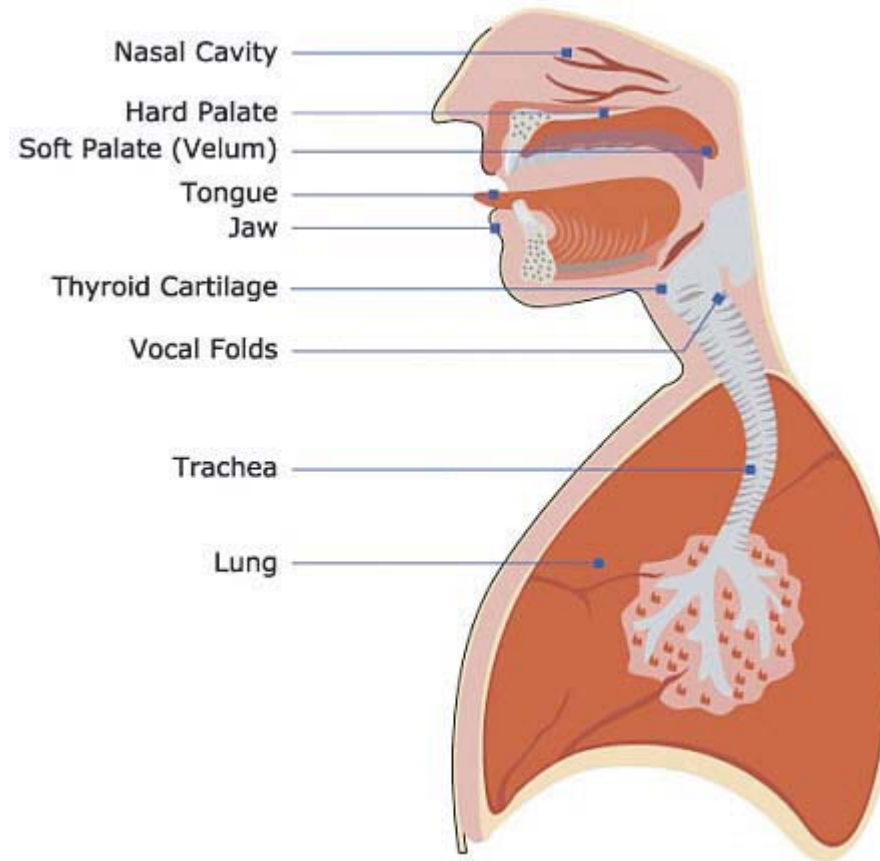


# Teoría acústica de la producción del habla

- Visión general
- Fuentes de sonido
- Función de transferencia del tracto vocal
  - Ecuaciones de onda
  - Propagación del sonido en un tubo acústico uniforme
- Cómo representar el tracto vocal con tubos acústicos simples
- Cómo calcular las frecuencias naturales a partir de funciones de área
- Cómo representar el tracto vocal con tubos uniformes múltiples

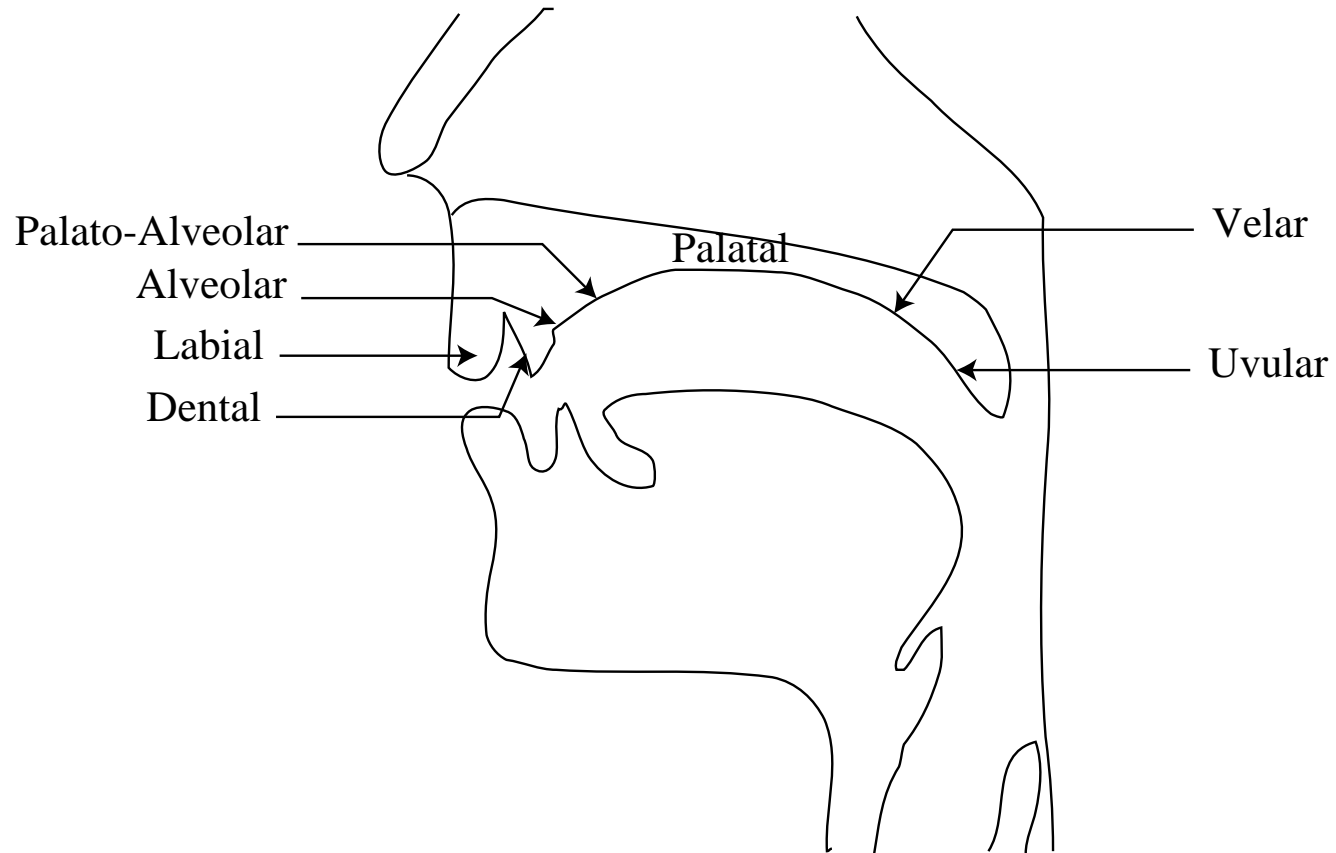
# Estructuras anatómicas para la producción del habla



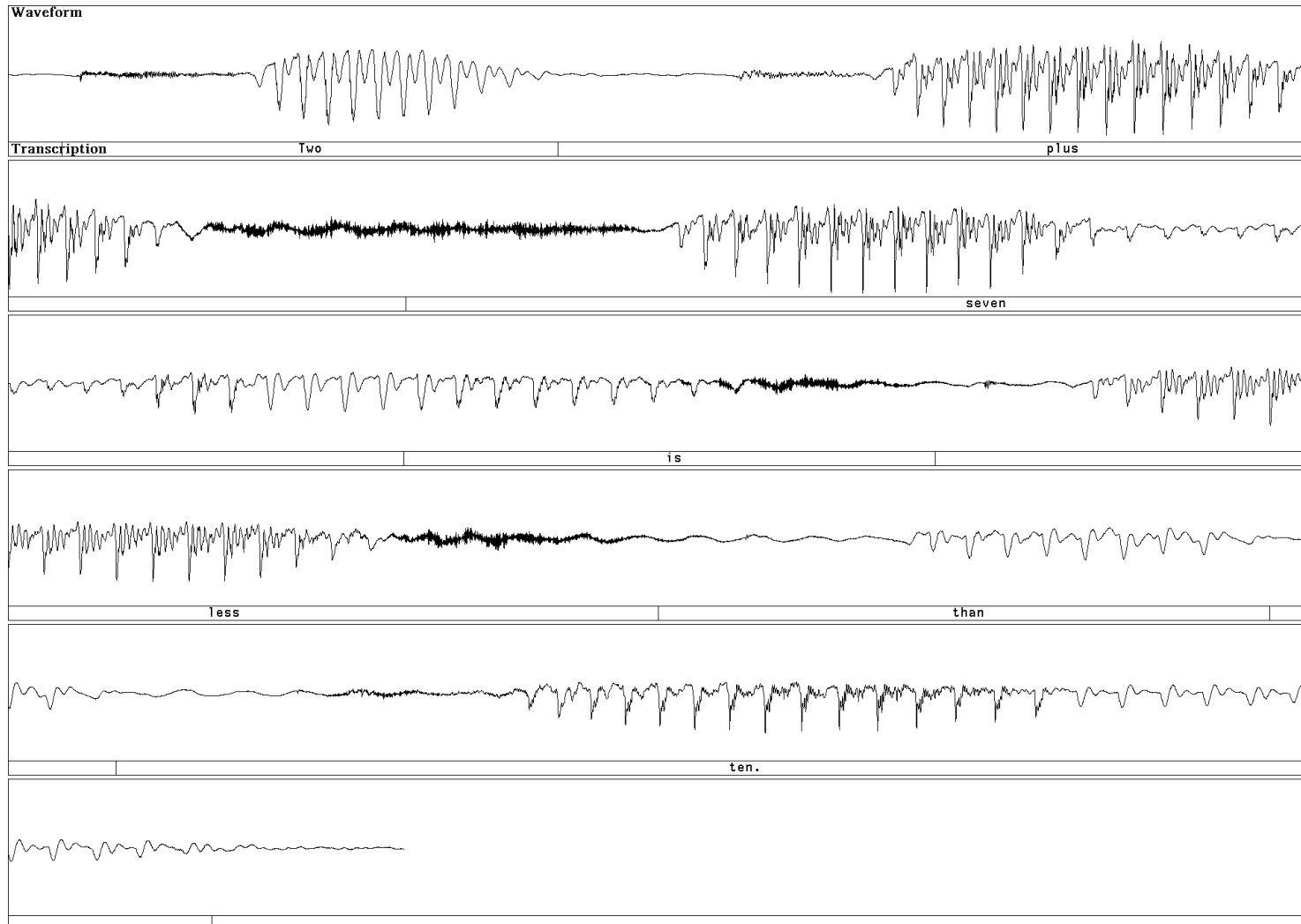
# Fonemas en inglés americano

<i>FONEMA</i>	<i>EJEMPLO</i>	<i>FONEMA</i>	<i>EJEMPLO</i>	<i>FONEMA</i>	<i>EJEMPLO</i>
/i <sup>y</sup> /	beat	/s/	see	/w/	wet
/ɪ/	bit	/ʃ/	she	/r/	red
/e <sup>y</sup> /	bait	/f/	fee	/l/	let
/ɛ/	bet	/θ/	thief	/y/	yet
/æ/	bat	/z/	z	/m/	meet
/ɑ/	Bob	/ʒ/	Gigi	/n/	neat
/ɔ/	bought	/v/	v	/ŋ/	sing
/ʌ/	but	/ð/	thee	/č/	church
/o <sup>w</sup> /	boat	/p/	pea	/j/	judge
/ʊ/	book	/t/	tea	/h/	heat
/u <sup>w</sup> /	boot	/k/	key		
/ɜ <sup>r</sup> /	Burt	/b/	bee		
/ɑ <sup>y</sup> /	bite	/d/	Dee		
/ɔ <sup>y</sup> /	Boyd	/g/	geese		
/ɑ <sup>w</sup> /	bout				
/ə/	about				

## Lugares de articulación de los sonidos del habla

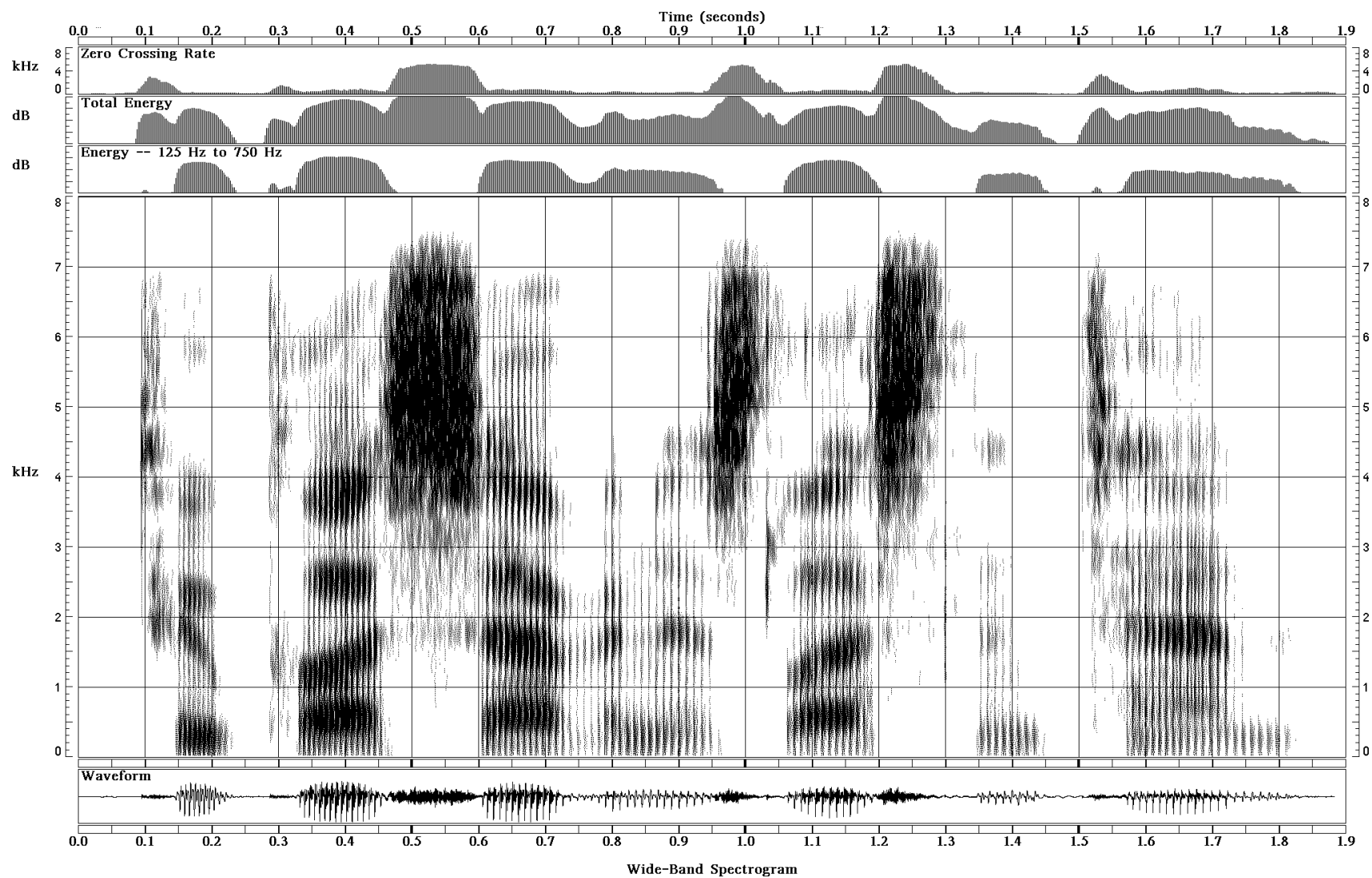


# Forma de onda de la voz: Un ejemplo



Dos más siete es menos que diez

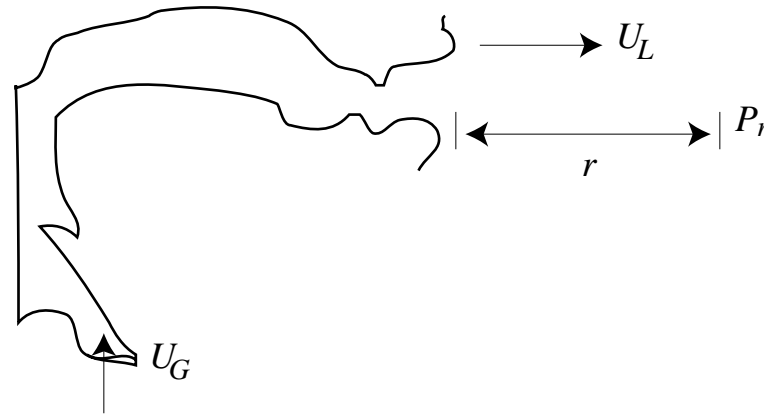
# Un espectrograma de banda ancha



Dos más siete es menos que diez

## Teoría acústica de la producción del habla

- Los rasgos acústicos del habla están modelados generalmente como una secuencia de la fuente, un filtro del tracto vocal y características de radiación.



$$P_r(j\Omega) = S(j\Omega) T(j\Omega) R(j\Omega)$$

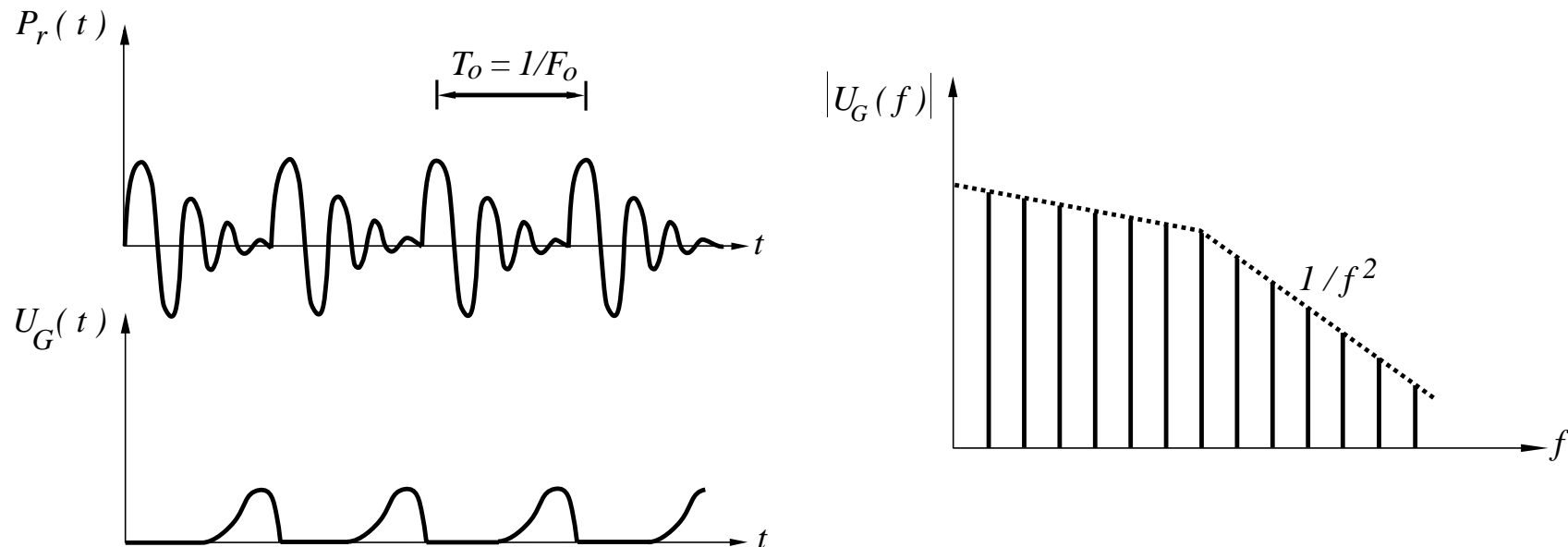
- Para la producción de las vocales:

$$\begin{aligned} S(j\Omega) &= U_G(j\Omega) \\ T(j\Omega) &= U_L(j\Omega) / U_G(j\Omega) \\ R(j\Omega) &= P_r(j\Omega) / U_L(j\Omega) \end{aligned}$$

# Fuente del sonido: Vibración de las cuerdas vocales

Modelada como fuente de la velocidad volumétrica en la glotis,

$$U_G(j\Omega)$$



	$F_0$ prom. (Hz)	$F_0$ min (Hz)	$F_0$ max (Hz)
Hombres	125	80	200
Mujeres	225	150	350
Niños	300	200	500

# Fuente del sonido: Ruido de turbulencia

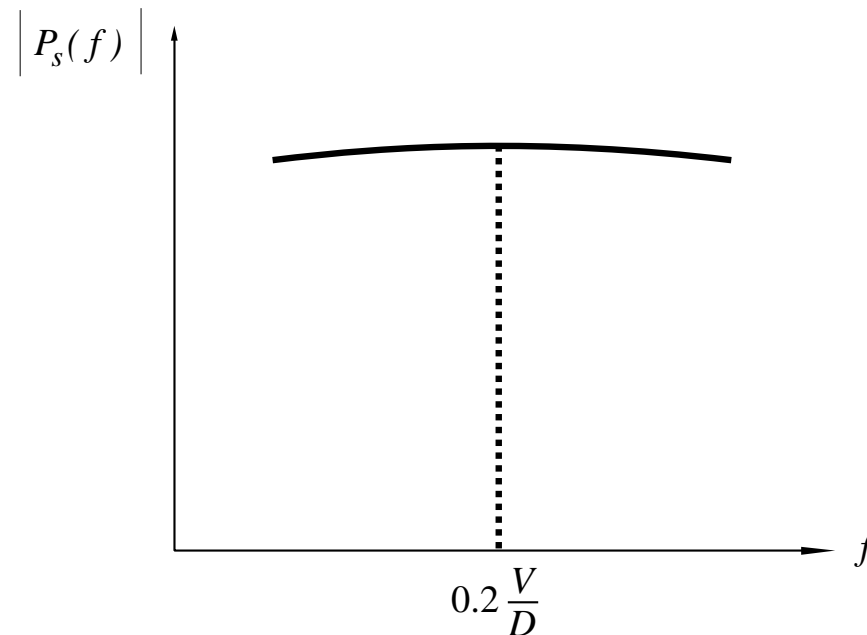
- El ruido de turbulencia se produce por una constricción en el tracto vocal.

- El ruido de **aspiración** se produce en la glotis.

- El ruido de **fricación** se produce encima de la glotis.

- Modelado como una **fente de presión sucesiva** en la constricción,  $P$

$$S(j\Omega)$$



$V$  : Velocidad de constricción

$D$ : Dimensión crítica

$$= \sqrt{\frac{4A}{\pi}} \approx \sqrt{A}$$

# Ecuaciones de onda en el tracto vocal

Definir:

$u(x, t)$	$\Rightarrow$	velocidad de la partícula	
$U(x, t)$	$\Rightarrow$	velocidad volumétrica ( $U = uA$ )	
$p(x, t)$	$\Rightarrow$	variación en la presión del sonido	$(P = P_0 + p)$
$\rho$	$\Rightarrow$	densidad de aire	
$c$	$\Rightarrow$	velocidad del sonido	

- Suponiendo una propagación de onda plana (para una dimensión cruzada  $\ll \lambda$ ) y un movimiento ondulatorio unidimensional, puede demostrarse que:

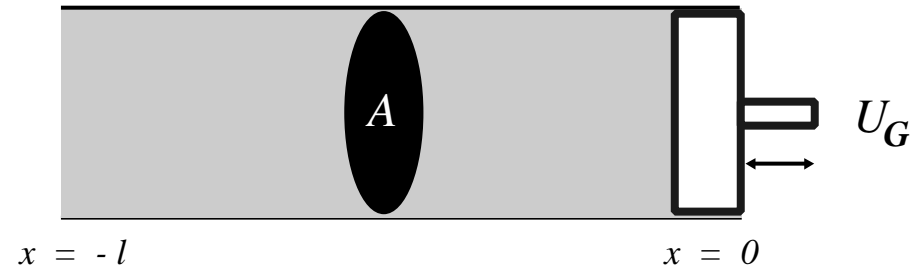
$$-\frac{\partial p}{\partial x} = \rho \frac{\partial u}{\partial t} \qquad -\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial p}{\partial t} \qquad \frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2}$$

- Las soluciones en el dominio del tiempo y la frecuencia presentan esta forma:

$$u(x, t) = u^+\left(t - \frac{x}{c}\right) - u^-\left(t + \frac{x}{c}\right) \qquad u(x, s) = \frac{1}{\rho c} \left[ P_+ e^{-sx/c} - P_- e^{sx/c} \right]$$

$$p(x, t) = \rho c \left[ u^+\left(t - \frac{x}{c}\right) + u^-\left(t + \frac{x}{c}\right) \right] \qquad p(x, s) = P_+ e^{-sx/c} + P_- e^{sx/c}$$

# Propagación del sonido en un tubo uniforme



- La función de transferencia del tracto vocal para las velocidades volumétricas es:

$$T(j\Omega) = \frac{U_L(j\Omega)}{U_G(j\Omega)} = \frac{U(-\ell, j\Omega)}{U(0, j\Omega)}$$

- Utilizando las condiciones de contorno  $U(0, s) = U_G(s)$  y  $P(-\ell, s) = 0$

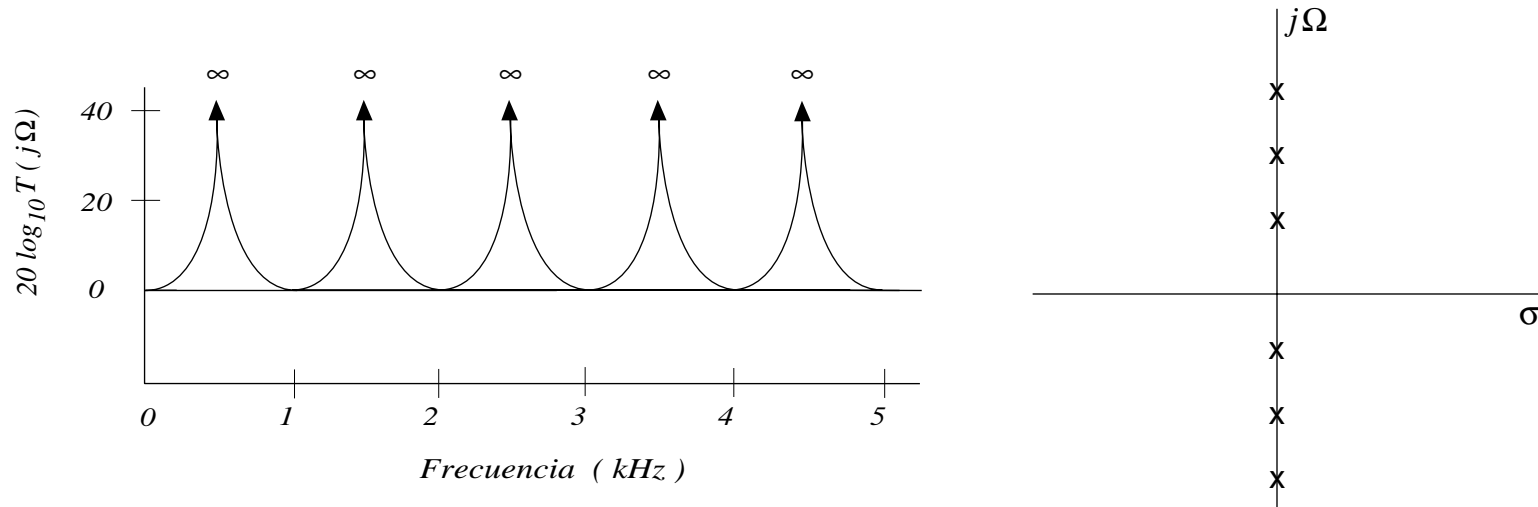
$$T(s) = \frac{2}{e^{s\ell/c} + e^{-s\ell/c}} \quad T(j\Omega) = \frac{1}{\cos(\Omega\ell/c)}$$

- Los polos de la función de transferencia  $T(j\Omega)$  están donde  $\cos(\Omega\ell/c) = 0$

$$\frac{(2\pi f_n)\ell}{c} = \frac{(2n-1)}{2}\pi \quad f_n = \frac{c}{4\ell}(2n-1) \quad \lambda_n = \frac{4\ell}{(2n-1)} \quad n = 1, 2, \dots$$

# Propagación del sonido en un tubo unidorme (cont.)

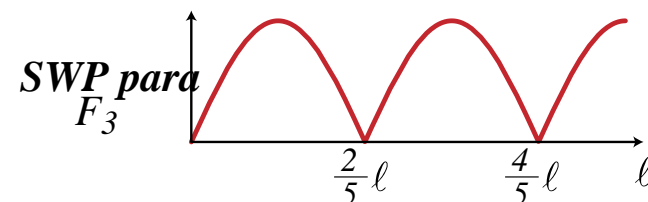
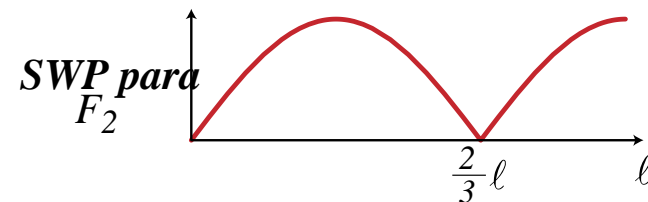
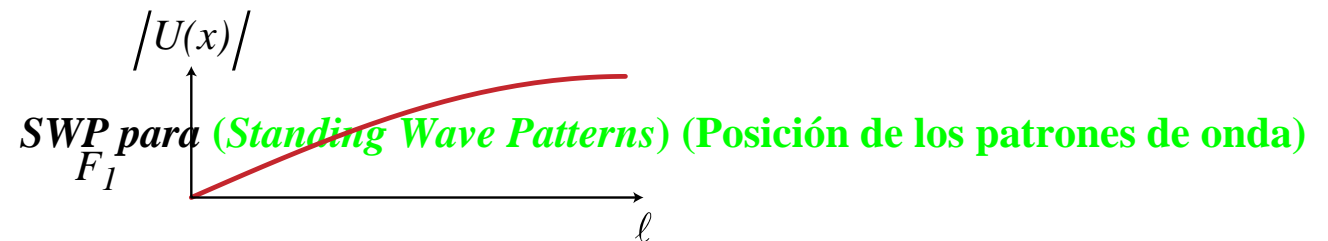
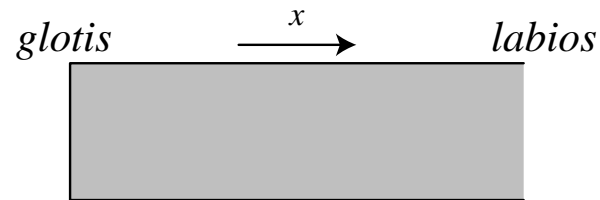
- Para  $c = 34,000 \text{ cm /seg}$ ,  $\ell = 17 \text{ cm}$ , las frecuencias naturales (conocidas también como formantes) aparecen en  $500\text{Hz}, 1500\text{Hz}, 2500\text{Hz}, \dots$



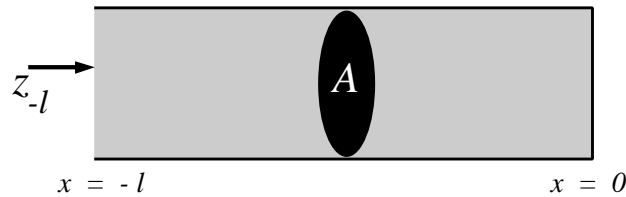
- La función de transferencia de un tubo **sin ramas laterales**, excitado en un extremo y con la respuesta medida en el otro, únicamente posee polos.
- Las frecuencias del formante tendrán un ancho de banda finito cuando se consideren las pérdidas del tracto vocal (ej., radiación, paredes, viscosidad, calor).
- La longitud del tracto vocal,  $\ell$ , corresponde a  $\frac{1}{4}\lambda_1^3, \frac{1}{4}\lambda_2^5, \frac{1}{4}\lambda_3^7, \dots$ , donde  $\lambda_i$  es la longitud de onda de la frecuencia natural de  $i^{\text{th}}$

# Posición de los patrones de onda en un tubo uniforme

Un tubo uniforme cerrado en un extremo y abierto en el otro, es conocido normalmente como un **resonador en cuarto de longitud de onda**.



## Frecuencias naturales de tubos acústicos simplificados



Resonador en un cuarto de longitud de onda

$$P(x, j\Omega) = 2P_+ \cos \frac{\Omega x}{c}$$

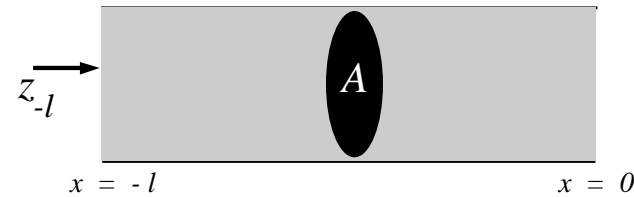
$$U(x, j\Omega) = -j \frac{A}{\rho c} 2P_+ \sin \frac{\Omega x}{c}$$

$$Y_{-l} = j \frac{A}{\rho c} \tan \frac{\Omega l}{c}$$

$$\approx j \Omega \frac{A l}{\rho c^2} = j \Omega C_A \quad \Omega l / c \ll 1$$

$$C_A = A l / \rho c^2 = \text{distensibilidad acústica}$$

$$f_n = \frac{c}{4l} (2n - 1) \quad n = 1, 2, \dots$$



Resonador en media longitud de onda

$$P(x, j\Omega) = -j 2P_+ \sin \frac{\Omega x}{c}$$

$$U(x, j\Omega) = \frac{A}{\rho c} 2P_+ \cos \frac{\Omega x}{c}$$

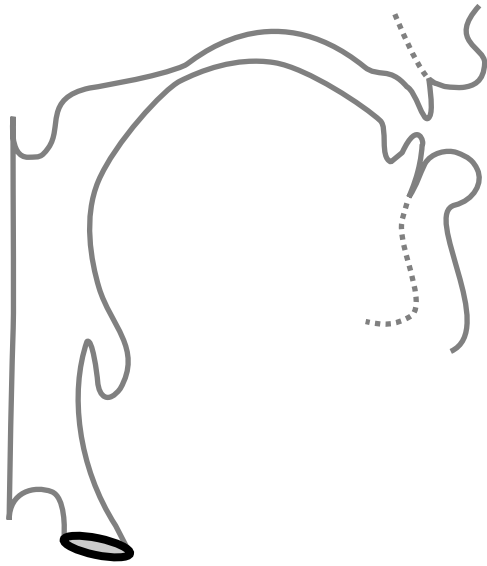
$$Y_{-l} = -j \frac{A}{\rho c} \cot \frac{\Omega l}{c}$$

$$\approx -j \frac{A}{\Omega \rho l} = -j \frac{1}{\Omega M_A} \quad \Omega l / c \ll 1$$

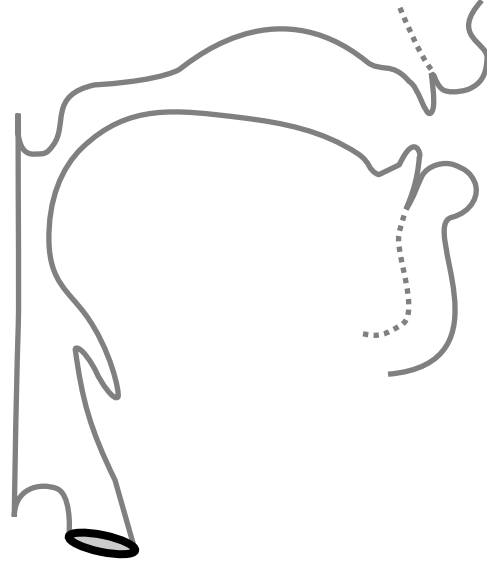
$$M_A = \rho l / A = \text{masa acústica}$$

$$f_n = \frac{c}{2l} n \quad n = 0, 1, 2, \dots$$

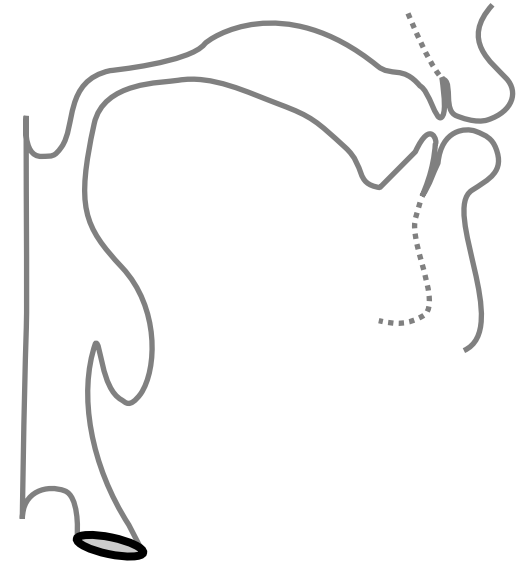
# Aproximación de las formas del tracto vocal



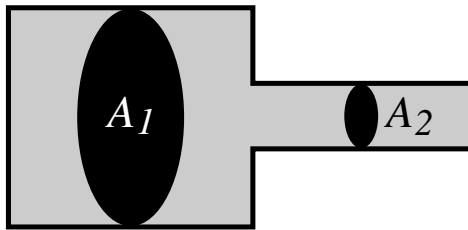
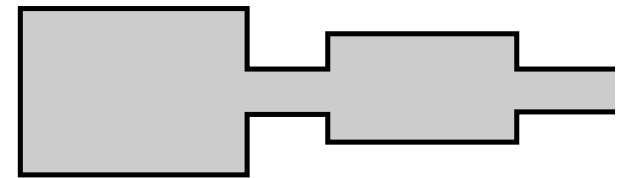
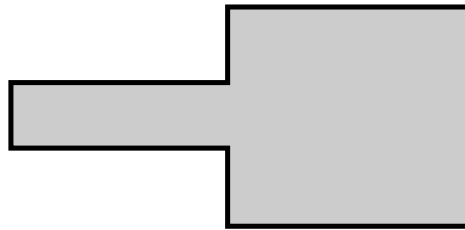
[ i ]



[ a ]

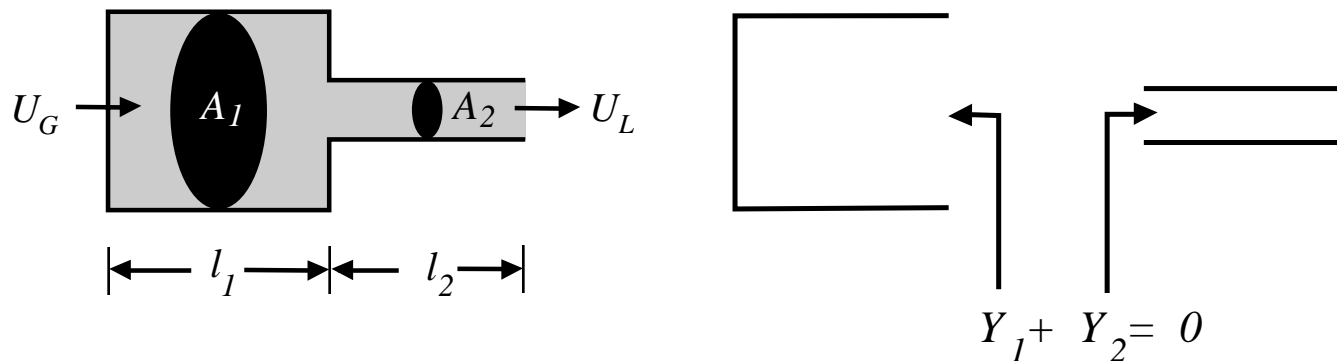


[ u ]


 $\left\langle \leftarrow l_1 \rightarrow \right\rangle \left\langle \leftarrow l_2 \rightarrow \right\rangle$ 


# Cómo calcular las frecuencias de resonancia natural

- Las frecuencias de resonancia se dan donde la función de obstaculización (o entrada) equivale a las condiciones de contorno naturales (ej., circuito abierto)



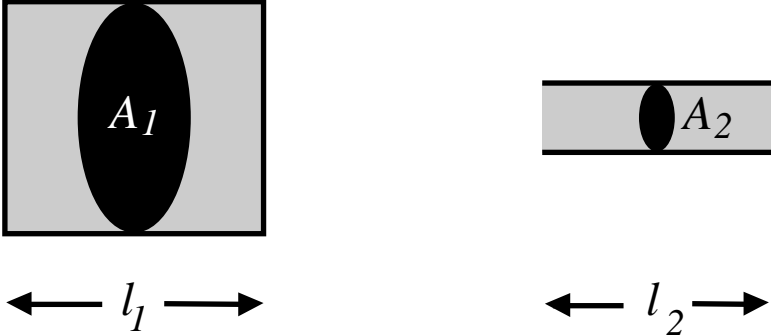
- Para una aproximación de dos tubos, es más fácil resolver  $Y_1 + Y_2 = 0$

$$j \frac{A_1}{\rho c} \tan \frac{\Omega \ell_1}{c} - j \frac{A_2}{\rho c} \cot \frac{\Omega \ell_2}{c} = 0$$

$$\sin \frac{\Omega \ell_1}{c} \sin \frac{\Omega \ell_2}{c} - \frac{A_2}{A_1} \cos \frac{\Omega \ell_1}{c} \cos \frac{\Omega \ell_2}{c} = 0$$

## Desacoplamiento de las aproximaciones de un tubo simple

- Si  $A_1 \gg A_2$ , o  $A_1 \ll A_2$ , los tubos pueden ser **desacoplados** y las frecuencias naturales de cada tubo se pueden calcular independientemente.
- Para la vocal /iʏ/, las frecuencias del formante se obtienen a partir de:



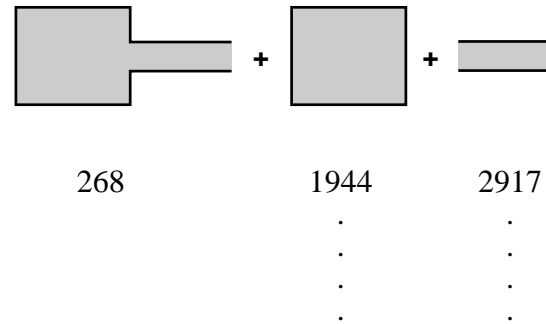
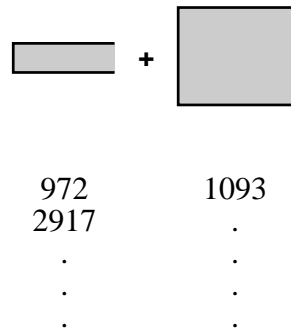
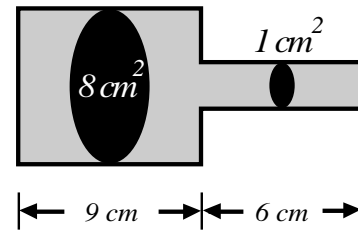
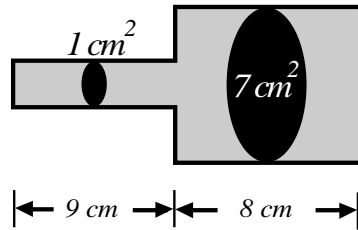
$f_n = \frac{c}{2\ell_1} n$       más       $f_n = \frac{c}{2\ell_2} n$

- Para frecuencias bajas:

$$f = \frac{c}{2\pi} \left[ \frac{A_2}{A_1 \ell_1 \ell_2} \right]^{1/2} = \frac{1}{2\pi} \left[ \frac{1}{C_{A_1} M_{A_2}} \right]^{1/2}$$

- Esta frecuencia de resonancia baja se denomina resonancia **Helmholtz**.

## Ejemplo de producción de la vocal



972  
2917  
.  
.  
.

1093  
.  
.  
.

268

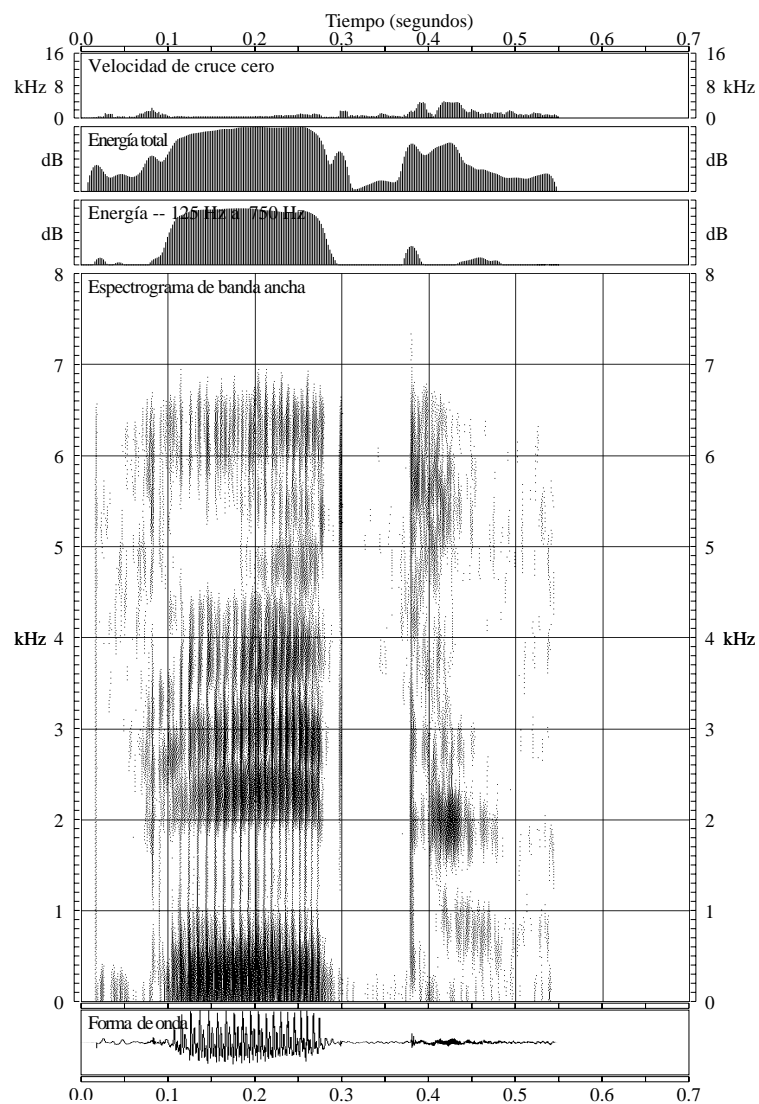
1944  
.  
.  
.

2917  
.  
.  
.

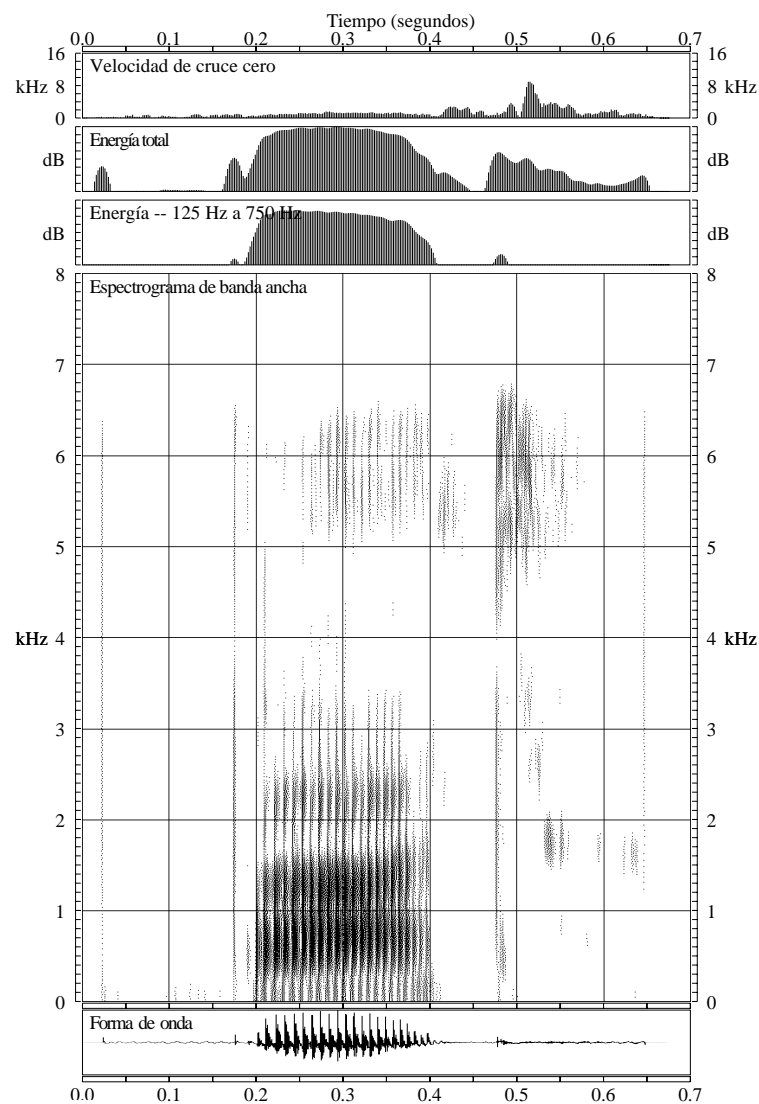
Formante	Real	Aproximado
F1	789	972
F2	1276	1093
F3	2808	2917
.	.	.
.	.	.

Formante	Real	Aproximado
F1	256	268
F2	1905	1944
F3	2917	2917
.	.	.
.	.	.

# Ejemplos de espectrogramas de vocal



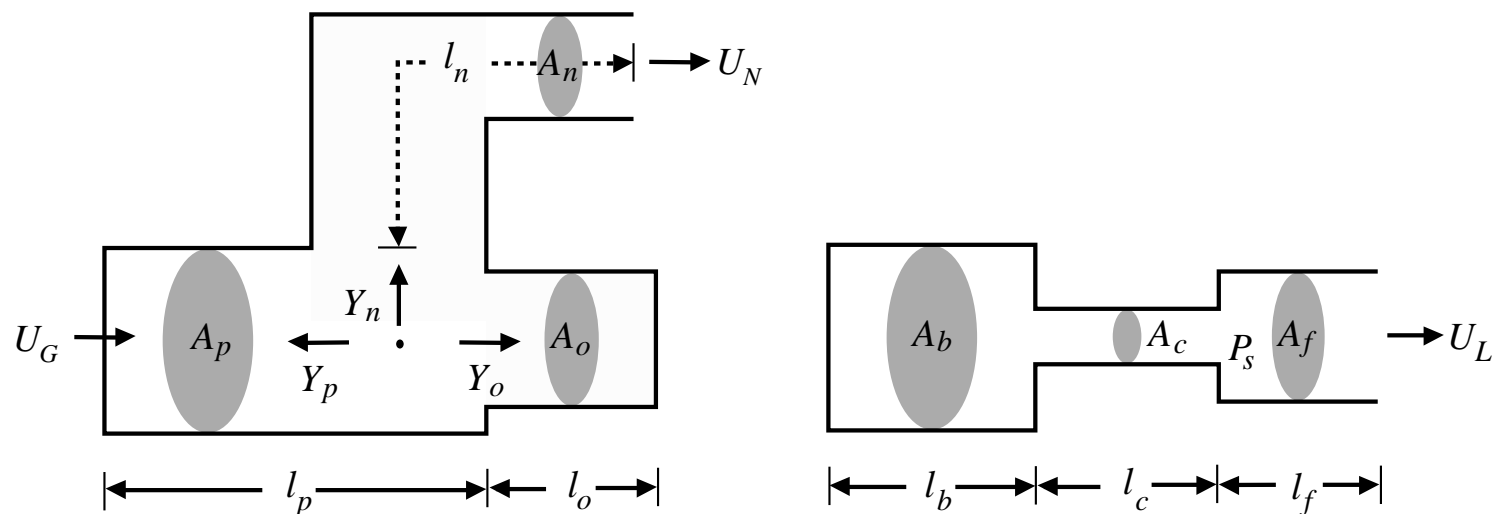
/bit/



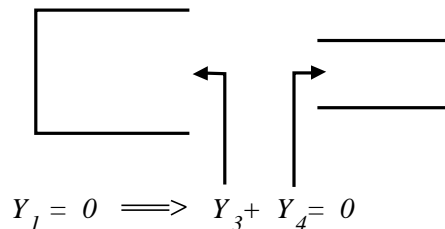
/bat/

## Cálculo de frecuencias antirresonantes (Ceros)

Los ceros se dan en frecuencias donde no hay salida mensurable.

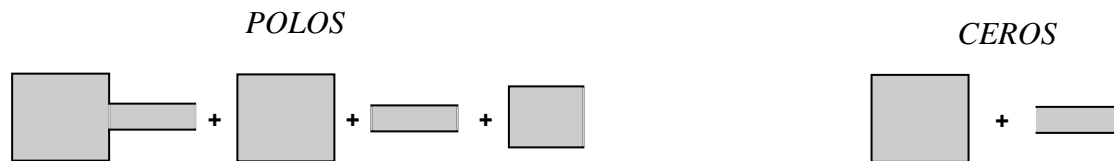
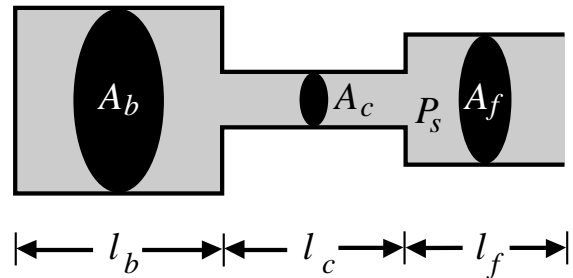


- Para consonantes nasales, los ceros en  $\mathcal{N}$  se dan donde  $Y_O = \infty$
- Para consonantes fricativas u oclusivas, los ceros en  $\mathcal{U}$  se dan donde la obstaculización tras la fuente es infinita (p.ej., **una pared dura en la fuente**).



- Los ceros se dan cuando las mediciones se realizan en el interior del tracto vocal.

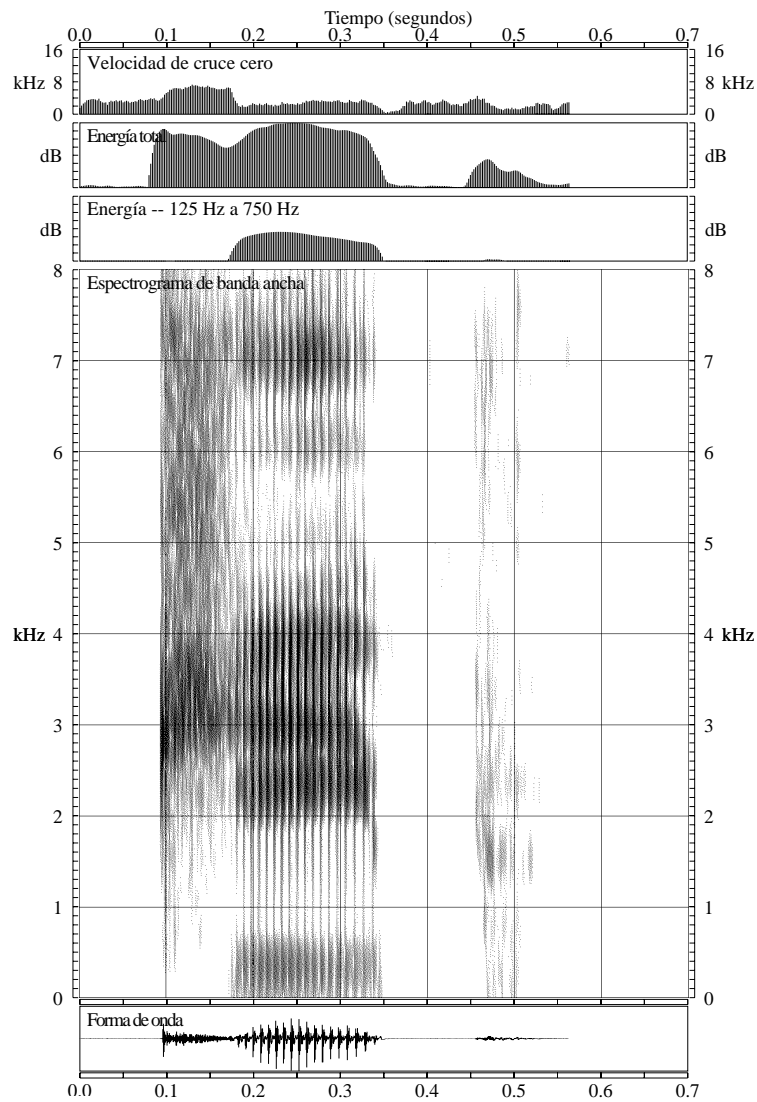
## Producción de la consonante



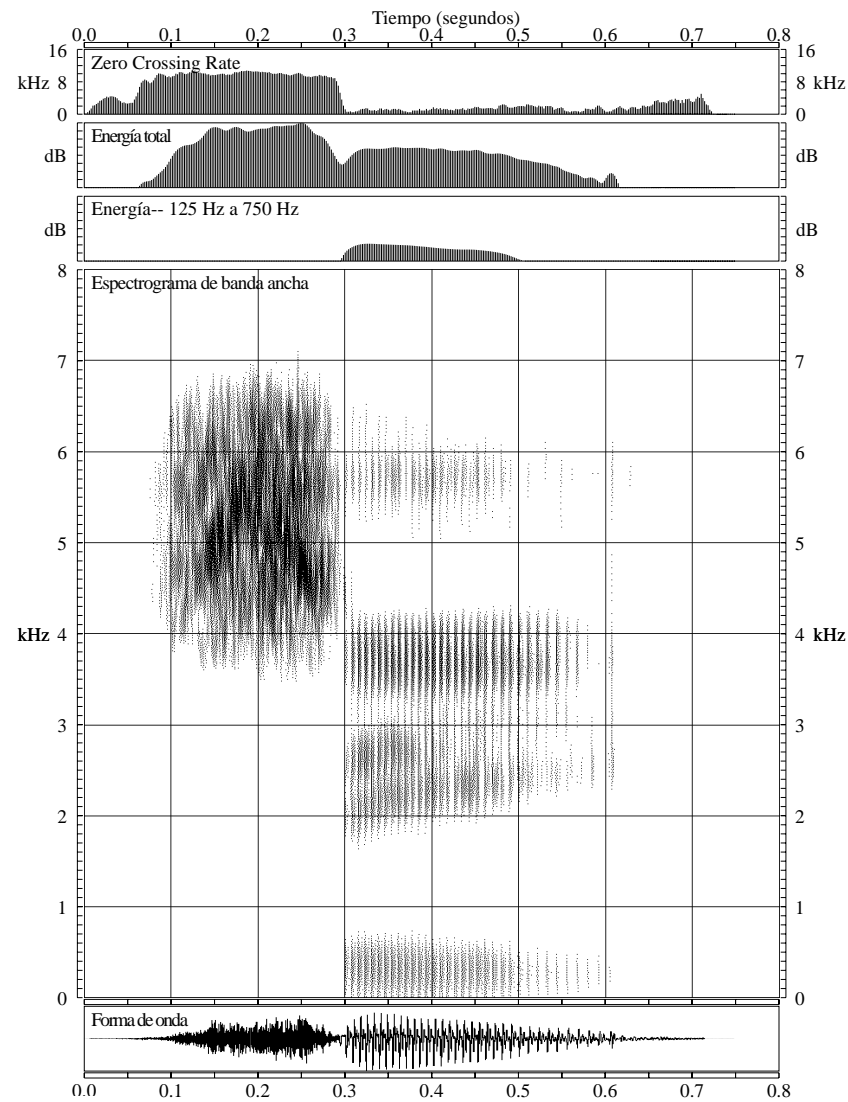
	$A_b$	$A_c$	$A_f$	$l_b$	$l_c$	$l_f$
[g]	5	0.2	4	9	3	5
[s]	5	0.5	4	11	3	2.5

[g]		[s]	
<i>polos</i>	<i>ceros</i>	<i>polos</i>	<i>ceros</i>
215	0	306	0
1750	1944	1590	1590
1944	2916	3180	2916
3888	3888	3500	3180
.	.	.	.
.	.	.	.

# Ejemplo de espectrograma de consonante



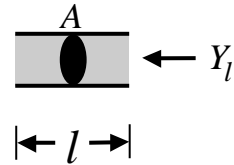
/kɪp/



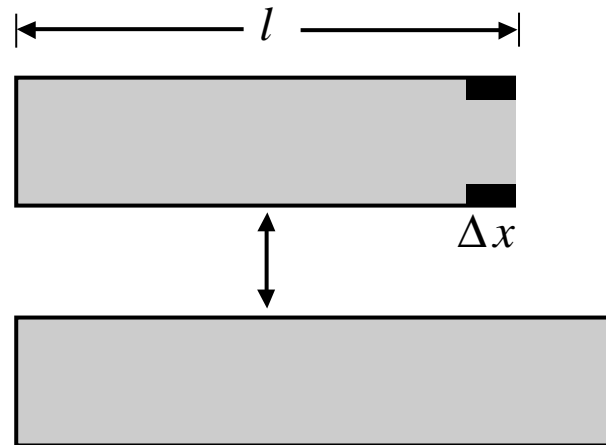
/sɪʏ/

## Teoría de la perturbación

$$Y_\ell \simeq -j \frac{A}{\Omega \rho \ell} \text{ para } \ell \text{ pequeña}$$

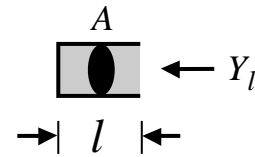


- Considere un tubo uniforme, cerrado en un extremo y abierto en el otro.

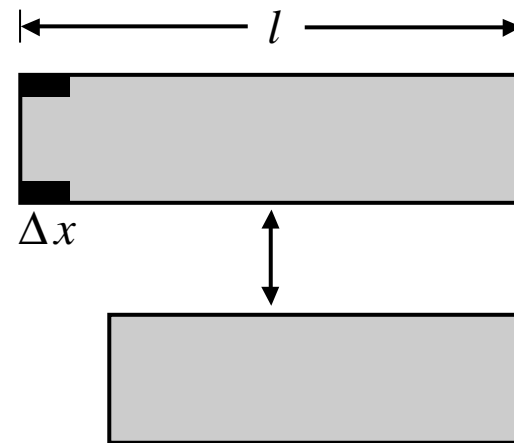


- Reducir el área de una pequeña parte del tubo cerca de la apertura (donde  $U$  alcanza el máximo), causa el mismo efecto que mantener el área fija y alargar el tubo.
- Dado que alargar el tubo disminuye las frecuencias resonantes, estrechar el mismo cerca de los puntos donde  $U(x)$  alcanza el máximo en la posición del patrón de onda para un determinado formante, reduce el valor de dicho formante.

## Teoría de la perturbación (cont.)

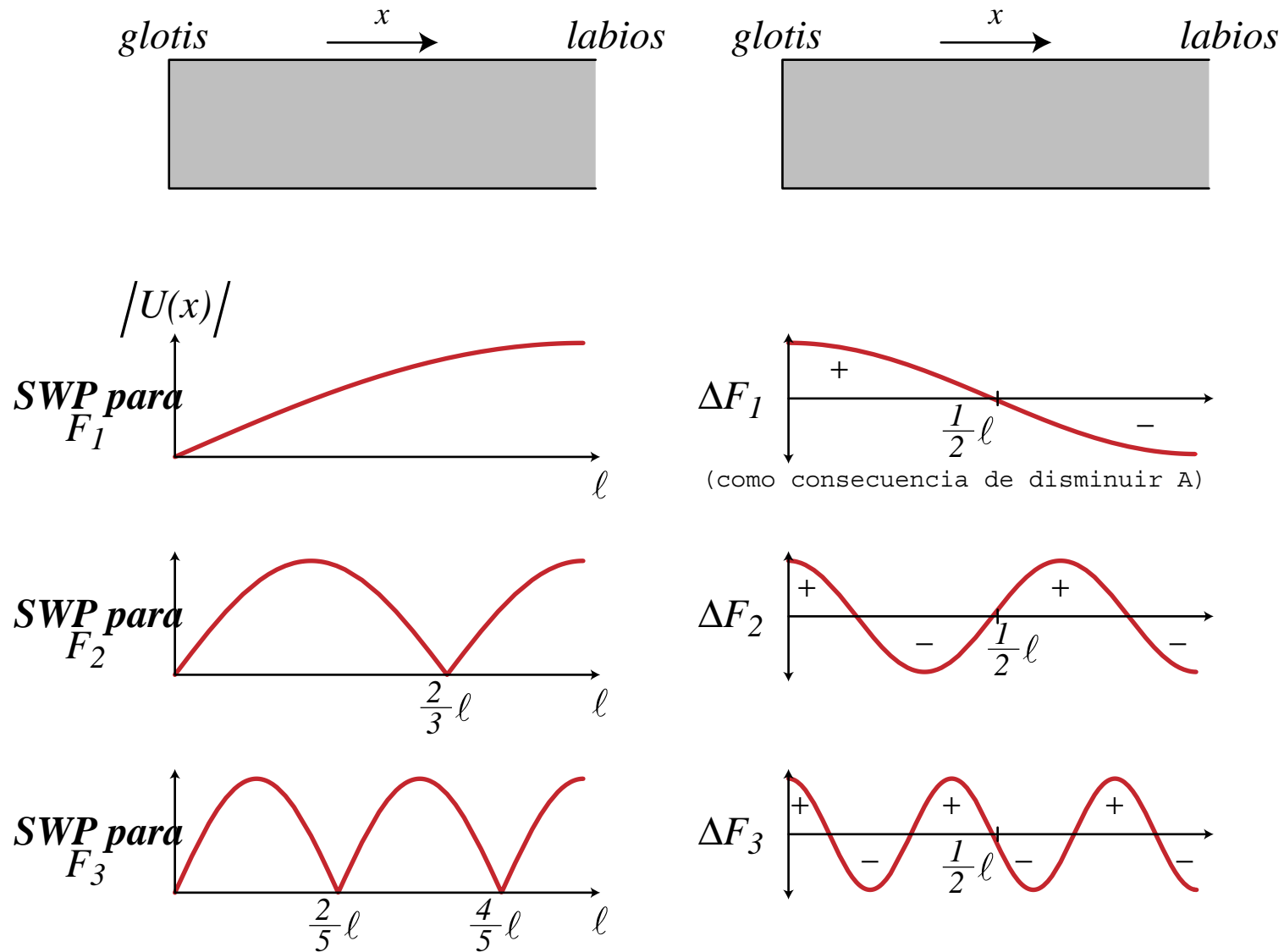


$$Y_l \approx j\Omega \frac{A\ell}{\rho c^2} \text{ para } l \text{ pequeña}$$

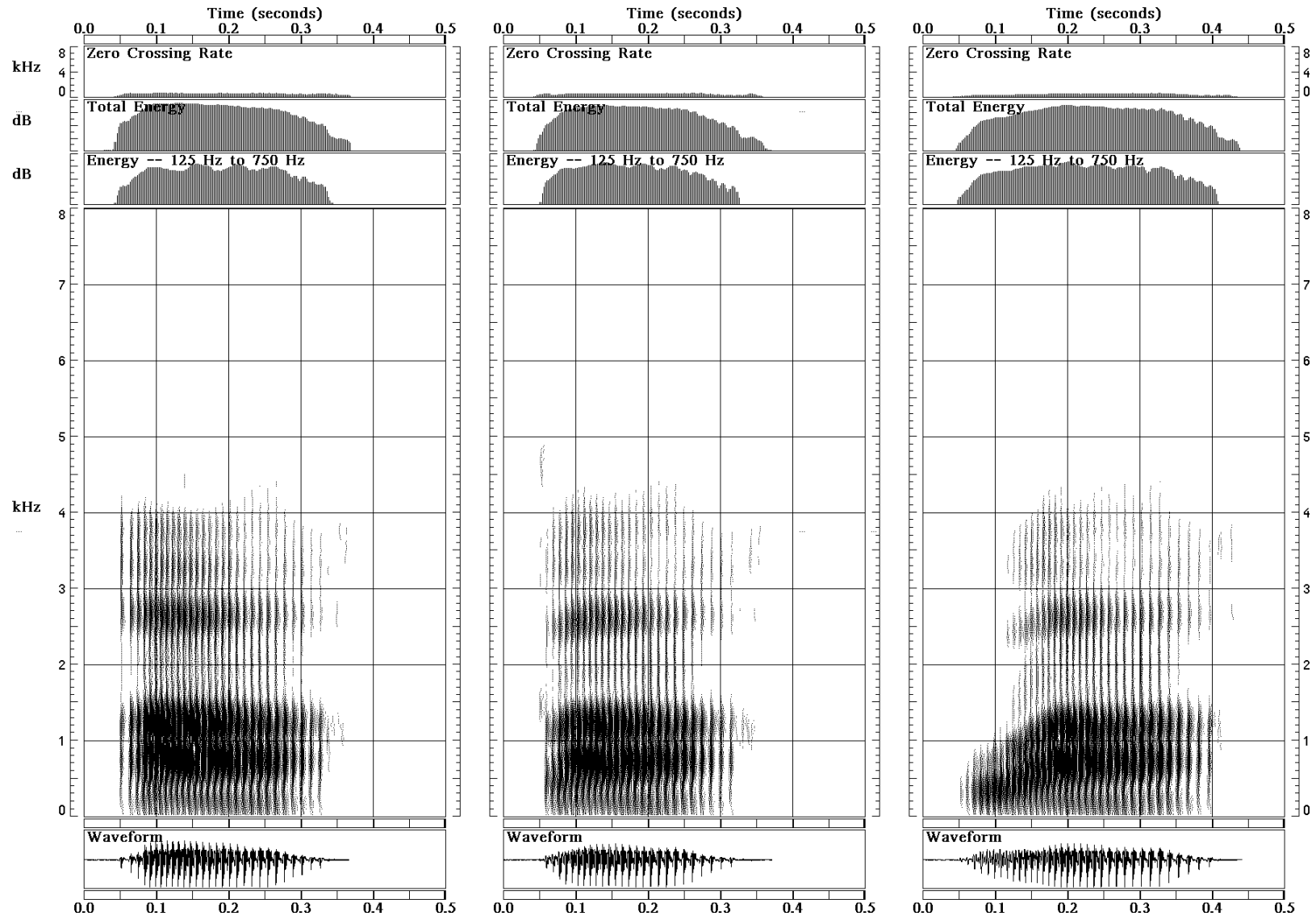


- Reducir el área de una pequeña parte del tubo próxima al cierre (donde  $p$  alcanza el máximo), posee el mismo efecto que mantener el área fija y acortar el tubo.
- Dado que acortar el tubo aumentará los valores de los formantes, estrecharlo cerca de los puntos donde  $p(x)$  alcanza el máximo en la posición del patrón de onda para un determinado formante, aumentará el valor de dicho formante.

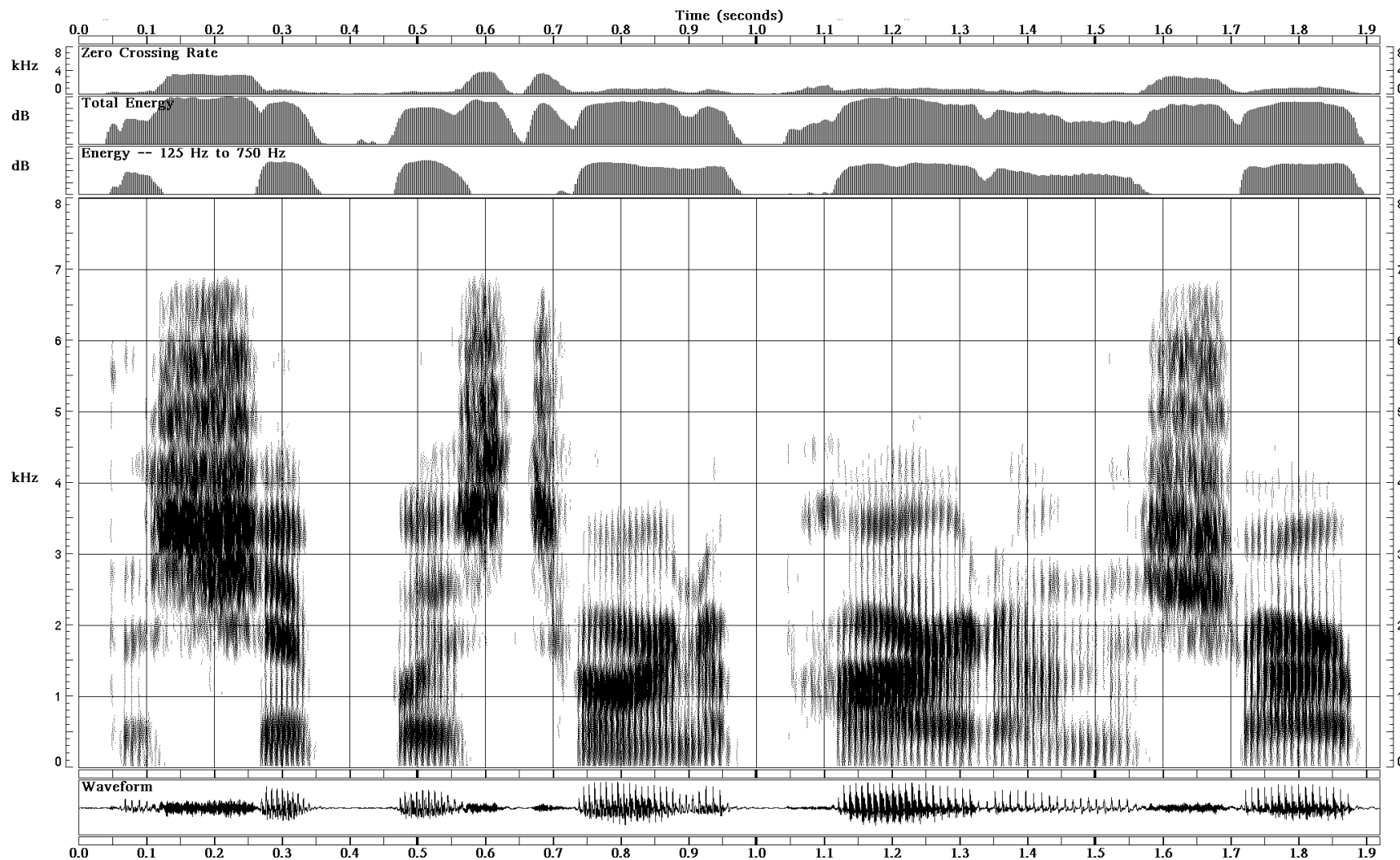
## Resumen de los resultados de la teoría de la perturbación



# Ilustración de la teoría de la perturbación

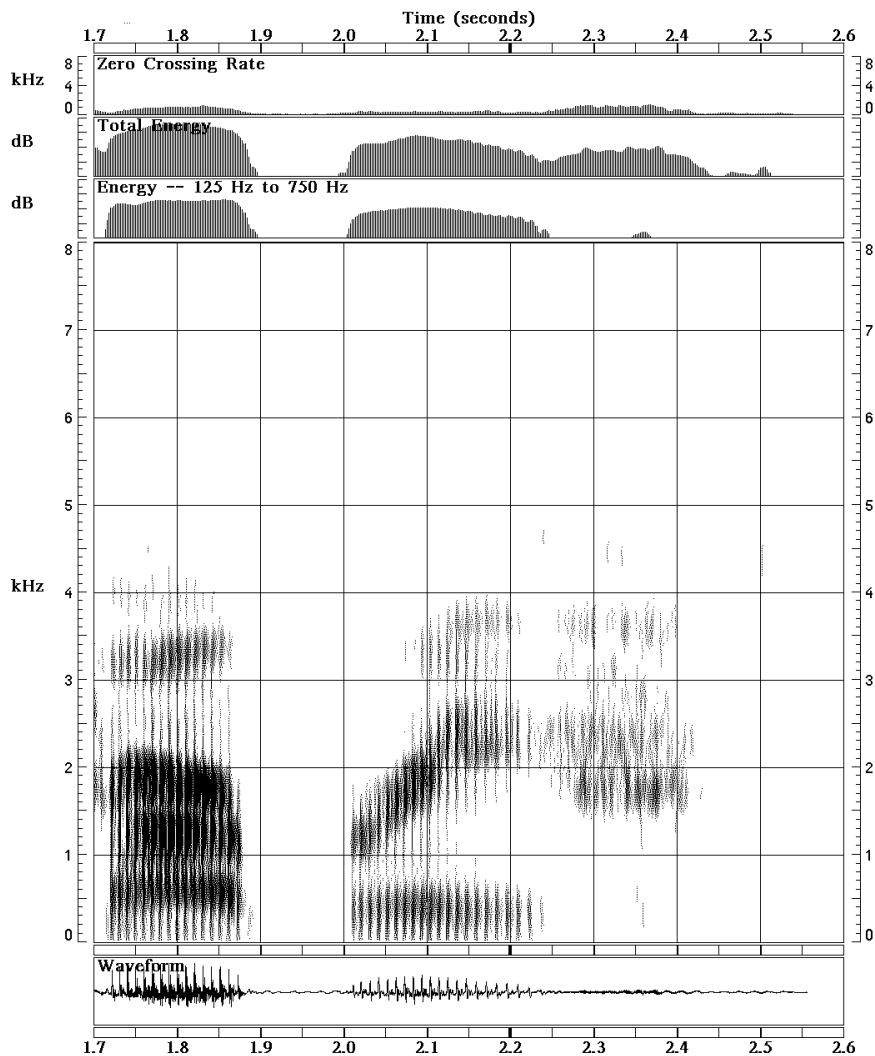


# Ilustración de la teoría de la perturbación



*The ship was torn apart on the sharp (reef)*(Desmantelaron el barco en el arrecife)

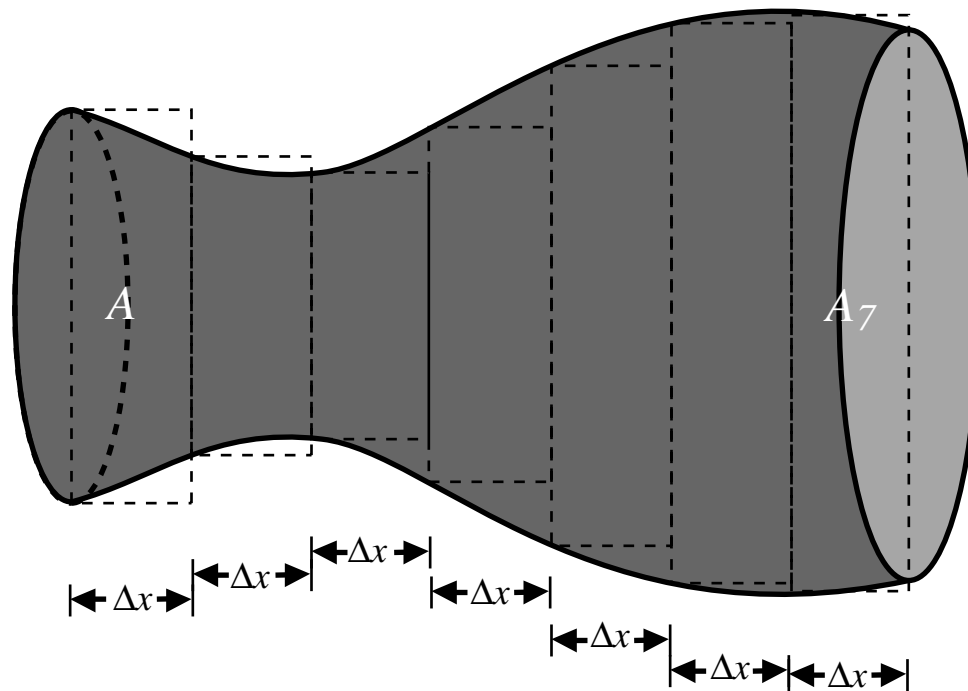
# Ilustración de la teoría de la perturbación



(The ship was torn apart on the sh)arp reef

# Aproximación de muchos tubos en el tracto vocal

- Podemos representar el tracto vocal como una concatenación de  $N$  tubos sin pérdidas, con un área constante  $\{A_k\}$  e igual longitud  $\Delta x = \ell/N$
- El tiempo de propagación de la onda por cada tubo es  $\tau = \Delta x / c = \frac{\ell}{Nc}$



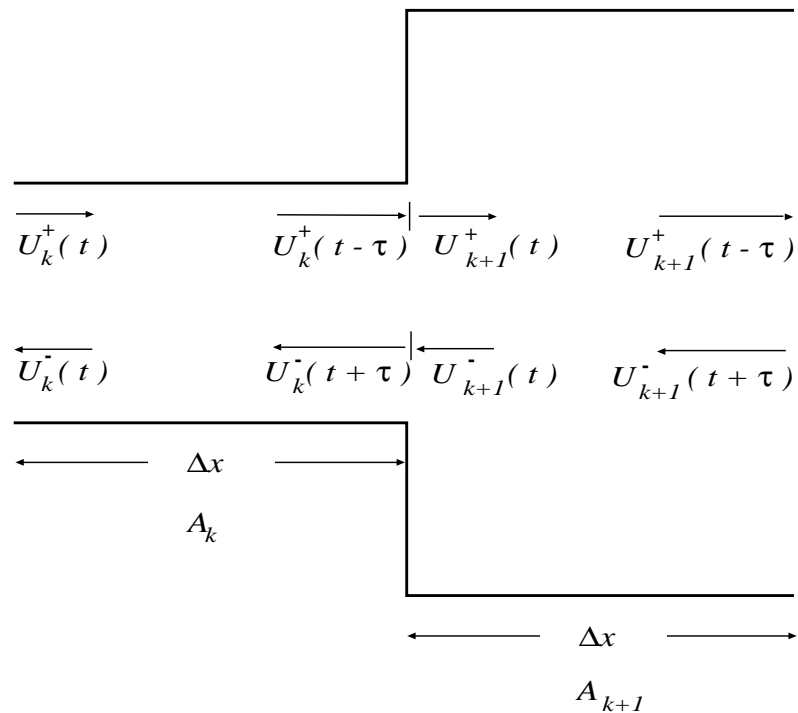
## Ecuaciones de onda para un tubo individual

Las ecuaciones de onda para el tubo  $k^{th}$  tienen la forma

$$p_k(x, t) = \frac{\rho c}{A_k} \left[ U_k^+ \left( t - \frac{x}{c} \right) + U_k^- \left( t + \frac{x}{c} \right) \right]$$

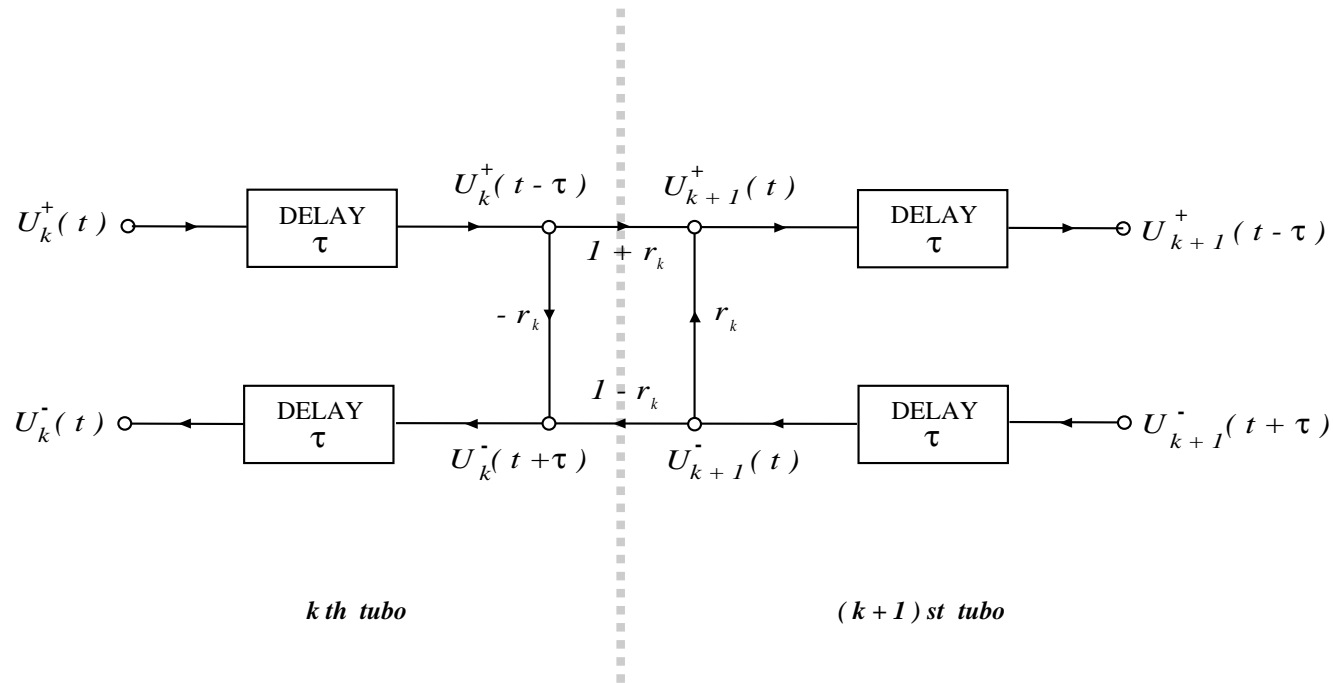
$$U_k(x, t) = U_k^+ \left( t - \frac{x}{c} \right) - U_k^- \left( t + \frac{x}{c} \right)$$

donde  $x$  se mide desde el lado de la izquierda ( $0 \leq x \leq \Delta x$ )



# Expresión de actualización en las fronteras del tubo

Podemos resolver expresiones de actualización utilizando las restricciones de continuidad en las fronteras del tubo, ej.,  $p_k(\Delta x, t) = p_{k+1}(0, t)$ , y  $U_k(\Delta x, t) = U_{k+1}(0, t)$



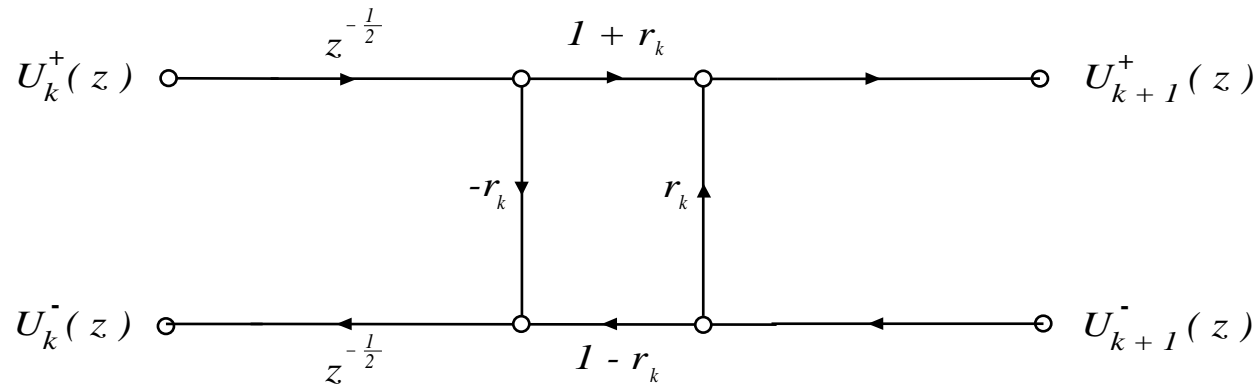
$$U_{k+1}^+(t) = (1 + r_k)U_k^+(t - \tau) + r_k U_{k+1}^-(t)$$

$$U_k^-(t + \tau) = -r_k U_k^+(t - \tau) + (1 - r_k)U_{k+1}^-(t)$$

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \quad \text{nota } |r_k| \leq 1$$

# Modelo digital del tracto vocal con muchos tubos

- Las actualizaciones en las fronteras del tubo se dan sincrónicamente cada  $2\tau$
- Si la excitación se limita a la banda, se pueden muestrear las salidas cada  $T = 2t$
- Cada sección del tubo posee un retraso de  $z^{-1/2}$



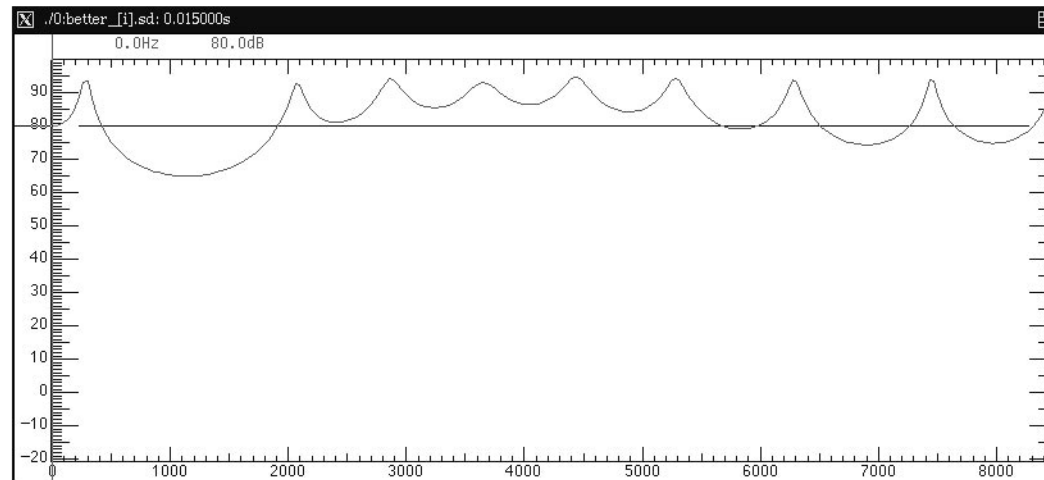
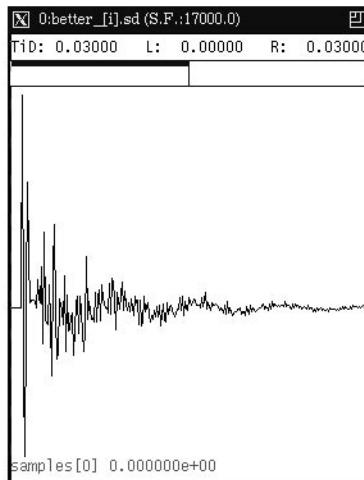
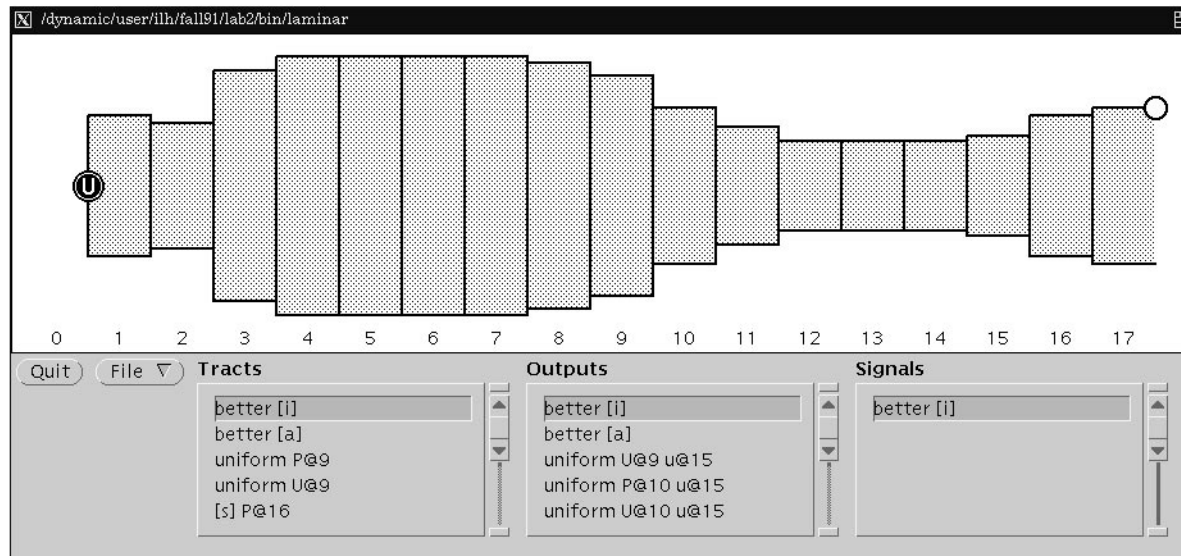
- La elección de  $N$  depende de la velocidad de muestreo  $T$

$$T = 2\tau = 2 \frac{\ell}{Nc} \quad \Rightarrow \quad N = \frac{2\ell}{cT}$$

- Las series y las pérdidas relegadas se pueden introducir también en las uniones del tubo
  - Los anchos de banda son proporcionales a la pérdida de energía en la velocidad de almacenamiento
  - La energía almacenada es proporcional a la longitud del tubo

# MIT

## Trabajo 1



- Zue, *Material del curso 6.345*
- Stevens, *Acoustic Phonetics*, MIT Press, 1998.
- Rabiner & Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.