

Procesamiento digital de señales de audio

Mecanismo y modelo de producción de voz

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

- ① **Introducción**
Comunicación a través de la voz

- ② **Mecanismo de producción de voz**
Fisiología del aparato vocal
Mecanismo de producción
Formantes
Clases de sonidos

- ③ **Modelo de producción de voz**
Modelo en tiempo discreto

Comunicación a través de la voz

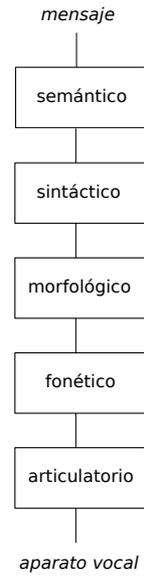
Niveles en la comunicación oral

- **semántico:** conceptos que forman el mensaje a comunicar
- **sintáctico:** vocabulario estructurado con reglas sintácticas
- **morfológico:** palabras individuales de acuerdo a una gramática
- **fonético:** secuencia de sonidos que conforman una palabra
- **articulatorio:** relación entre fonemas y estados del aparato vocal

procesamiento de voz

determinar relación entre secuencia de fonemas y señal de voz

dificultad: no hay correspondencia uno a uno entre fonema y configuración del aparato vocal

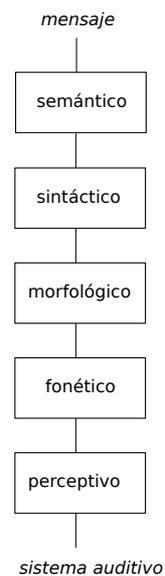


E. Chilton, Speech Analysis

Comunicación a través de la voz

Percepción del habla [Basso, 2006]

- no depende simplemente de las características de la señal acústica
- los patrones de una palabra se modifican en función de los sonidos que la preceden y la suceden
- un fonema altamente probable en un cierto contexto va a ser "oído" aún en ausencia de evidencia acústica
- el reconocimiento de voz depende de indicadores semánticos y sintácticos (y su importancia aumenta al aumentar ruido)



E. Chilton, Speech Analysis

Comunicación a través de la voz

Procesamiento digital de voz

- extraer información de la señal de voz
- mayor conocimiento posible de la estructura de los sonidos de la voz
i.e. de qué forma está codificada la información en la señal

resulta útil conocer:

- mecanismo de producción de voz
- principales clases de sonidos

objetivo: modelos de tiempo discreto para señales digitales de voz

Mecanismo de producción de voz

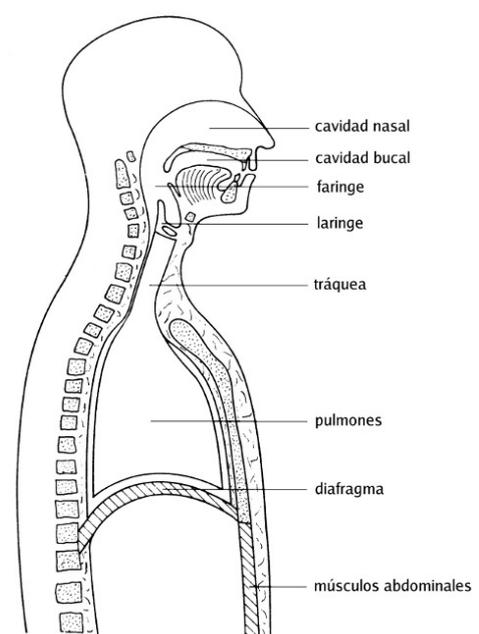
Fisiología del aparato vocal

de forma esquemática:

- **pulmones**
proveen flujo de aire
- **conducto**
pasaje de aire, que puede ser interrumpido
- **cavidad resonante**
modifica características espectrales

tracto vocal:

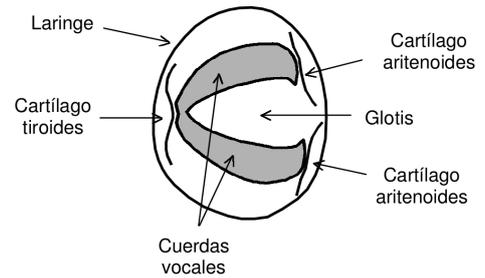
unido a través de la laringe, comprende la faringe y las cavidades bucal y nasal



Mecanismo de producción de voz

Cuerdas vocales

- dos membranas en la laringe, que restringen el pasaje de aire
- si comienzan a cerrarse, el aire experimenta una turbulencia, y se emite un sonido de origen aerodinámico (aspiración)
- al cerrarse más las cuerdas vocales comienzan a vibrar produciendo un sonido periódico frecuencia controlada por:
 - presión subglótica
 - tensión de los músculos
 - masa de las cuerdas

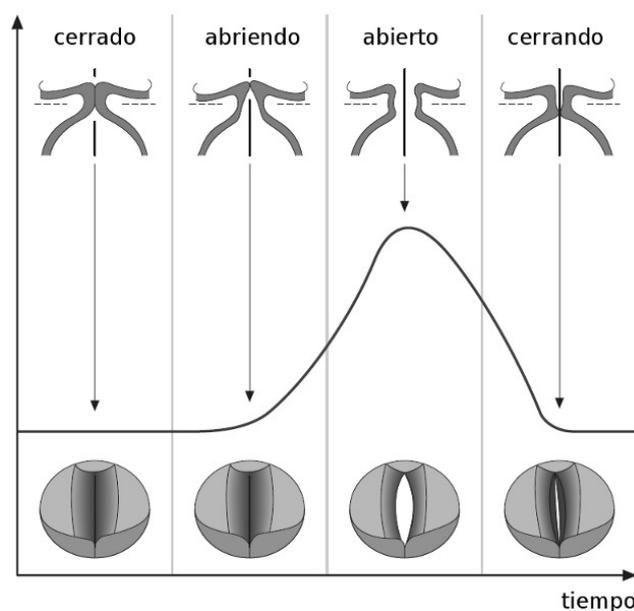


[Miyara, 2003]



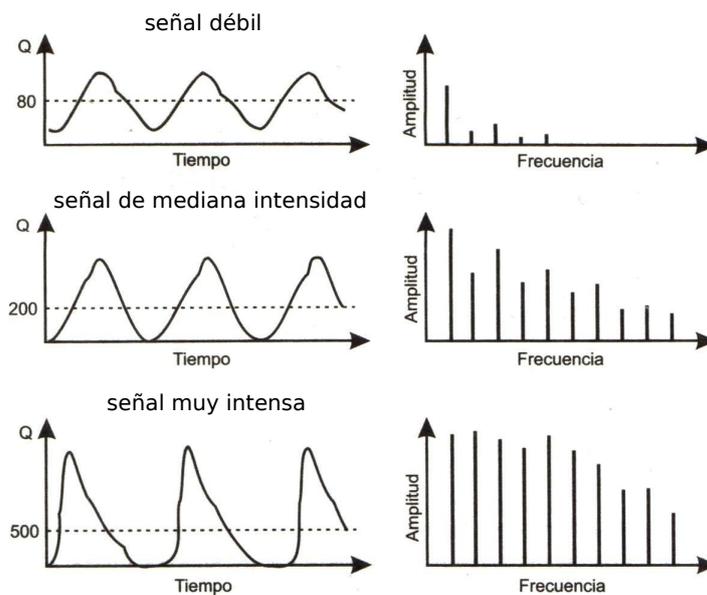
glotis abierta (respiración) y cerrada (vibración)

Mecanismo de producción de voz



esquema de generación de pulso glotal

Mecanismo de producción de voz

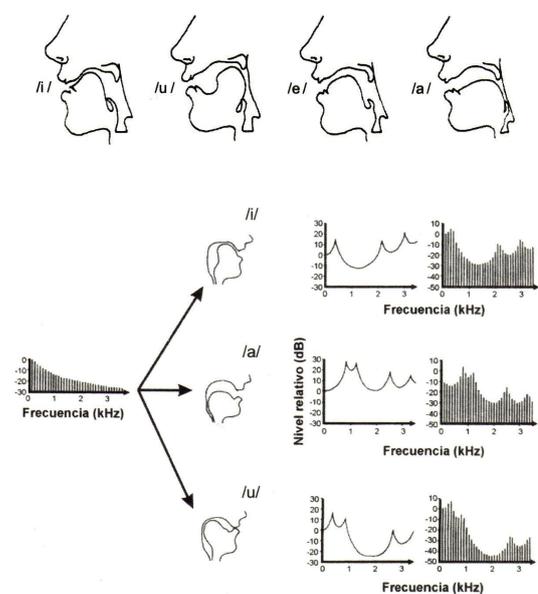


forma de onda y espectro en la laringe [Basso, 2006]

Mecanismo de producción de voz

Formantes

- **tracto vocal** tubos de sección no uniforme, modifican contenido espectral por su selectividad en frecuencia
- **formantes** frecuencias de resonancia, dependen de forma y dimensiones del tracto vocal
- **articulación** posición de partes móviles del aparato fonador, determina ubicación de frecuencias de resonancia (lengua, mandíbula inferior, labios, velo del paladar)

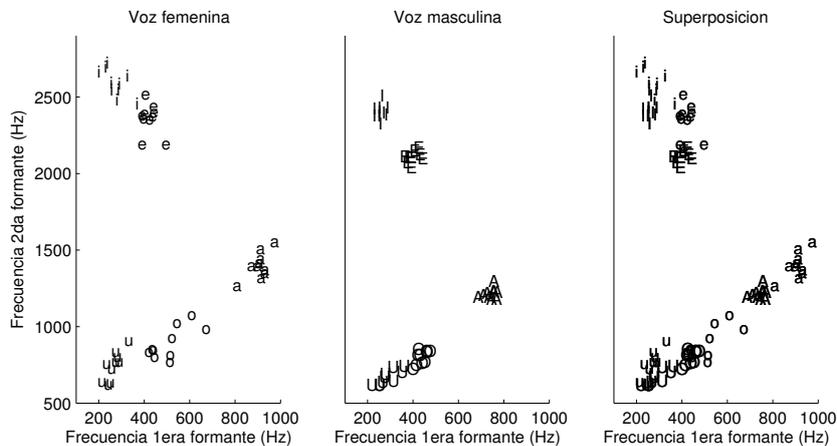


varias configuraciones del tracto vocal y efecto de filtrado [Basso, 2006]

Mecanismo de producción de voz

Formantes

- cada **vocal** corresponde a una combinación particular de formantes, y es posible distinguirlas incluso sólo a partir de las dos primeras
- varía entre diferentes personas y de acuerdo a su género y edad

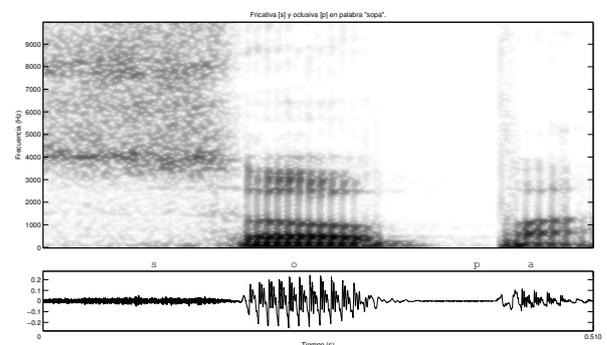


mapa de formantes para una voz femenina, una voz masculina y su superposición

Mecanismo de producción de voz

Mecanismos de producción de sonido

- **sonoros o tonales**: haciendo vibrar las cuerdas vocales (**voiced**) pulsos glotales aproximadamente periódicos, altura tonal definida
- **sordos o fricativos**: expeler aire a través de una restricción (**unvoiced**) fuente de excitación del tracto vocal es ruido de banda ancha
- **oclusivos o plosivos**: completo bloqueo y liberación repentina



fricativa [s], oclusiva [p] y vocales [o] y [a] en la palabra *sopa*

Mecanismo de producción de voz

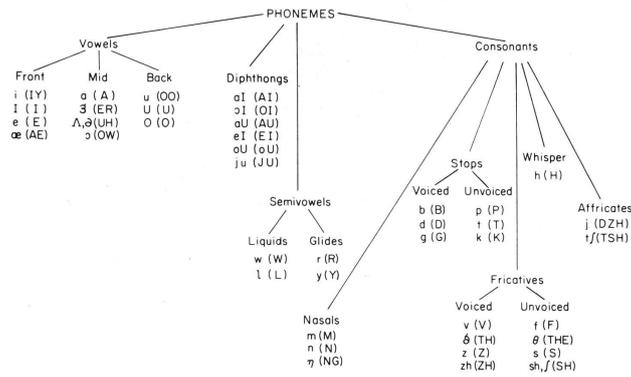
Clases de sonidos de la voz

fonemas: clases de sonidos de un idioma

clasificación según diversos criterios:

- tonalidad, oralidad o nasalidad, lugar y modo de articulación, etc

categorías más importantes: vocales, consonantes, diptongos, semi-vocales y semi-consonantes



[Rabiner and Schafer, 2011]

Mecanismo de producción de voz

Clases de sonidos de la voz [Rabiner and Schafer, 2011, Phonetics, 2005]

- **vocales/consonantes** si hay restricción al pasaje de aire o no
- **continuidad** de un fonema
 - configuración fija del tracto vocal (e.g. vocales, consonantes fricativas, consonantes nasales)
 - configuración variable del tracto vocal (e.g. diptongos, consonantes oclusivas y africadas)
- **sonoridad** vibran cuerdas vocales (e.g. vocales, consonantes nasales)
- **nasalidad/oralidad** el aire pasa principalmente por la nariz o boca

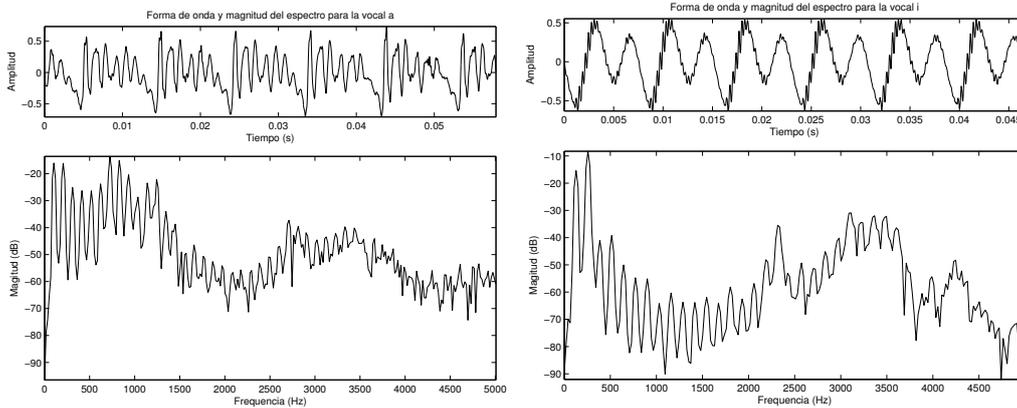


[Phonetics, 2005]

Mecanismo de producción de voz

Vocales

- articulación fija, no hay restricción al pasaje de aire
- fonemas sonoros (vibran las cuerdas vocales), altura definida
- forma de onda aprox. periódica, espectro de carácter discreto
- 4 o 5 formantes hasta los 4 kHz

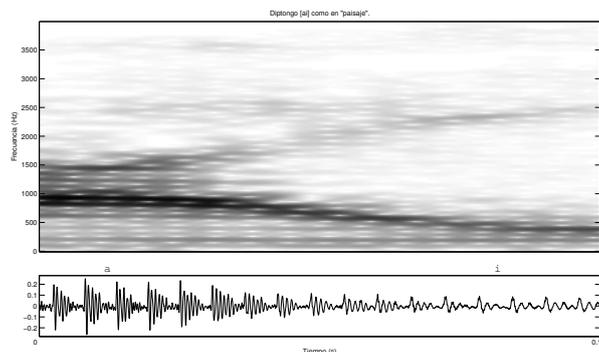


vocales [a] e [i]

Mecanismo de producción de voz

Diptongos

- secuencias de dos vocales que pertenecen a una misma sílaba
- articulación de la primera vocal y se mueve hacia la segunda
- vocal de mayor apertura es el núcleo de la sílaba
- la otra es semi-consonante o semi-vocal según si la precede o sigue

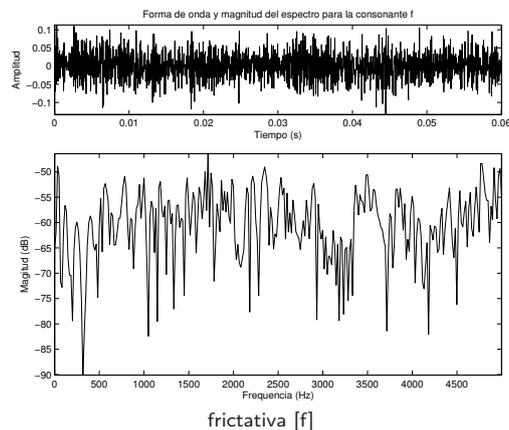


diptongo [ai] como en *paisaje* 

Mecanismo de producción de voz

Consonantes fricativa

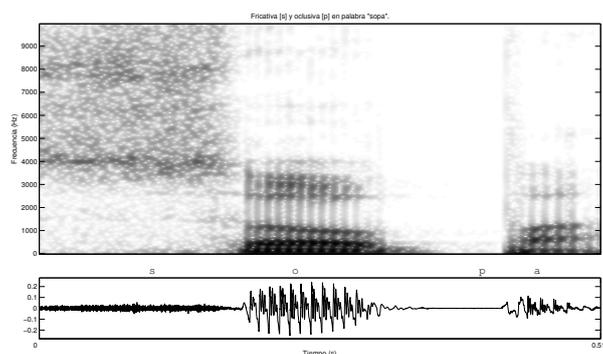
- forzando pasaje de aire a través de pequeña hendidura (e.g. [s],[f])
- sonido sordo, no tiene altura definida
- forma de onda parece aleatoria, espectro de banda ancha
- 2 ó 3 formantes, menos pronunciadas respecto a los sonidos sonoros



Mecanismo de producción de voz

Consonantes oclusivas

- bloqueo completo del tracto vocal y apertura súbita
- ataque impulsivo intenso seguido de vibración que se extingue rápido
- el transitorio tiene un espectro de banda ancha
- se dividen a su vez en sordas (e.g. [p],[k]) y sonoras (e.g. [b],[g])

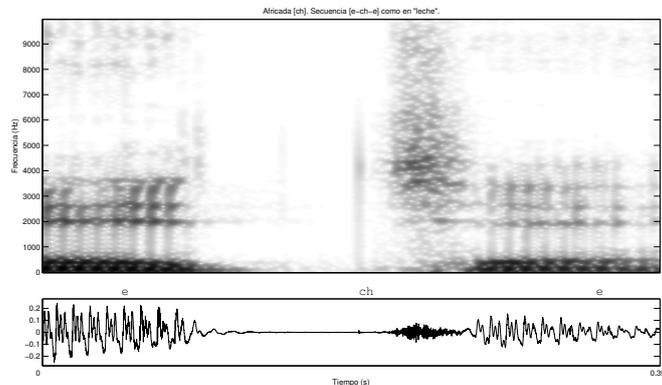


fricativa [s], oclusiva [p] y vocales [o] y [a] en la palabra *sopa*

Mecanismo de producción de voz

Consonantes africadas

- bloqueo completo seguido de una fase de fricción del aire
- comportamiento asimilable a consonante oclusiva seguida de consonante fricativa
- ejemplos: la letra “y” seguida de vocal, y la letra “ch”

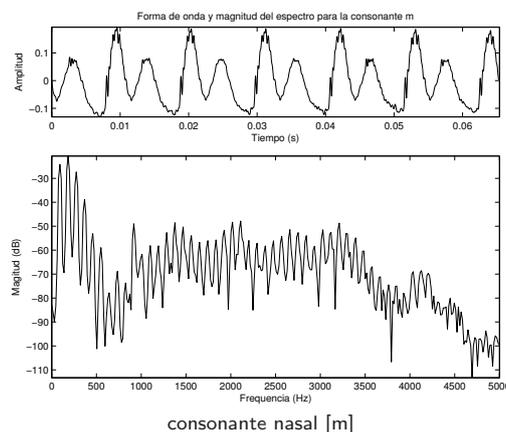


africada [ch] como en la palabra *leche* 

Mecanismo de producción de voz

Consonantes nasales

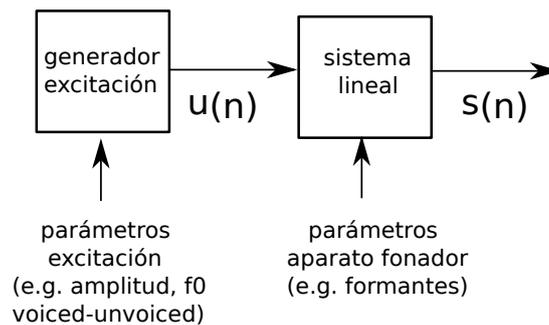
- cerrar pasaje de aire por la boca y descender el velo del paladar
- forma de onda aproximadamente periódica, fonema sonoro
- los varios sonidos producidos por las letras “m”, “n” y “ñ”
- la boca atenúa ciertas frecuencias (anti-resonancias)



Modelo de producción de voz

Modelo en tiempo discreto

- señal modelada como salida de sistema no lineal variante en el tiempo
- modelo en tiempo discreto simplificado:
sistema lineal y variante en el tiempo (invariante en tiempo corto)
sistema representa la función de transferencia del aparato fonador



Modelo de producción de voz

Modelo en tiempo discreto

- función de transferencia de estado estacionario:

$$H(z) = \frac{S(z)}{U(z)} = G \frac{1 + \sum_{l=1}^q b_l z^{-l}}{1 + \sum_{k=1}^p a_k z^{-k}}, \quad (1)$$

donde $S(z)$ y $U(z)$ son la transformada Z de la señal de voz y de la excitación respectivamente y G es un factor de ganancia del sistema.

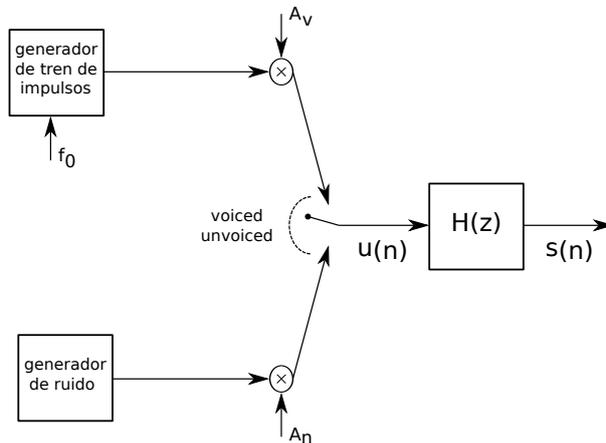
- corresponde a un modelo autorregresivo de media móvil (ARMA)

$$s(n) = - \sum_{k=1}^p a_k s(n-k) + G \sum_{l=0}^q b_l u(n-l), \quad \text{con } b_0 = 1. \quad (2)$$

Modelo de producción de voz

Modelo en tiempo discreto

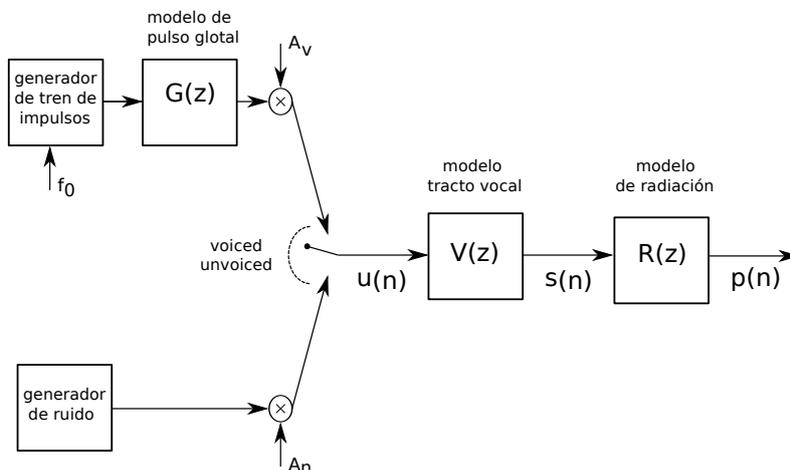
- excitación depende del tipo sonido
 - sonoro señal periódica (tren de impulsos)
 - sordo señal aleatoria (ruido de banda ancha)



Modelo de producción de voz

Modelo en tiempo discreto

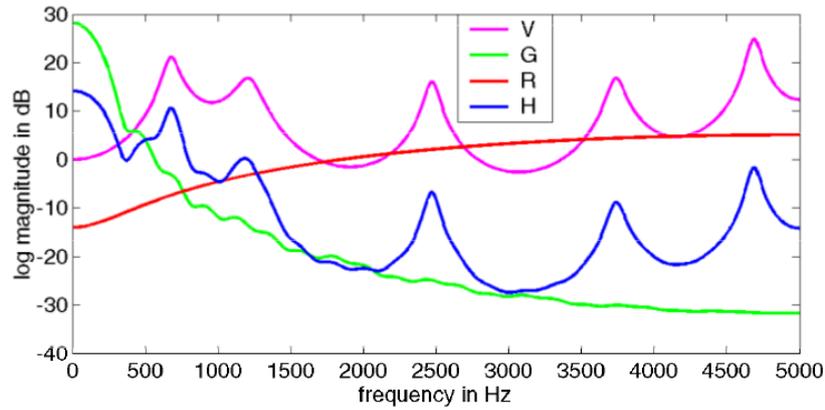
- refinamiento básico del modelo
 - modelo de pulso glotal filtro pasa bajos $G(z)$
 - modelo de radiación filtro pasa altos $R(z)$



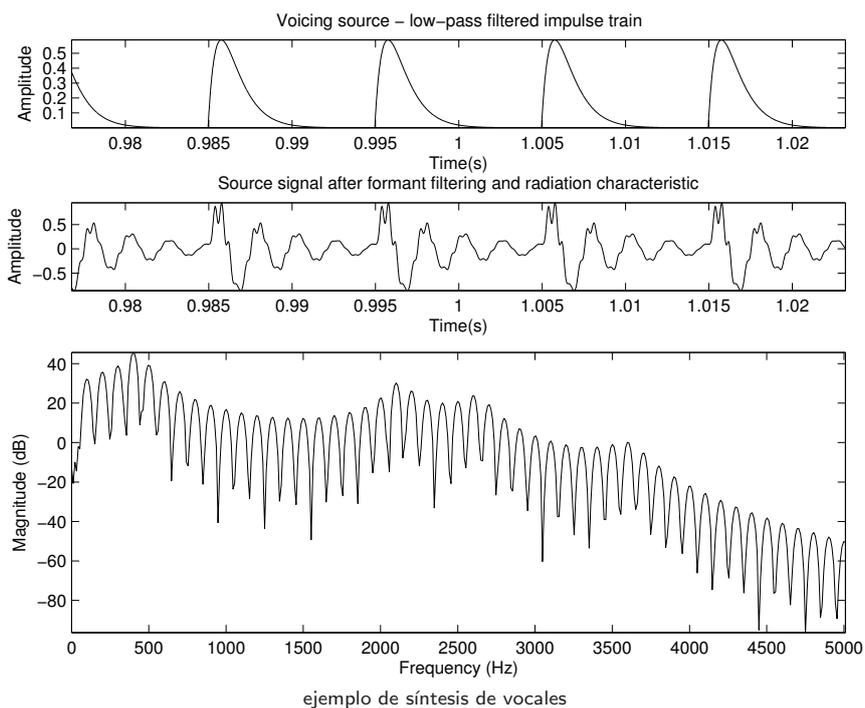
Modelo de producción de voz

Modelo en tiempo discreto

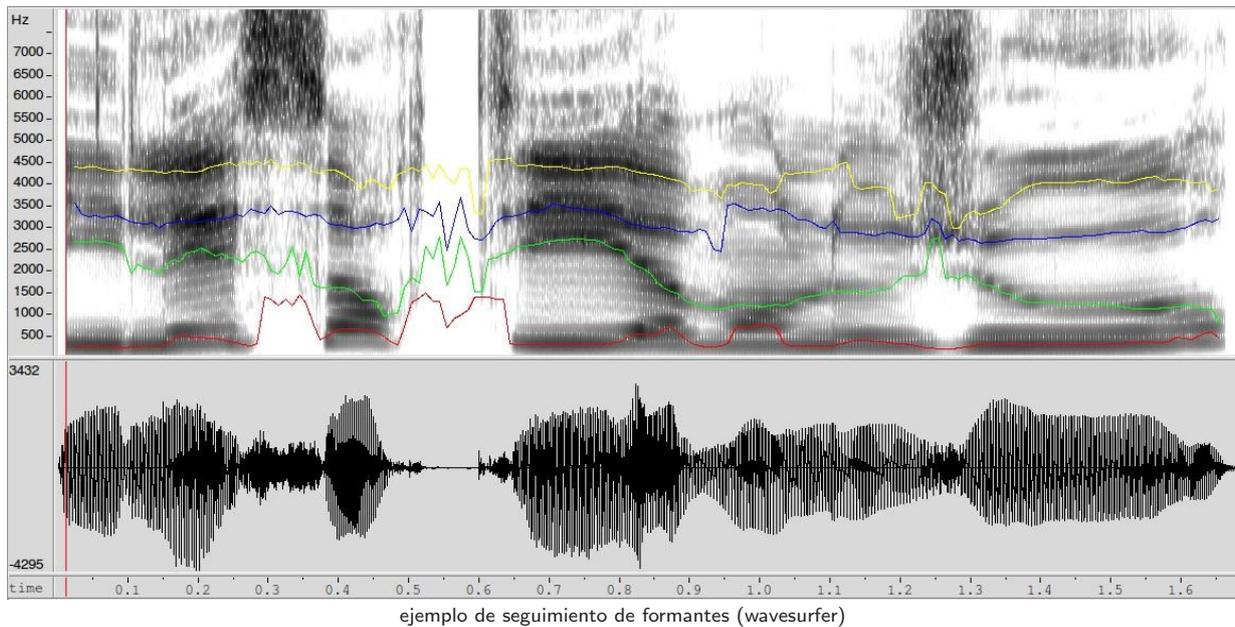
- refinamiento básico del modelo
 - modelo de pulso glotal filtro pasa bajos $G(z)$
 - modelo de radiación filtro pasa altos $R(z)$



Modelo de producción de voz



Modelo de producción de voz



Modelo de producción de voz

Algunos parámetros

- f_0 : voz hablada normal, mujeres 140-400 Hz, hombres 70-200 Hz
estos rangos pueden extenderse una octava o más en el canto
- **espectro**: máximo en 1000/500 Hz mujeres/hombres
cae aproximadamente -8dB/oct hasta 4 ó 5 kHz
- **rango dinámico**: aprox. 30-40 dB, SNR > 30 dB, implica unos 12 bits
- **tasa de fonemas**: más de 30 fonemas por segundo al hablar rápido

Referencias

-  Basso, G. (2006).
Percepción auditiva.
Universidad Nacional de Quilmes.
-  Miyara, F. (2003).
La voz humana.
Universidad Nacional de Rosario.
www.fceia.unr.edu.ar/acustica/biblio/fonatori.pdf.
-  Phonetics (2005).
Sounds of speech.
The University of Iowa.
<https://soundsofspeech.uiowa.edu>.
-  Rabiner, L. R. and Schafer, R. W. (2011).
Theory and Applications of Digital Speech Processing.
Prentice Hall, 1st edition.
Chapter 3 - Fundamentals of human speech production
Chapter 5 – Sec. 5.5 Digital models for sampled speech signals.