

6.542J, 24.966J, HST.712J CURSO PRÁCTICO EN FISIOLÓGÍA, ACÚSTICA Y PERCEPCIÓN DEL HABLA**Otoño 2001**

Práctica 15

15/11/01

Síntesis de alto nivel con un sintetizador por formantes usando parámetros quasi articulatorios**Libro de consulta**

Hanson, H.M. y K.N. Stevens, "Control of acoustic source parameters in speech synthesis: A quasi-articulatory approach". 2001. Enviado a *J. Acoust. Soc. Am.*

Introducción

En esta práctica realizará algunos ejercicios de síntesis de voz usando un sintetizador quasi articulatorio llamado HLsyn (para la síntesis de "alto nivel"). Este sintetizador está controlado por un conjunto de 13 parámetros, mucho mejores que los 40 parámetros raros utilizados para controlar las fuentes acústicas y las funciones de transferencia en KLSyn, y que se emplearon en anteriores prácticas de síntesis de voz. Los parámetros HL, a través de un mapeo, se convierten en parámetros KL, como aparece en el esquema de la Fig.2 (tomada de Hanson y Stevens). Esta figura también enumera a los 13 parámetros HL, y muestra cómo estos parámetros están relacionados con la forma del tracto vocal, con la configuración de la laringe y con la presión respiratoria. Los parámetros también aparecen enumerados en la Tabla 1.

En lugar de controlar directamente los parámetros de fuente acústica y el filtrado, HLsyn controla los parámetros articulatorios que dan lugar a las fuentes y a los filtros. Un componente central del mapeo es el cálculo de flujos de aire y presiones en el tracto vocal, y la estimación de las fuentes acústicas (ruido periódico o de turbulencia) que resultan de estos flujos de aire y presiones en las constricciones de las vías respiratorias. Los métodos para calcular las fuentes están descritos en Hanson y Stevens (ej. observe la Fig. 4).

Procedimientos

En la práctica en la que se utiliza el sintetizador HLsyn, realizará ejercicios que demuestran cómo éste funciona. El sintetizador está instalado en tres ordenadores con plataforma Windows NT, y será necesario realizar esta práctica en estaciones de trabajo distintas a las que se han venido usando en otras prácticas. Los tres ordenadores que usted debe usar son Formant (en 36-549), Glottis (en la clase de las prácticas de digitalización de sonidos) y Brogino (en la Biblioteca del Grupo de Habla). Si los ordenadores NO están ejecutándose sobre Windows NT, debe reiniciarlos.

Cuando se encienda el ordenador, use las flechas del teclado que indican hacia arriba y hacia abajo para seleccionar Windows NT. Puede entrar en el ordenador utilizando la cuenta y la clave del curso. Cuando esté en Windows NT, encontrará dos iconos en el escritorio que usará para la práctica: SpeechStation y HLsyn. Usará HLsyn para llevar a cabo la síntesis, y SpeechStation (un software para el procesamiento de la voz desarrollado por Sensimetrics) para examinar el espectrograma y el contorno de f_0 de los enunciados sintetizados.

Tarea: cómo modificar la sonoridad y el modo de articulación de las consonantes labiales.

Cuando haga doble clic en HLsyn le aparecerá una ventana de diálogo. Haga clic sobre el botón “Open” del menú y encuentre el archivo **abit_grpx.hld** (donde “x” representa el número de su grupo, por ejemplo, el grupo 1 debería utilizar el archivo **abit_grp1.hld**). El fichero contiene los parámetros HL que se utilizan para sintetizar (a mano) el enunciado corto “a bit”. Se adjunta un espectrograma de este enunciado sintetizado. Se manipularon un número de parámetros HL para sintetizar este enunciado. Puede ver estos parámetros en forma numérica o en modo gráfico (haga clic en “view” y elija modo gráfico HL o parámetros HL). Si hace clic en “view” también puede seleccionar “Spectrogram” para ver el espectrograma de los enunciados sintetizados. Si quiere usar SpeechStation, debe salvar el archivo sintetizado como fichero .wav y abrirlo con SpeechStation.

Dése cuenta de que el parámetro al (área de sección transversal de la apertura del labio en mm^2) va hasta cero para hacer la /b/, y luego es liberado en un tiempo de 250ms. La presión subglótica ps es aproximadamente de 8 cm H₂O durante la mayoría del enunciado, pero es más baja en el principio y en el final, reflejando el inicio y el cese del enunciado. El área de la apertura glótica ag permanece en aproximadamente 4 mm^2 durante la mayoría del enunciado, especialmente durante el cierre de la /b/. El área seccional transversal de la apertura an permanece en cero dado que no hay sonidos nasales en este enunciado. Dése cuenta de que los parámetros dc y ue se utilizan para ayudar a mantener la vibración de las cuerdas vocales durante el cierre de la /b/.

Puede hacer clic sobre los parámetros de presión y del flujo de aire (modo gráfico PF o parámetros PF, situados debajo de “view”). Observe cómo la presión intraoral (P_m) aumenta en el cierre de la /b/, y tenga en cuenta el pico en el flujo de aire de la boca (U_{acx}) siguiendo inmediatamente la explosión de la /b/. Explique estos cambios en P_m y U_{acx} .

También puede considerar los parámetros KL que resultan del mapeo. Por ejemplo, dése cuenta de cómo la AV (amplitud de sonoridad) cambia con tiempo durante el cierre de la /b/ y justo antes y después de este cierre. Observe también la AF cerca de la explosión, y estudie los parámetros HL y KL utilizados para sintetizar la /t/ final.

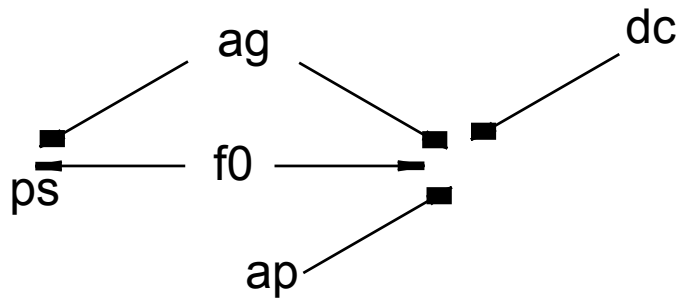
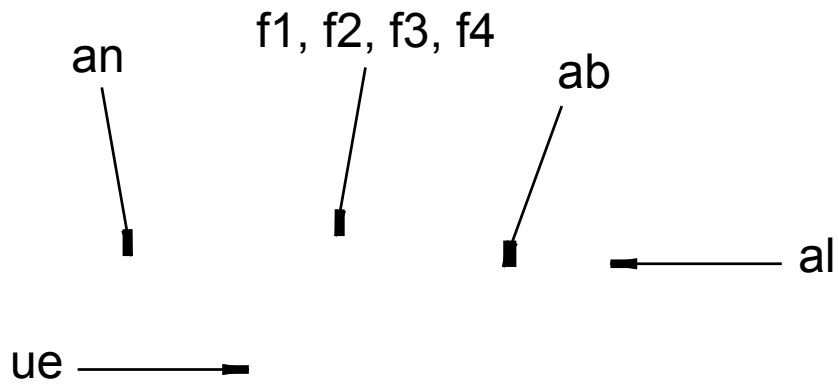
- (i) Como primer ejercicio, cambie el enunciado de “a bit” por “a mitt”. Para hacer esto, introduzca el parámetro an (área de la apertura velofaríngea). Comience con an aumentando aproximadamente a 40 ms antes del cierre labial, hasta un máximo de aproximadamente 20 cm^2 durante el cierre, disminuyendo luego a cero aproximadamente a 40 ms de la vocal /I/. Es posible que quiera deshacer los parámetros ue , dc y ap que se han introducido en “a bit”. Después de que haya realizado los cambios, haga clic en “Synthesize”. También puede hacer clic en “Play” para oír el enunciado. Experimente con el tiempo y el máximo valor de an . Observe el flujo de aire y la presión intraoral durante la producción de la /m/.

- (ii) Como segundo ejemplo comience con el enunciado original “a bit” y cámbielo por “a pit”. Esto debería hacerse aumentando la apertura glótica ag , comenzando aproximadamente 20 ms antes del cierre labial. Haga que aumente linealmente desde este punto hasta un máximo de aproximadamente 30 mm^2 en el momento de la explosión labial. Luego vuelva a disminuir ag a su valor nominal de 4 mm^2 , más o menos 50 ms después de la explosión. Esto debería conducir a un intervalo de aspiración. Debería deshacer el parámetro ue utilizado para la /b/, y debería cambiar el signo de dc para simular un entumecimiento aumentado de las cuerdas vocales y de las paredes del tracto vocal. Observe y grabe los distintos aspectos de las presiones, flujos y amplitudes de la fuente KL (AV, AF, AH).
- (iii) Lleve a cabo algunos cambios en los parámetros del enunciado original “a bit”. Algunos cambios aconsejados serían:
- Seleccione $ue = 0$ en la /b/.
 - ajuste $dc = 0$ en todas partes.
 - ajuste la $f\theta =$ una constante (digamos 1200) en todas partes. Dése cuenta que el parámetro KL F0 no es constante debido a otros efectos. Explíquelo.
 - ajuste $ps =$ una constante en todas partes.

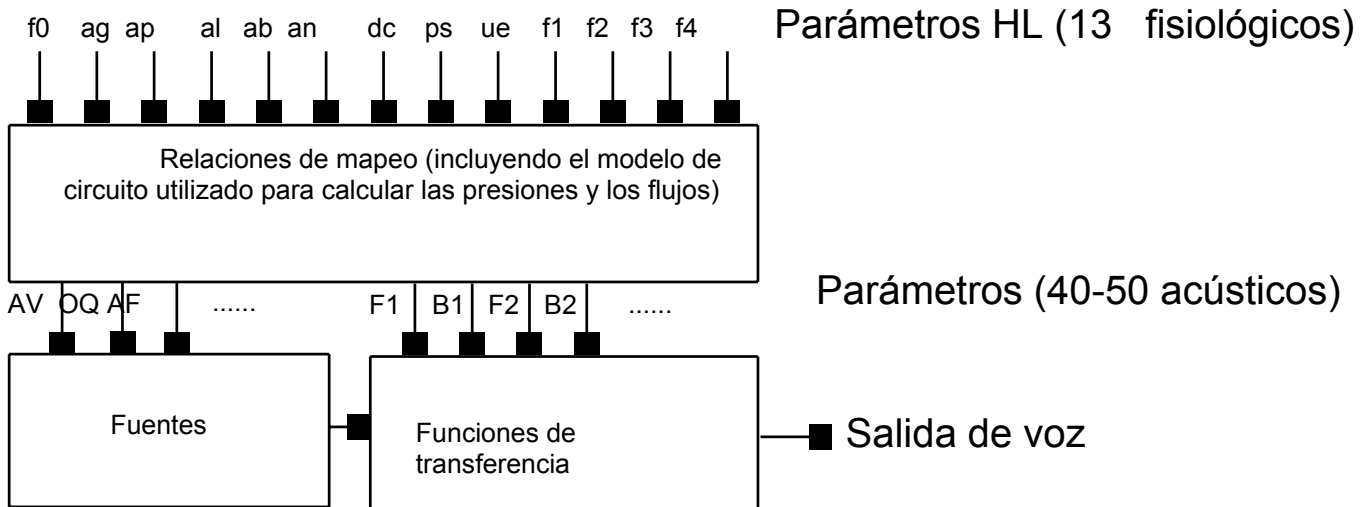
Observe el resultado en cada caso y coméntelo.

Tabla I. Descripción de los parámetros HLsyn

Parámetro	Rango	Descripción
f1	150-1300	Primera frecuencia natural del tracto vocal, suponiendo que no existen constricciones locales. (Hz)
f2	550-3000	Segunda frecuencia natural del tracto vocal, suponiendo que no existen constricciones locales. (Hz)
f3	1200-4800	Tercera frecuencia natural del tracto vocal, suponiendo que no existen constricciones locales (Hz)
f4	2400-4990	Cuarta frecuencia natural del tracto vocal, suponiendo que no existen constricciones locales (Hz)
f0	0-5000	Frecuencia fundamental debido a ajustes activos de las cuerdas vocales (Hz x 10)
ag	0-40	Área media de apertura glótica entre la porción membranosa de las cuerdas vocales (mm ²)
ap	0-40	Área de la apertura glótica posterior (mm ²)
ps	0-20	Presión subglótica (cm H ₂ O)
al	0-200	Área de sección transversal de constricción en los labios (mm ²)
ab	0-200	Área de sección transversal de la constricción del dorso de la lengua (mm ²)
an	0-100	Área de sección transversal de la pared velofaríngea (mm ²)
ue	-200 - 200	Relación de aumento del volumen del tracto vocal (cm ³ /s)
dc	-100 - 100	Cambio en las resistencias a la vibración de la pared o cuerdas vocales (%)



(a)



(b)

