

J

6.542J, 24.966J, HST.712J CURSO PRÁCTICO EN FISIOLÓGÍA, ACÚSTICA Y PERCEPCIÓN DEL HABLA

Otoño 2001

Práctica 13

30/10/01

Síntesis de voz mediante el uso de un sintetizador por formantes

Libros de consulta

Klatt, D.H., "Software for a Cascade/Parallel Formant Synthesizer", *J. Acoust. Soc. Am.* n° 67, 1980, págs. 971-995.

Klatt, D.H. y L.C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers", *J. Acoust. Soc. Am.*, n° 87, 1990, págs. 820-857 (opcional).

Klatt, D.H. "Description of the cascade/parallel formant synthesizer". **Capítulo 3 de un libro en desarrollo.**

Procedimientos

Los ordenadores se utilizarán para sintetizar una palabra monosilábica inglesa o (incluso mejor) un sintagma, mediante el uso del sintetizador por formantes que se describe en el libro de Klatt (1980) y en el de Klatt y Klatt (1990). Acabar esta síntesis nos llevará dos sesiones de prácticas.

Elija una palabra breve que contenga dos consonantes "interesantes" y haga que un compañero de prácticas la grabe con el micrófono de la sala de ordenadores (o también es posible usar un enunciado ya grabado). Utilice el programa LSPECTO para producir un espectrograma, un trazado de frecuencia fundamental y un trazado de frecuencias por formantes, para medir y tabular los movimientos de frecuencias por formante de cada 25 a 50 mseg, y para medir la frecuencia fundamental. Es posible que quiera conseguir una copia impresa de la frecuencia fundamental y de las frecuencias por formantes. (Los comandos son:

lspecto fichero para el espectrograma *lpr file.ps*

lspecto fichero -f0 para la frecuencia fundamental: *lpr file.ps_f0*

lspecto fichero -pf para las frecuencias por formantes *lpr file.ps_pf*

lspecto file -syn para la copia impresa: *print file.doc*)

Si está sintetizando una voz femenina debería utilizar una velocidad de muestreo de 11500 Hz, con el número de formantes (NF) igual a 5. Para una voz masculina utilice una velocidad de muestreo por defecto de 10000 Hz. (Las razones de por qué hay que proceder de esta manera se explicarán en clase).

Para conseguir que la calidad de la vocal sea correcta, obtenga una copia impresa xkl del espectro en la mitad de una vocal de la grabación. Mediante el uso del sintetizador con una duración corta ($DU=50$), intente ajustar los distintos parámetros glotales y del formante (OQ, TL, AH, F0, AV y los anchos de banda del formante, especialmente B1) para sintetizar un segmento vocálico con un espectro que se ajuste a la vocal emitida. Utilice estos parámetros glotales como base para la síntesis de la vocal o vocales del enunciado.

A continuación, utilice el comando *d* (*draw*) del sintetizador para introducir los valores del parámetro del sintetizador sobre el tiempo. Si un parámetro de control del sintetizador no puede medirse con exactitud a partir de la copia impresa del análisis, (ej. anchos de banda del formante, amplitudes de la fuente relativas), haga uso de la teoría para estimar un valor inicial, y a base de probar y cometer errores, realice comparaciones espectrales de formas de ondas sintéticas y naturales para mejorar los cálculos del parámetro de entrada. Después de 2 ó 3 iteraciones de pruebas y errores, la forma de onda sintética debería ser inteligible e idéntica a la del hablante después de la que se ha modelado. Elabore un espectrograma de su síntesis final.

El informe de prácticas debería contener una breve discusión sobre los problemas que usted se encontró, y debería mostrar los dos espectrogramas, además de ejemplos de espectros muestreados en lugares cruciales dentro de la palabra. Haremos una grabación de algunos de los enunciados originales y sintéticos para reproducirlos en clase.

6.542J, 24.966J, HST.712J CURSO PRÁCTICO EN FISIOLÓGÍA, ACÚSTICA Y PERCEPCIÓN DEL HABLA

Otoño del 2001

Kenneth N. Stevens (36-517, x3209)
stevens@speech.mit.edu

Joseph S. Perkell (36-591, x3223)
perkell@speech.mit.edu

Stefanie Shattuck-Hufnagel (36-523, x3201) stef@speech.mit.edu

Secretario del curso: Arlene Wint (36-511, x7309) wint@speech.mit.edu

Suplemento para la práctica 13

30/10/01

Suplemento para la síntesis de voz mediante un sintetizador por formantes

Sugerencias para la copia de síntesis

1. Comience por fijar la duración del enunciado que va a ser sintetizado (parámetro *du*). Para una voz masculina utilice una velocidad de muestreo por defecto de 11.5 kHz, tanto para la muestra de discurso hablado como para la del sintetizador (parámetro *sr*).
2. Seleccione un punto en la mitad de la vocal. Haga un espectro de la vocal en este punto. Luego, fije los parámetros constantes del sintetizador para la F0, las frecuencias de formantes, OQ, TL y los anchos de banda del formante, para que haya un buen ajuste entre el espectro sintetizado y el espectro del enunciado original.
3. Trace los parámetros para F0, F1, F2, y F3 como una función de tiempo. Dése cuenta de cuando F0 está fijada en cero, no hay fuente de sonoridad. Los parámetros de un formante deben ser continuos y **no** deberían fijarse en cero o en valores extraños. Normalmente, F4 y F5 deberían fijarse como valores constantes. Después de la iteración de cada síntesis, renombre los ficheros *.doc* y *.wav*.
4. Una vez que los formantes y F0 estén acabados, siga con los parámetros de ruido de fricación, ruido de aspiración y nasalización. Le hará falta hacer algún ajuste de espectros entre el enunciado original y el sintetizado.

Cómo utilizar el sintetizador

Entre en la cuenta *labc*.

Escriba *xkl*

Haga clic sobre lo que haya por defecto, o si ya tiene un fichero *.doc*, encuéntrelo.

Hay un conjunto de ajustes por defecto para el sintetizador. Para verlos, escriba *p*.

Para cambiar un valor por defecto (como F0, F1, ...AV...), escriba c

Para dibujar un parámetro de variación de tiempo, escriba d . Introduzca el parámetro que va a cambiar. La función de tiempo del parámetro saldrá visualizada. Se le pedirá que introduzca los números en pares: un valor de tiempo y un valor de parámetro. El gráfico mostrará una función lineal por tramos. Los números deben introducirse secuencialmente.