ELO 385 – Laboratorio de Procesamiento Digital de Señales **Laboratorio # 6:**

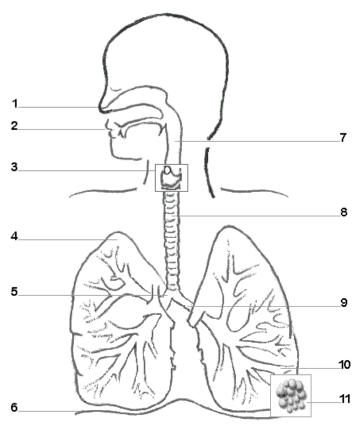
Análisis y Compresión de la Voz Humana

Matías Zañartu, Ph.D.

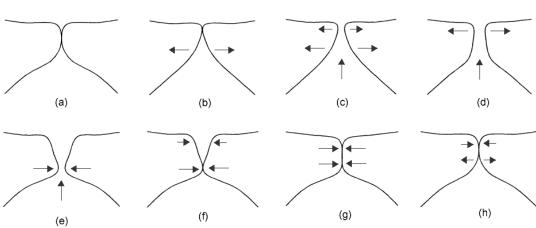
Departamento de Electrónica Universidad Técnica Federico Santa María

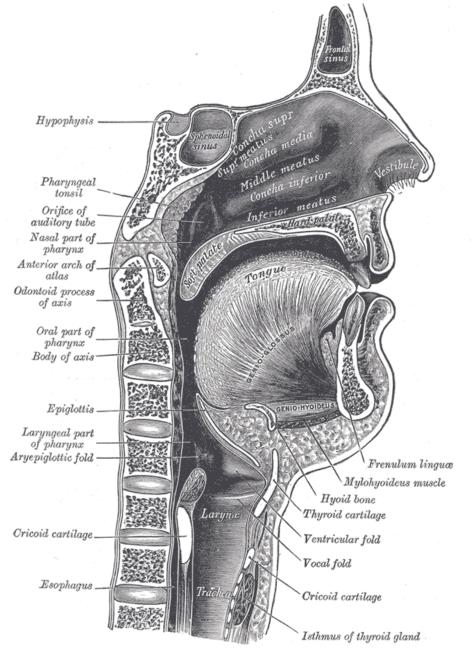


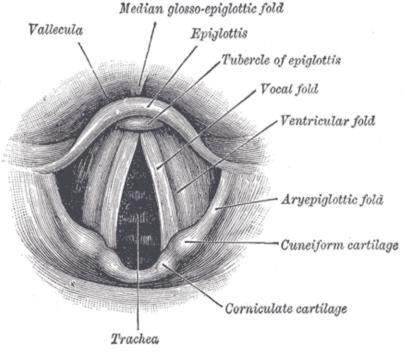
Sistema completo

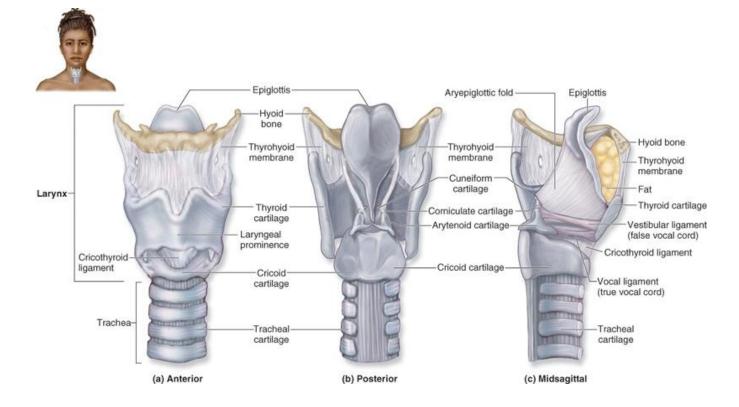


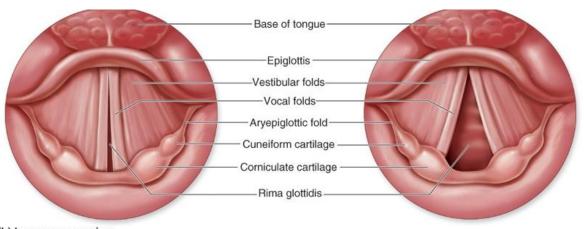
Ciclo típico de vibración de las cuerdas vocales







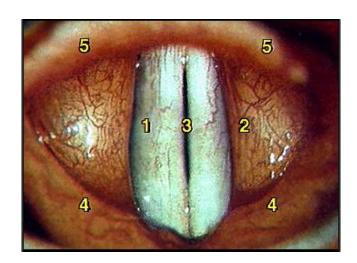






(b) Laryngoscope view

Patologías en las cuerdas vocales



Normal



Edema de Reinke

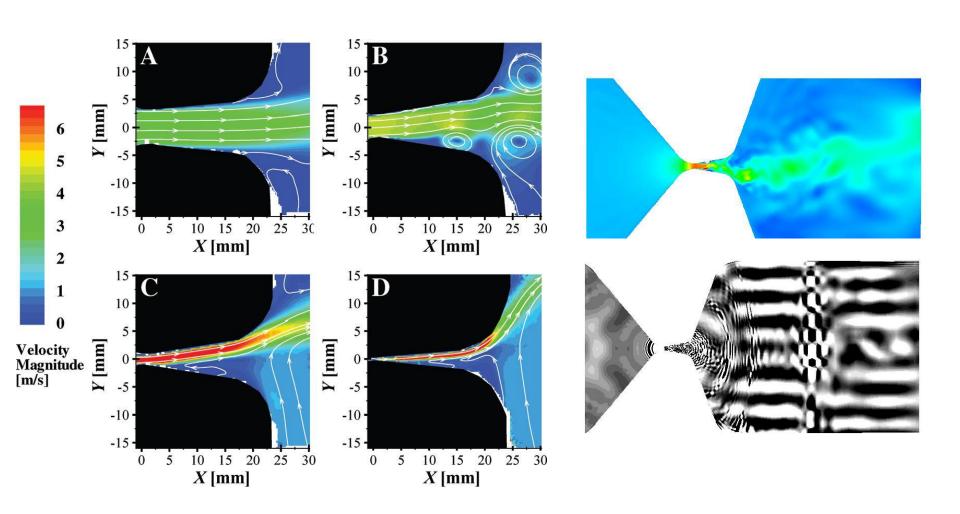


Pólipos

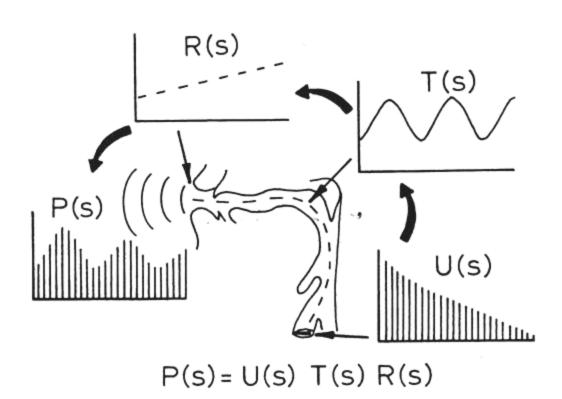


Cáncer Laringeo

Generación de sonido en la glotis



Representación básica de la producción de la voz humana

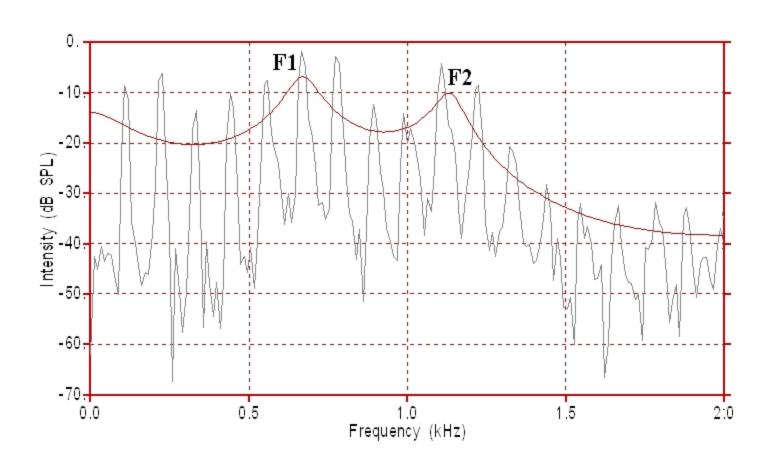




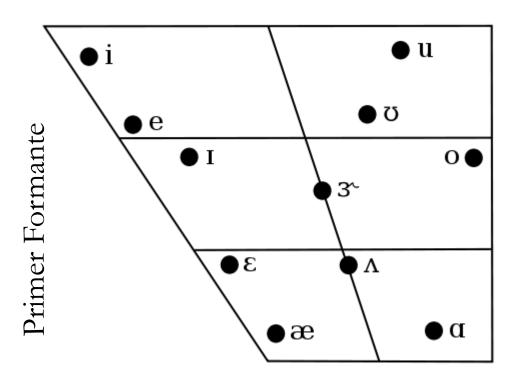
• • • Aspectos Generales

- Para analizar las señales de la voz humana es necesario considerar nuevas herramientas para estudiar su espectro
- FFT: Representa a la señal modulada, es decir harmónicos del sonido de las cuerdas vocales con amplitudes controladas por la modulación del tracto vocal
- Tracto: Es posible visualizar sólo el efecto modulador del tracto mediante herramientas tales como LPC, FFT con N pequeño, Cepstrum, etc.
- Espectrograma: Una representación de la STFT, la cual muestra la composición espectral de una señal a lo largo del tiempo

Formantes de Vocales



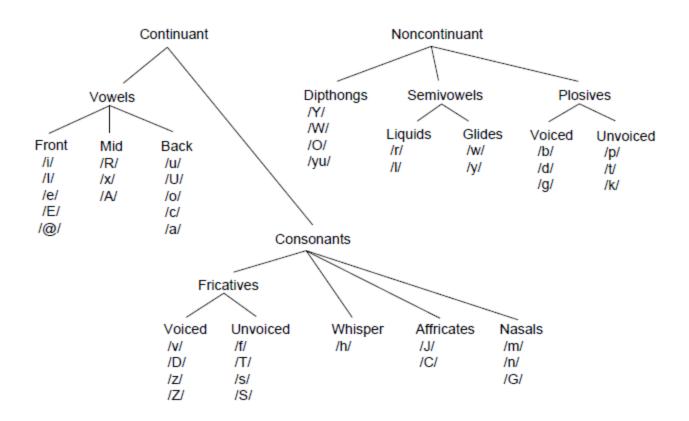
Formantes de Vocales en Inglés (Las Vocales del Español son un subconjunto)



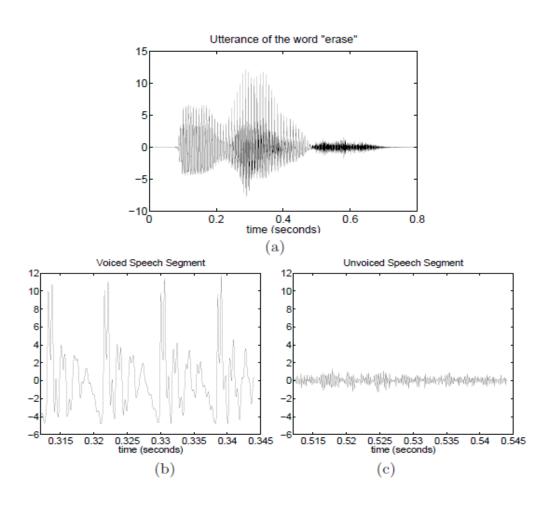
Formant Frequencies for the Vowels				
Typewritten Symbol for the Vowel	Typical Word	F1 (Hz)	F2 (Hz)	F3 (Hz)
IY I E AE UH A OW U OO ER	(beet) (bit) (bet) (bat) (but) (hot) (bought) (foot) (boot) (bird)	270 390 530 660 520 730 570 440 300 490	2290 1990 1840 1720 1190 1090 840 1020 870 1350	3010 2550 2480 2410 2390 2440 2410 2240 2240 1690

Segundo Formante

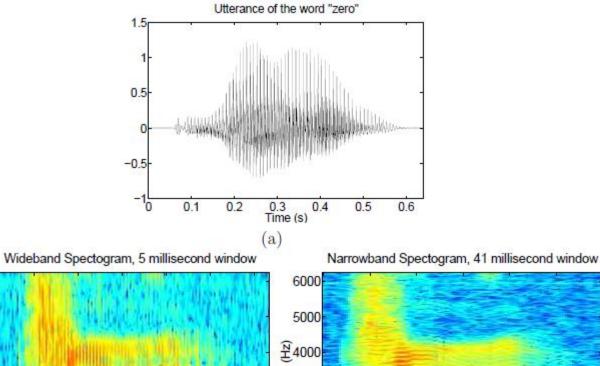
Fonemas del inglés

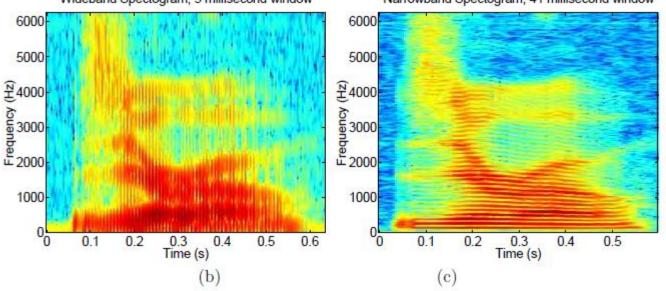


Variación Temporal de la señal de la Voz Humana



Espectrograma



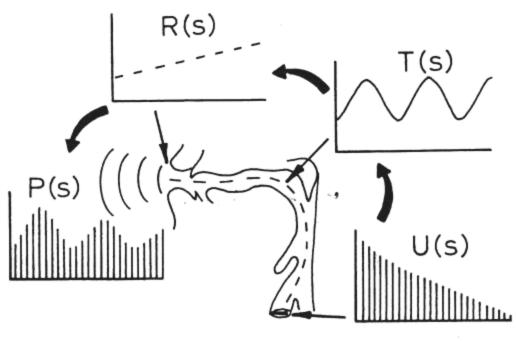




• • Aspectos Generales

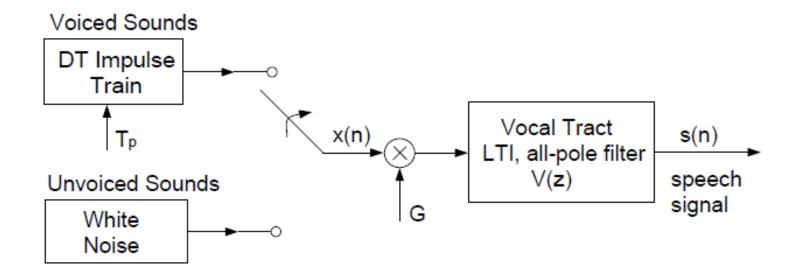
- Compresión de la voz se basa en representar la señal nuevamente con un modelo conocido y eficiente
- Reconstrucción: Síntesis o Vocoder
- La transmisión de voz se asocia a la transmisión de los parámetros del modelo, y no de la señal de audio
- El modelo normalmente utilizado busca representar la fuente del sonido y la modulación del tracto vocal
 - Fuente de sonido: vibración de cuerdas vocales (tono definido) o turbulencias de aire (ruido)
 - Modulación del tracto vocal: Resonancias se modelan con un filtro IIR (sólo polos)

Síntesis de Voz



P(s) = U(s) T(s) R(s)

Síntesis de Voz



• • • Fuente de Sonido

- Con vibración de las cuerdas: Señal con un tono definido y muchos armónicos (tren de impulsos)
- O Sin vibración de las cuerdas: Ruido de banda ancha
- O Detección de segmentos de la señal con vibración de las cuerdas:
 - Vocales o consonantes
 - Alta periodicidad y mayor energía
 - Cruces de ceros y umbral RMS
- Síntesis requiere estimar el tono de la señal:
 - FFT, autocorrelación

• • • Modulación del Tracto

- Estimación de los parámetros del filtro IIR se obtiene con los mismos datos de la señal
- Método: Predicción Lineal Modelo autoregresivo

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

$$s(n) = \sum_{k=1}^{P} a_k s(n-k) + G \cdot x(n)$$

- El filtro tiene orden P el cual suele ser P=fs/1000+2
- La estimación de los coeficientes **a**_k se repite en el tiempo para capturar las distintas modulaciones



• • • Actividades Programadas

• Parte I:

- Representar espectros de sonidos
- Analizar y clasificar sonidos de vocales
- Crear su propio espectrograma
- Estudiar espectrogramas de distinto ancho de banda

• Parte II:

- Crear su propio LPC
- Crear un vocoder para sintetizar una grabación de voz
- Evaluar compresión de la señal