	Bw1	Bw2	Bw3	Bw4	Bw5
/a/	130	70	160	250	200
/e/	100	60	110	250	200
/i/	45	200	400	250	200
/o/	80	70	70	250	200
/u/	50	55	100	250	200

Tabla 2. Anchos de banda de los resonadores, en Hz .

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 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

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

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

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 pre-warping.
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 que la modelización de la señal generada sea más parecida a las señales reales, a los 5 resonadores se les agrega un sexto que simula el efecto de radiación del sonido a la salida de los labios. El modelo es un sistema de primer orden dado por

$$y(n) = x(n) - x(n - 1)$$

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