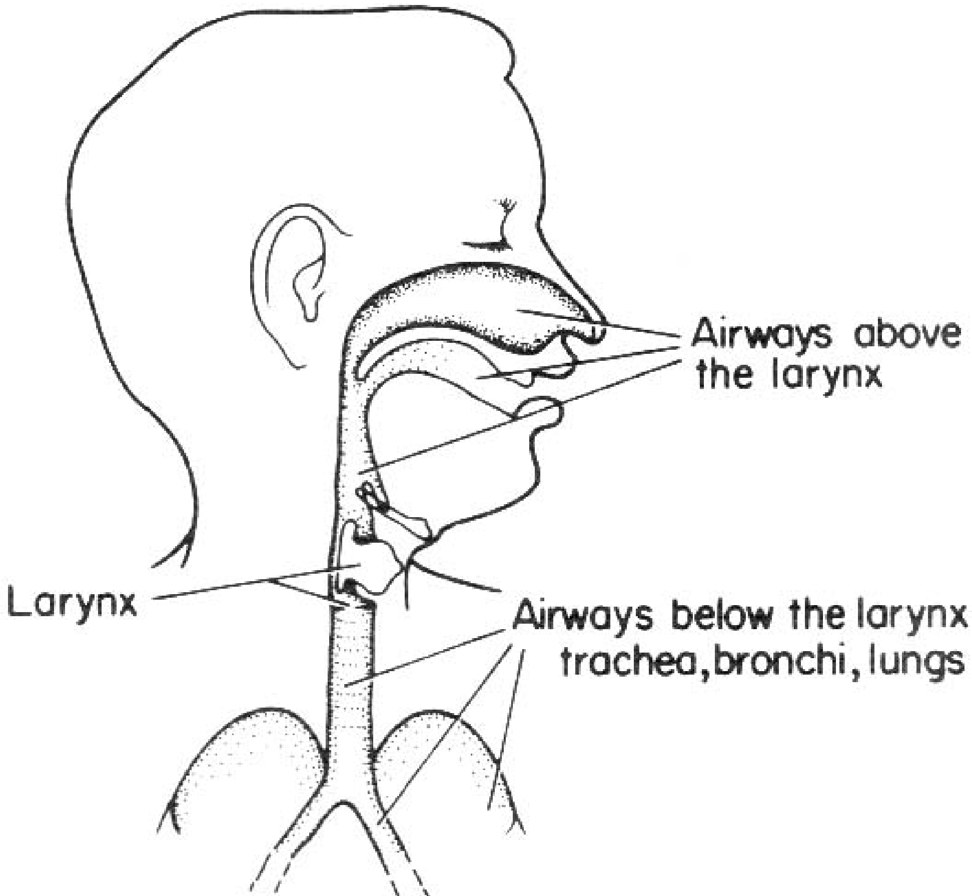
04 Speech Production & Sounds in Languages

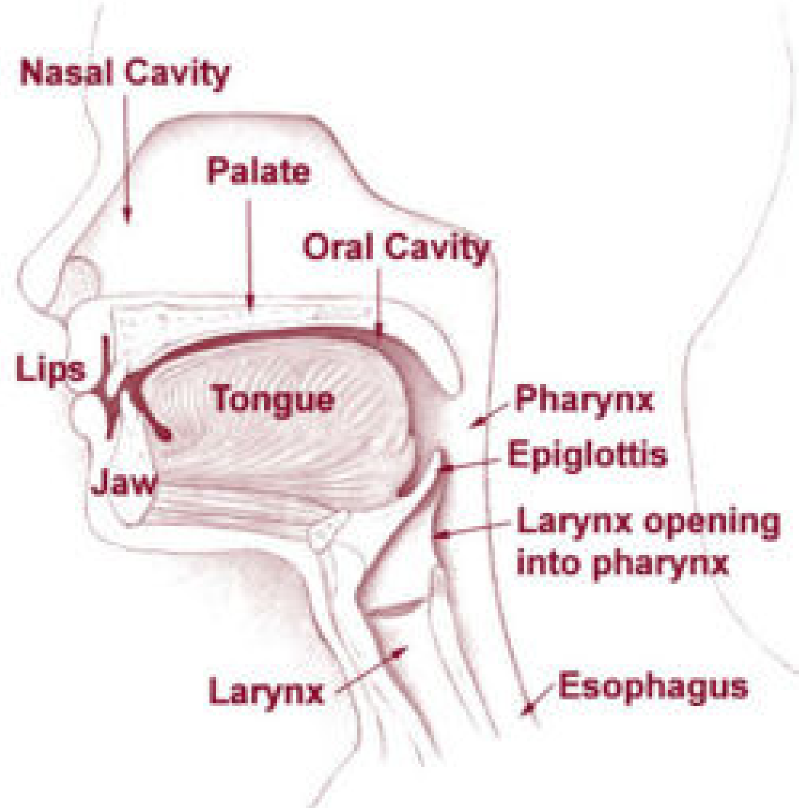
This Lecture

- Physiology of the human speech production mechanism
  - How we generate speech
- A “Source-Filter” model of human speech production
  - A link between what happens inside your mouth and the speech that comes out of it
- Organization of sounds in languages

Speech Production System



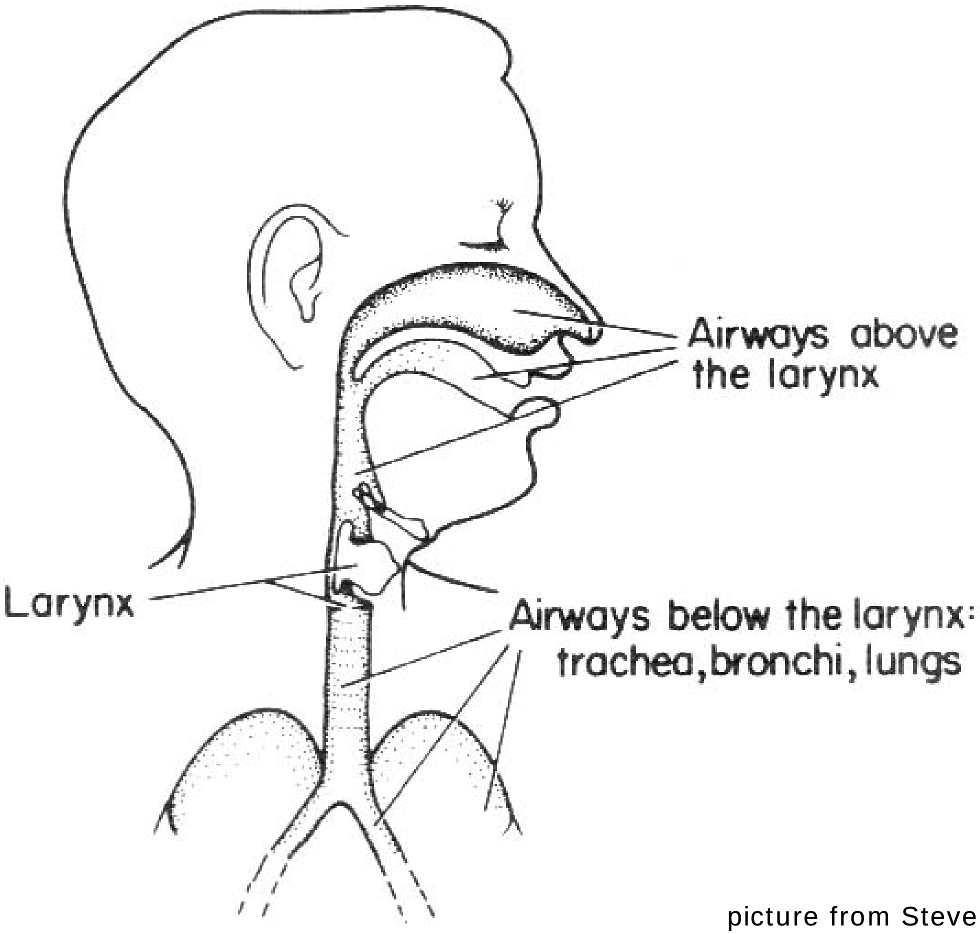
picture from Stevens 1999



en.wikipedia.org

Speech Production System

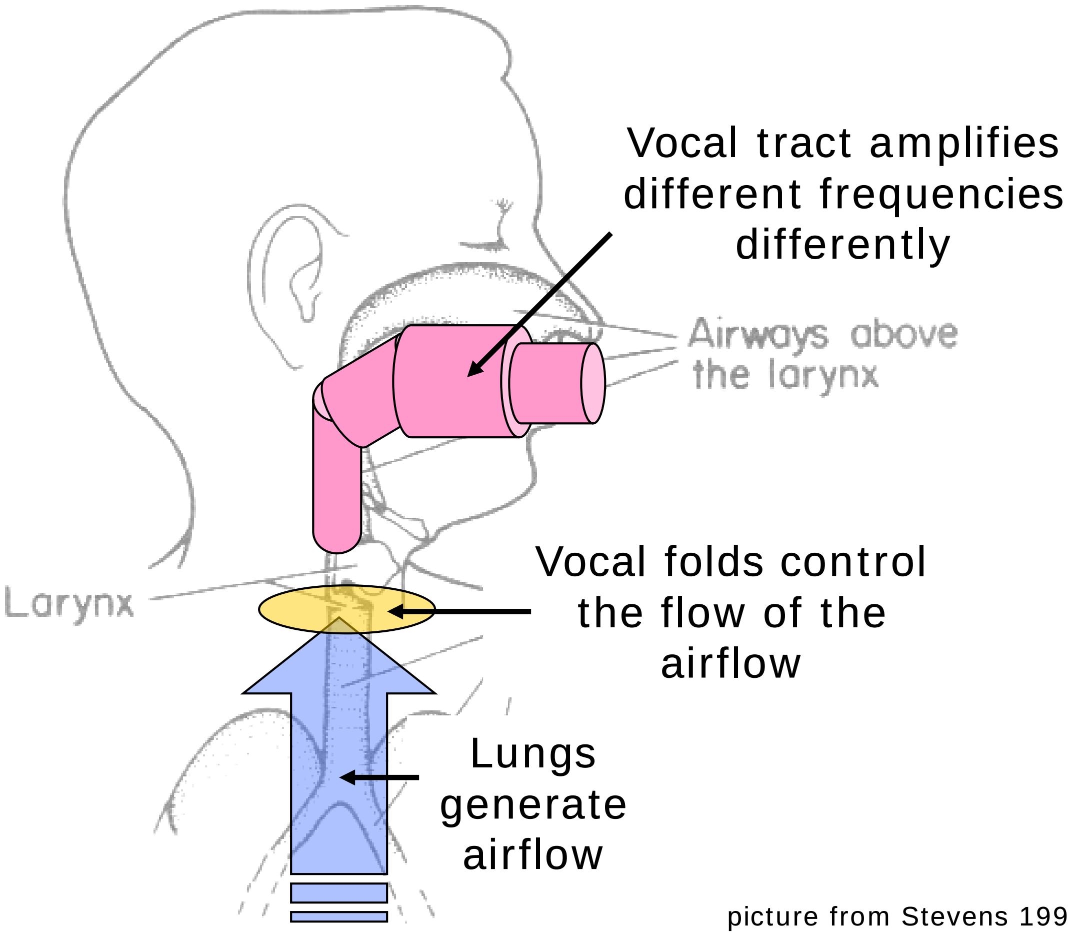
- 3 parts
  - below the larynx
    - the subglottal system
  - the larynx
    - the vocal folds
  - above the larynx
    - the supraglottal system



picture from Stevens 1999

Speech Production System

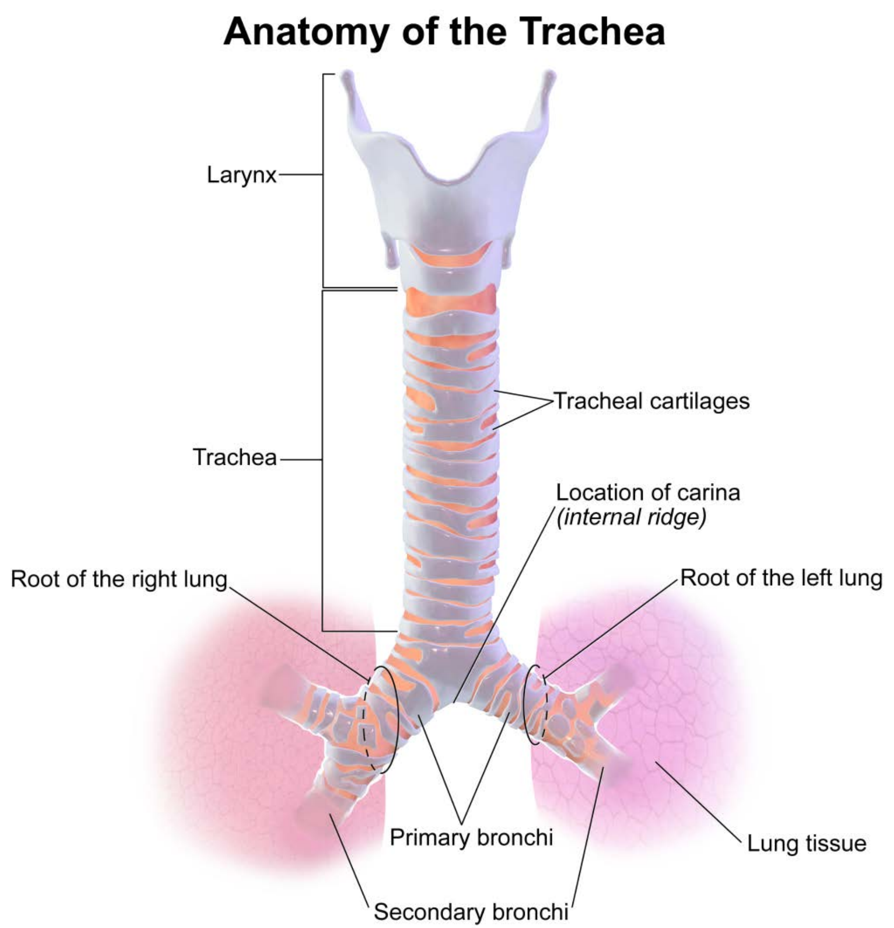
- 3 parts
  - below the larynx
    - the subglottal system
  - the larynx
    - the vocal folds
  - above the larynx
    - the supraglottal system



picture from Stevens 1999

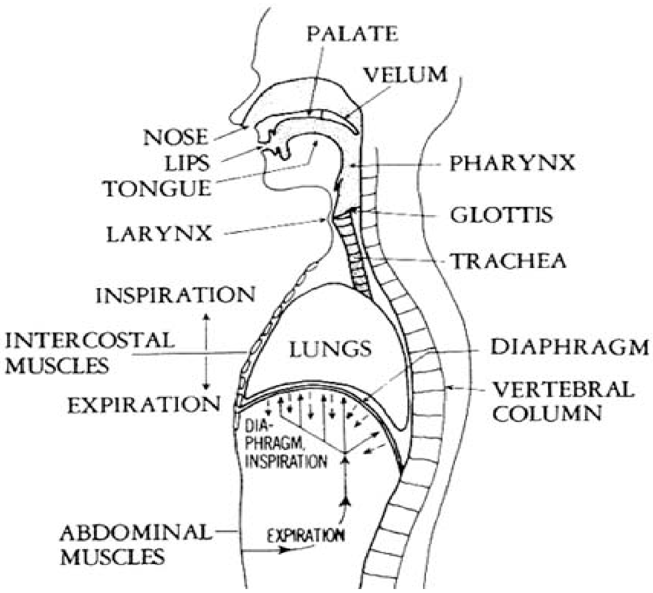
Anatomy of the Subglottal System

- below the larynx
- the trachea
  - 2.5 cm^2 cross-sectional area
  - 10 to 12 cm length for an adult speaker
- Two bronchi
  - each with one-half the cross-sectional area of the trachea
  - turn into series of smaller airways terminated in the lungs
- Provide airflow for the production of speech



https://upload.wikimedia.org/wikipedia/commons/4/4c/Blausen_0865_TracheaAnatomy.png

Generating the airflow

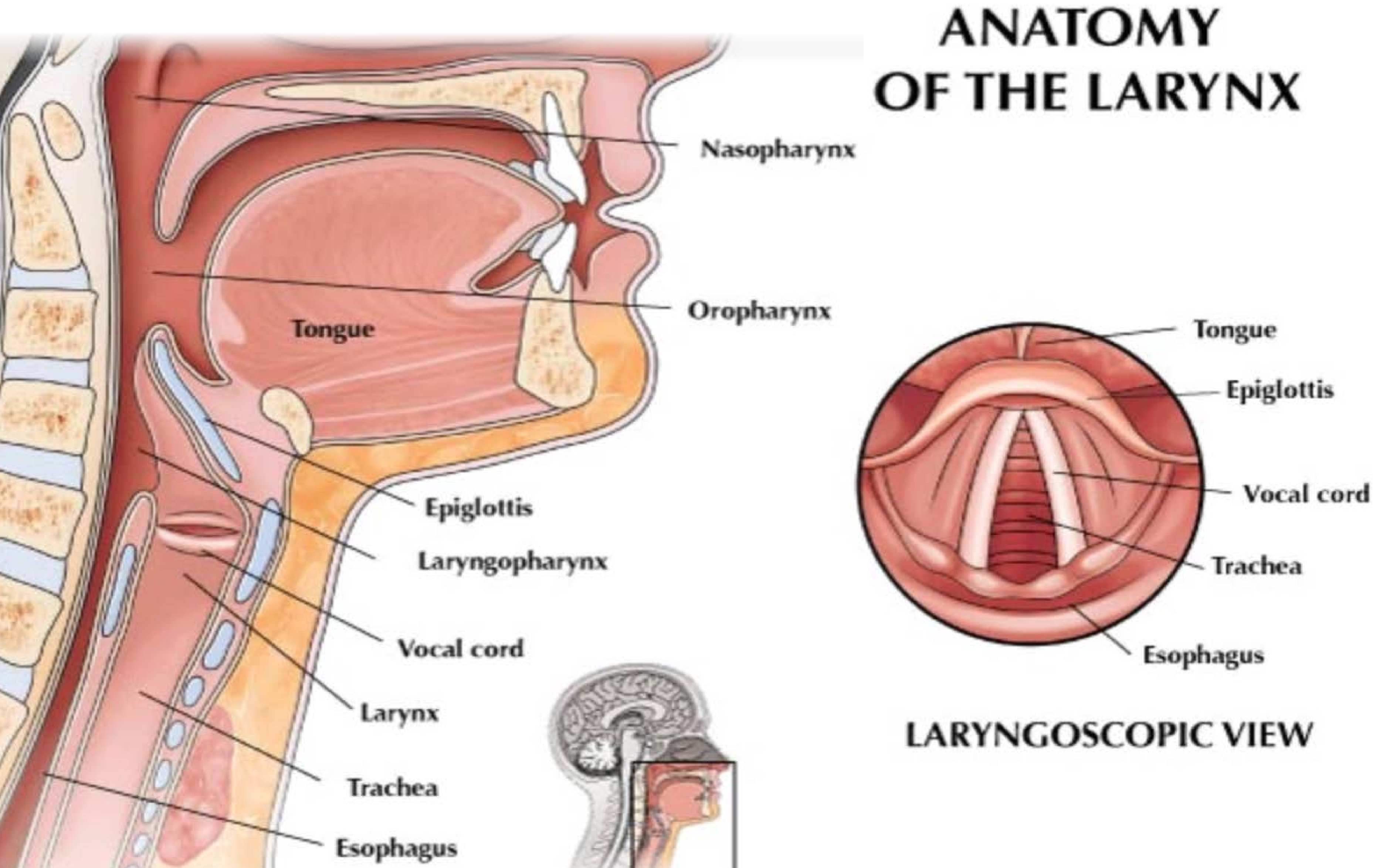


schematic representation of the methods for controlling respiration

picture from Stevens 1999

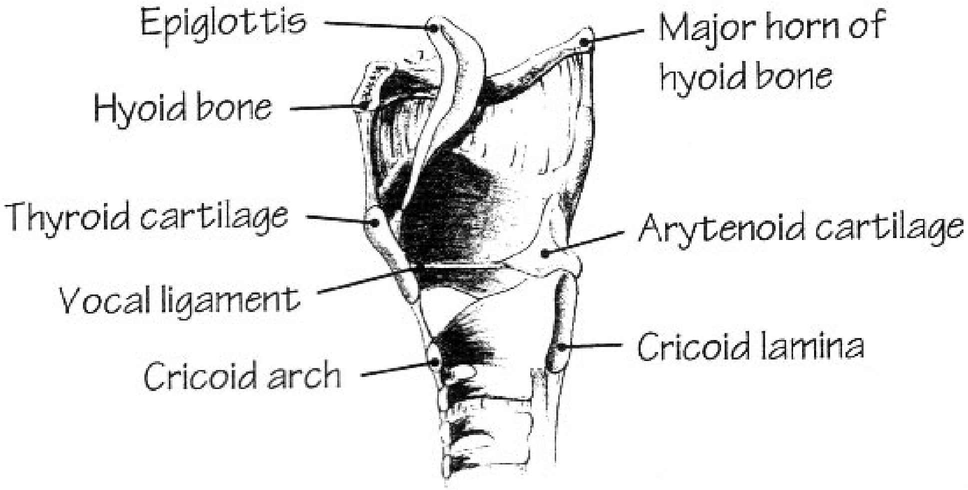
- mechanical movements of the respiratory system are controlled by:
  - diaphragm
  - abdominal muscles
  - chest wall
- the movement changes the lung’s volume
- reflected in the change in the pressure of the air in the lung

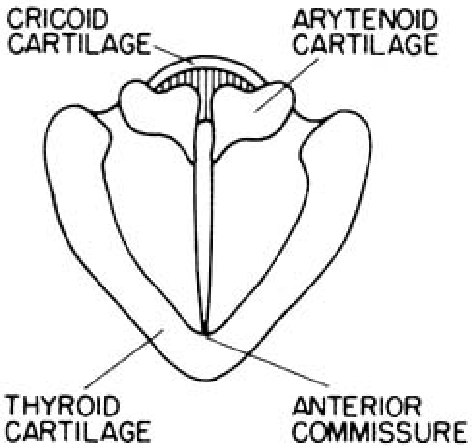
http://cdn.transgenderuniverse.com/wp-content/uploads/2016/03/vocal-feminization-larynx.png



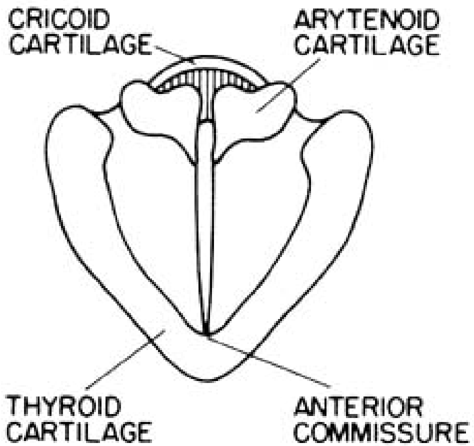
The Larynx

- principle structure for speech production are the “vocal folds” (Also known as “Vocal Cords”)
  - two bands of tissue of length 1.0-1.5 cm
  - thickness of 2-3 mm
  - The two vocal folds are arranged roughly parallel to each other in an anteroposterior direction.

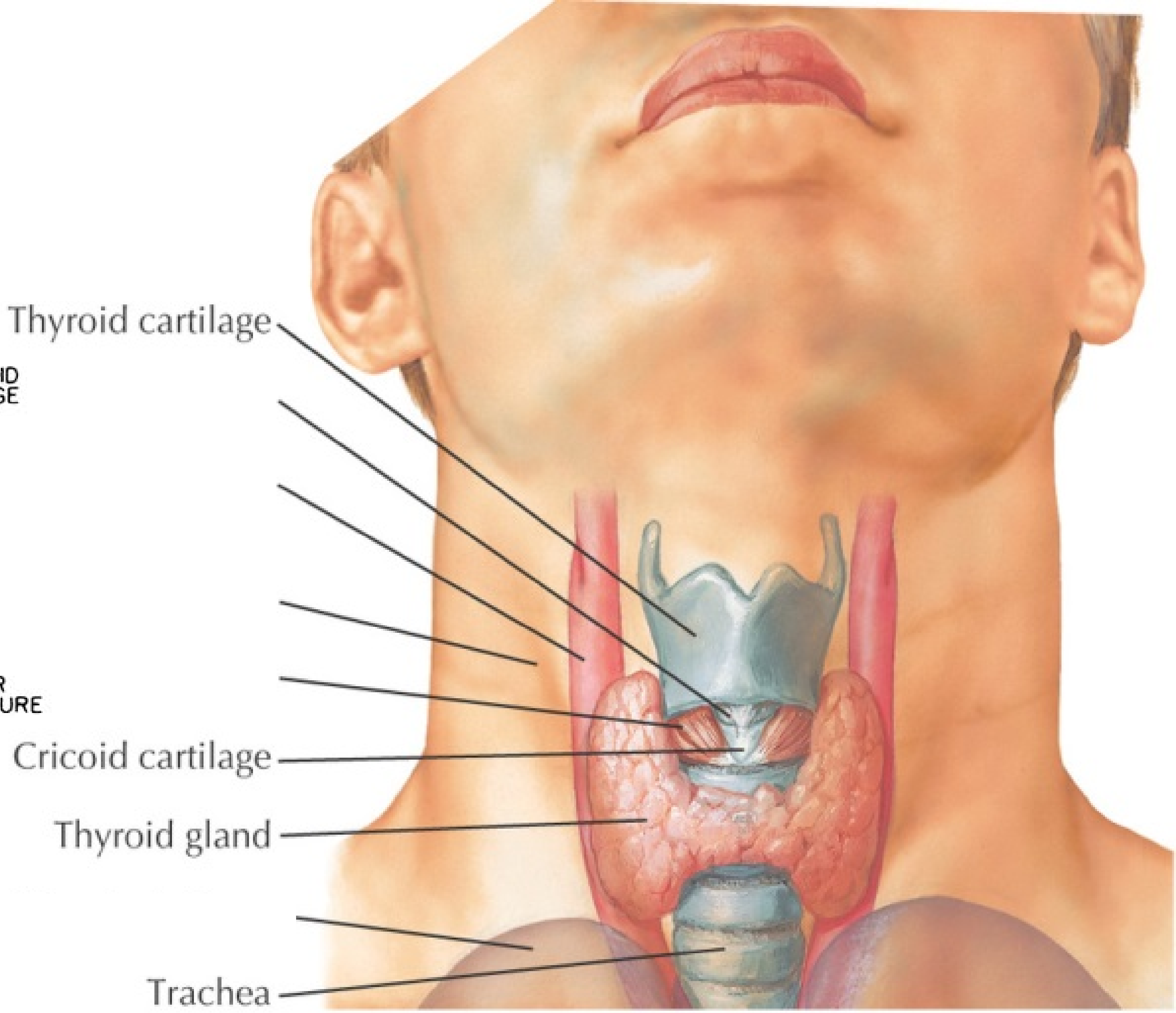




pictures from Stevens 1999



Stevens 1999



http://www.voicescienceworks.org/uploads/5/2/6/0/52608601/48528_orig.jpg

Ep.9: Singing Basics: How do Vocal Cords Work?

Power To Sing

Subscribe 11,770

https://www.youtube.com/watch?v=P2pLJfWUjc8

Retrieved: Feb 8th, 2017

613,092 views

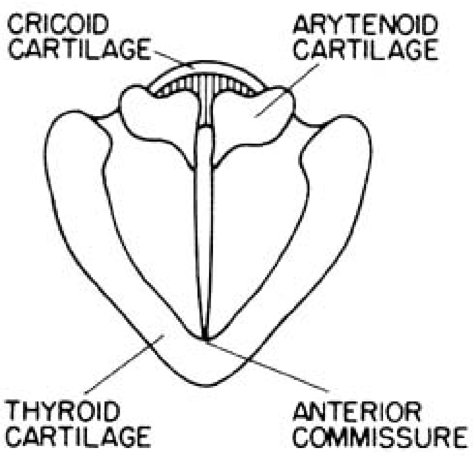
Add to   Share   More   1,774   131

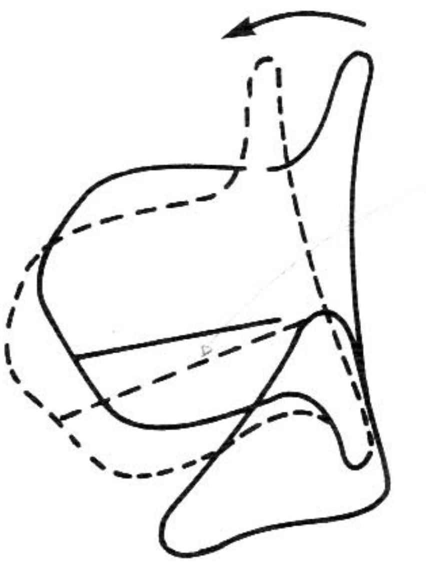
Vocal Folds

- position and stiffness controlled by some surrounding muscles
- “slack” vocal folds → easy to vibrate
- “strict” vocal folds → hard to vibrate

“Feel” Your Vocal Folds

Try ‘sssss…’ vs. ‘zzzzz…’

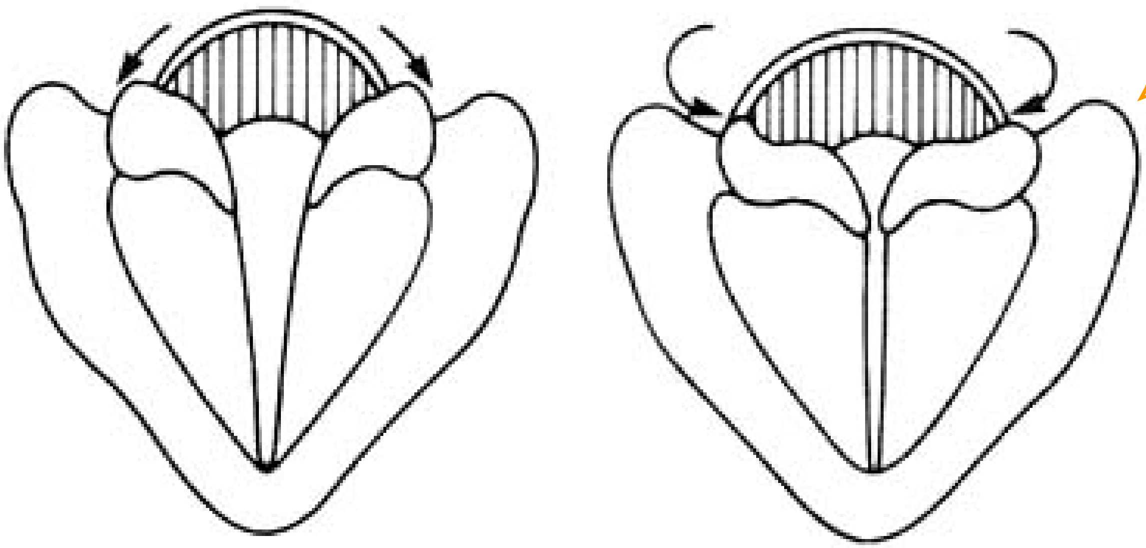




picture from Stevens 1999

Vocal Folds

- abduction and adduction

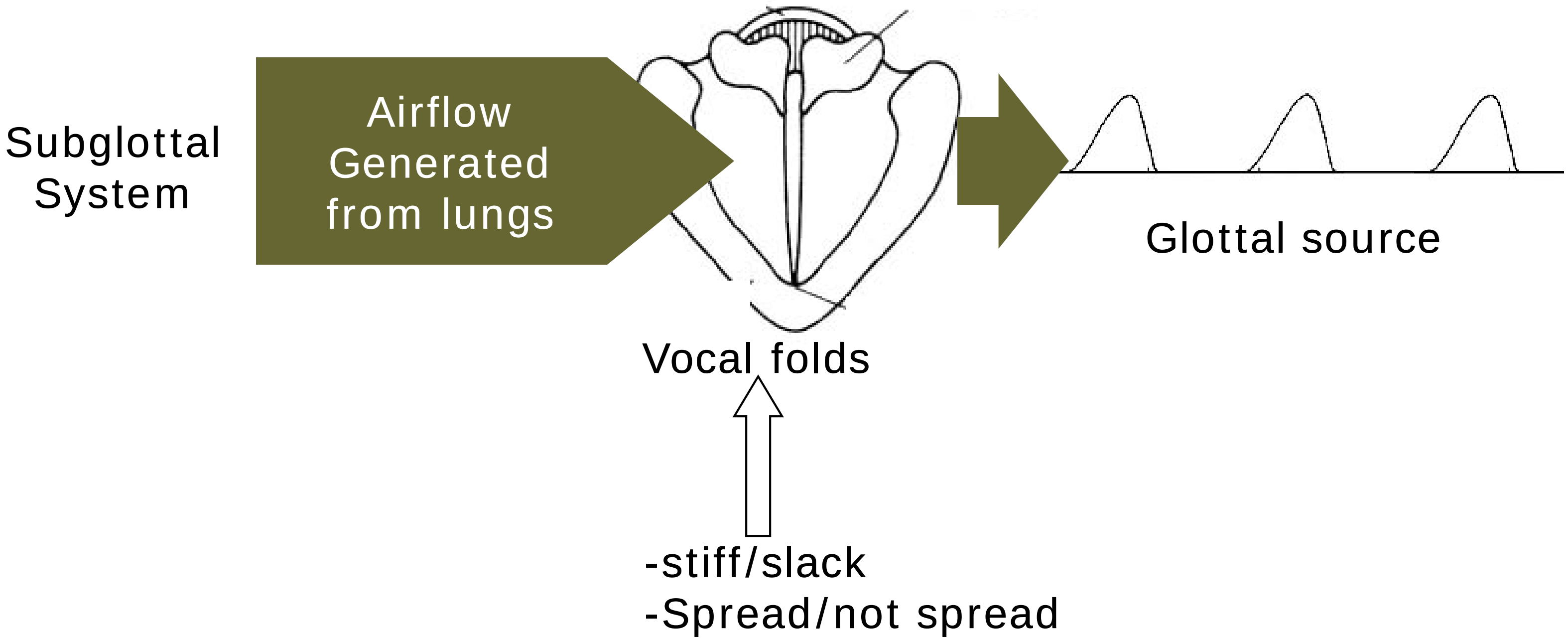


“Feel”
Your Vocal Folds
Try ‘haaaaa’

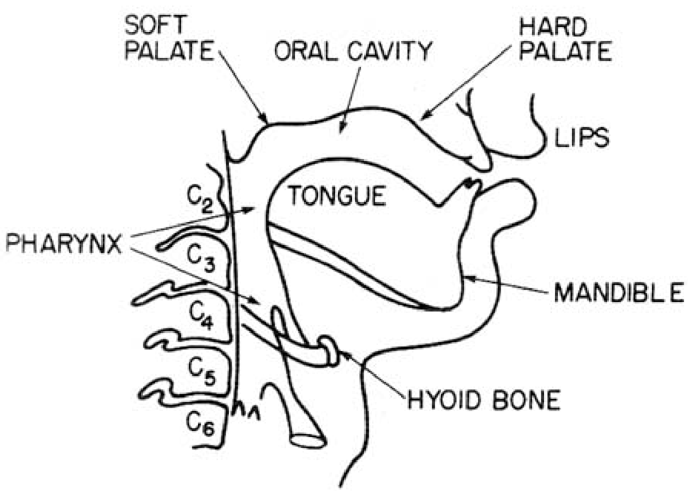
picture from Stevens 1999

The glottal source

- pressure from the lungs + position/stiffness of the vocal folds



The Supraglottal system



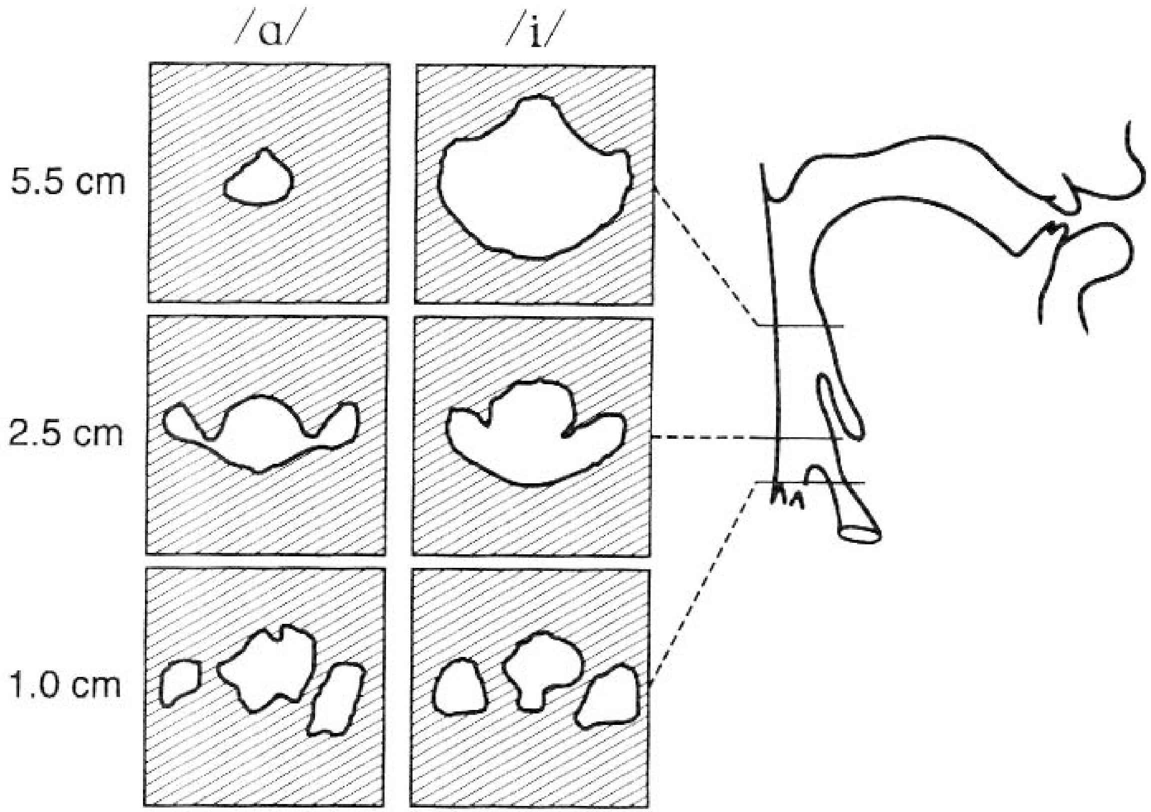
midsagittal section of the vocal tract and surrounding structures

picture from Stevens 1999

- Three pharyngeal constrictor muscles extending from the laryngeal region to the soft palate
- contraction of these muscles produces wide range of shapes and cross-sectional area of the pharyngeal airway

> “Knowledge of the
>
> **cross-sectional area**
>
> is important for determining the aerodynamic and acoustic behavior of the vocal tract”

Pharynx



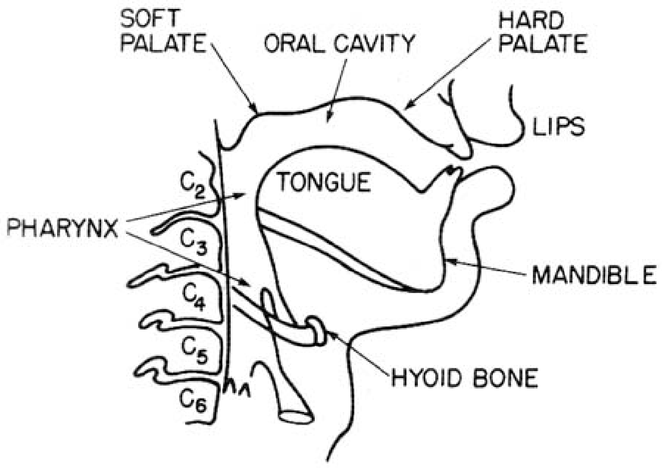
Try
'อา' vs. 'อี'

picture from Stevens 1999

cross-sectional shape of the airway at three different positions in the pharyngeal region

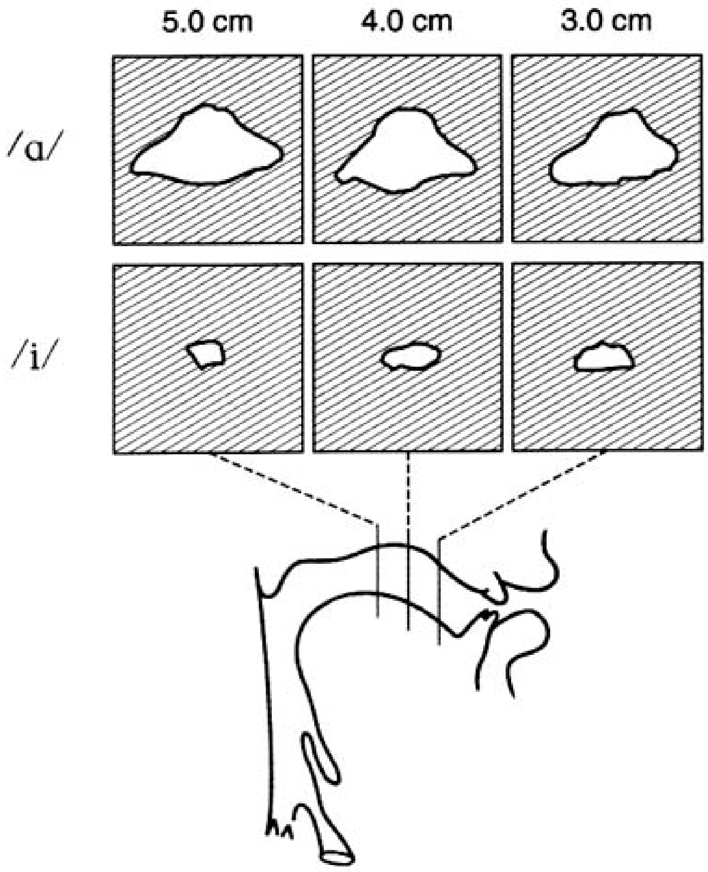
Oral Cavity

- displacement of the tongue body forward and backward changes vocal tract area in the pharyngeal region
- raising and lowering tongue body change vocal tract area in the oral cavity



picture from Stevens 1999

Oral Cavity

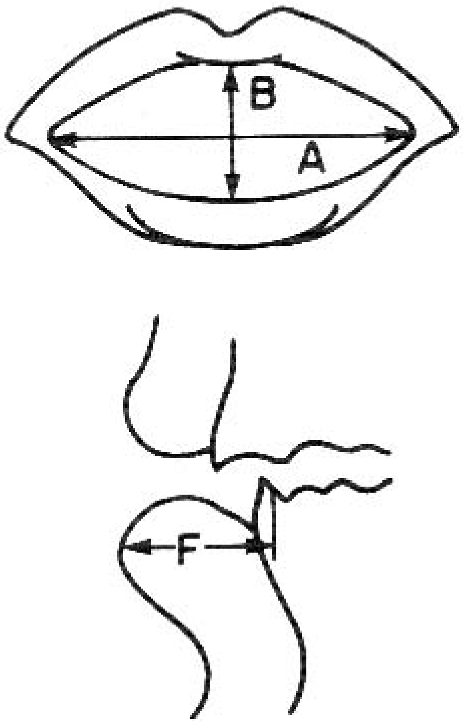


cross-sectional shape of the airway

picture from Stevens 1999

Try
'อา' vs. 'อี'

Lips



dimensions of the lips

- The spread and rounding of the lips affect the properties of sound radiated from the mouth opening.

picture from Stevens 1999

Vocal Tract Length

- glottis to the lips opening
- depends on the position of the larynx and the configuration of the lips

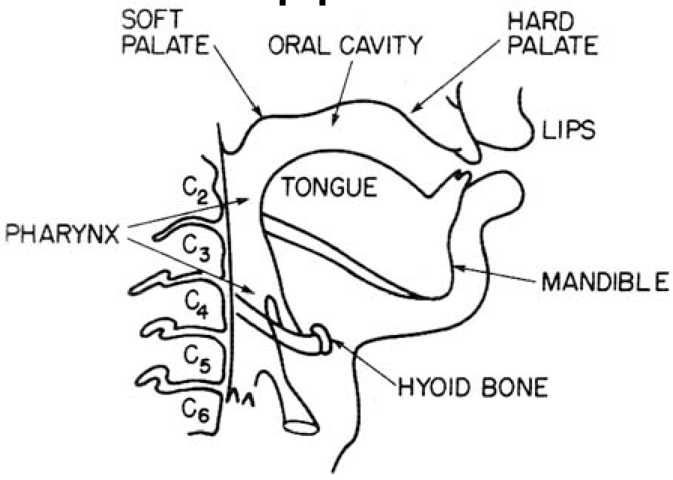
	Adult female	Adult male
Vocal Tract Length	14.1 cm	16.9 cm
Pharynx Length	6.3 cm	8.9 cm
Oral Cavity Length	7.8 cm	8.1 cm

- pharynx length:cavity length ratio is much less in children

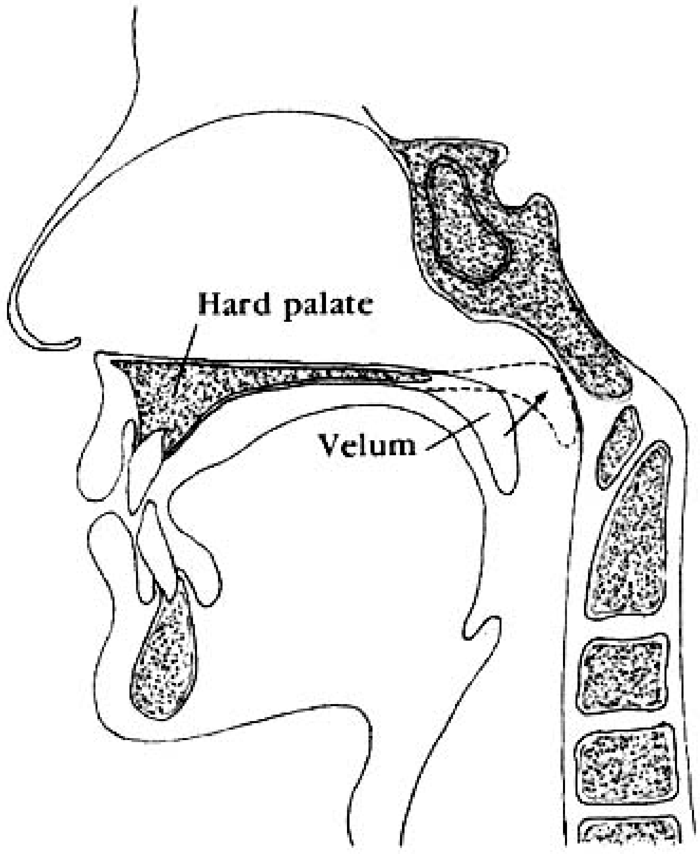
Do you think you can make your vocal tract longer or shorter?

Soft Palate and Nasal Cavity

- upper end of the pharynx extends vertically into the nasal cavity
- the opening into the nasal cavity called the “velopharyngeal opening”
- controlled by the position of the soft palate (velum) and the lateral wall of the upper pharynx



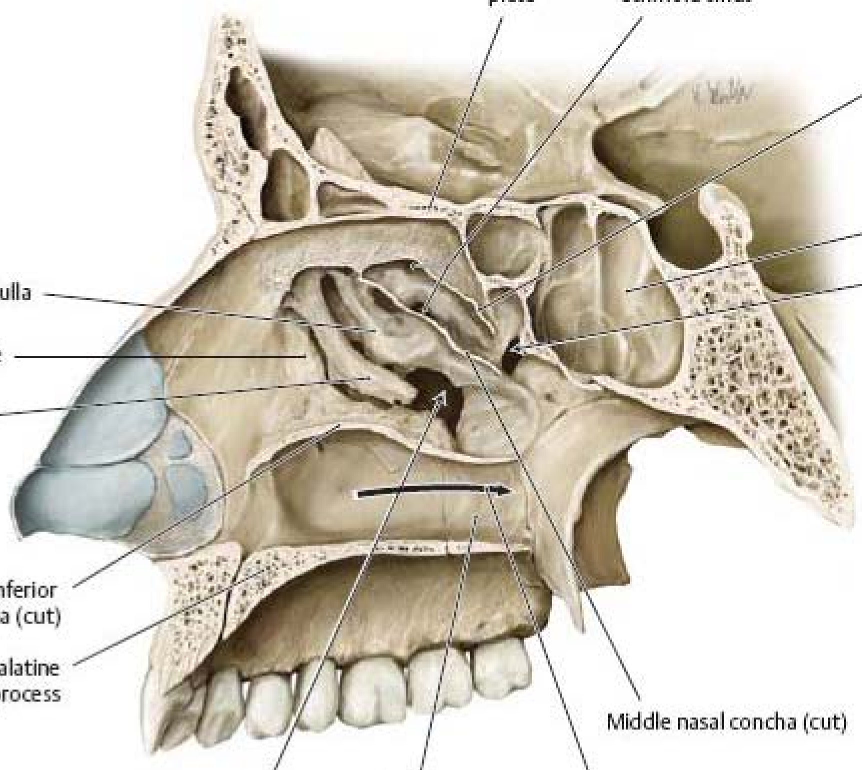
Velopharyngeal Opening



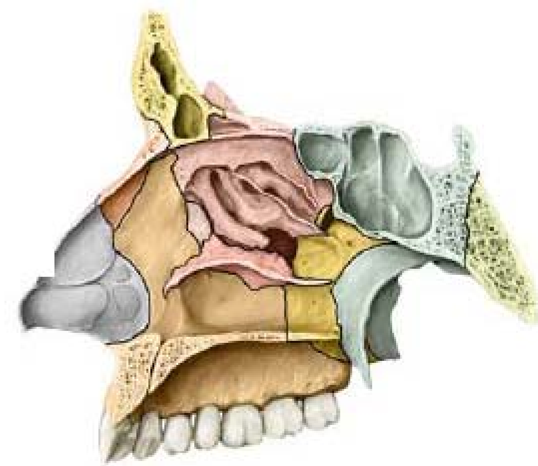
lateral view
picture from Stevens 1999

- raising and lowering of the soft palate
- cross-sectional area of as large as 1.0 cm^2
- 0.2-0.8 cm^2 when producing sounds that require the participation of the nasal cavity

Try
‘กา’ vs. ‘งา’



C Lateral wall of the right nasal cavity with the conchae removed. Sagittal section, medial view. *Revealed:* Paranasal sinuses (p. 522).



http://doctorlib.info/medical/anatomy/anatomy.files/image931.jpg

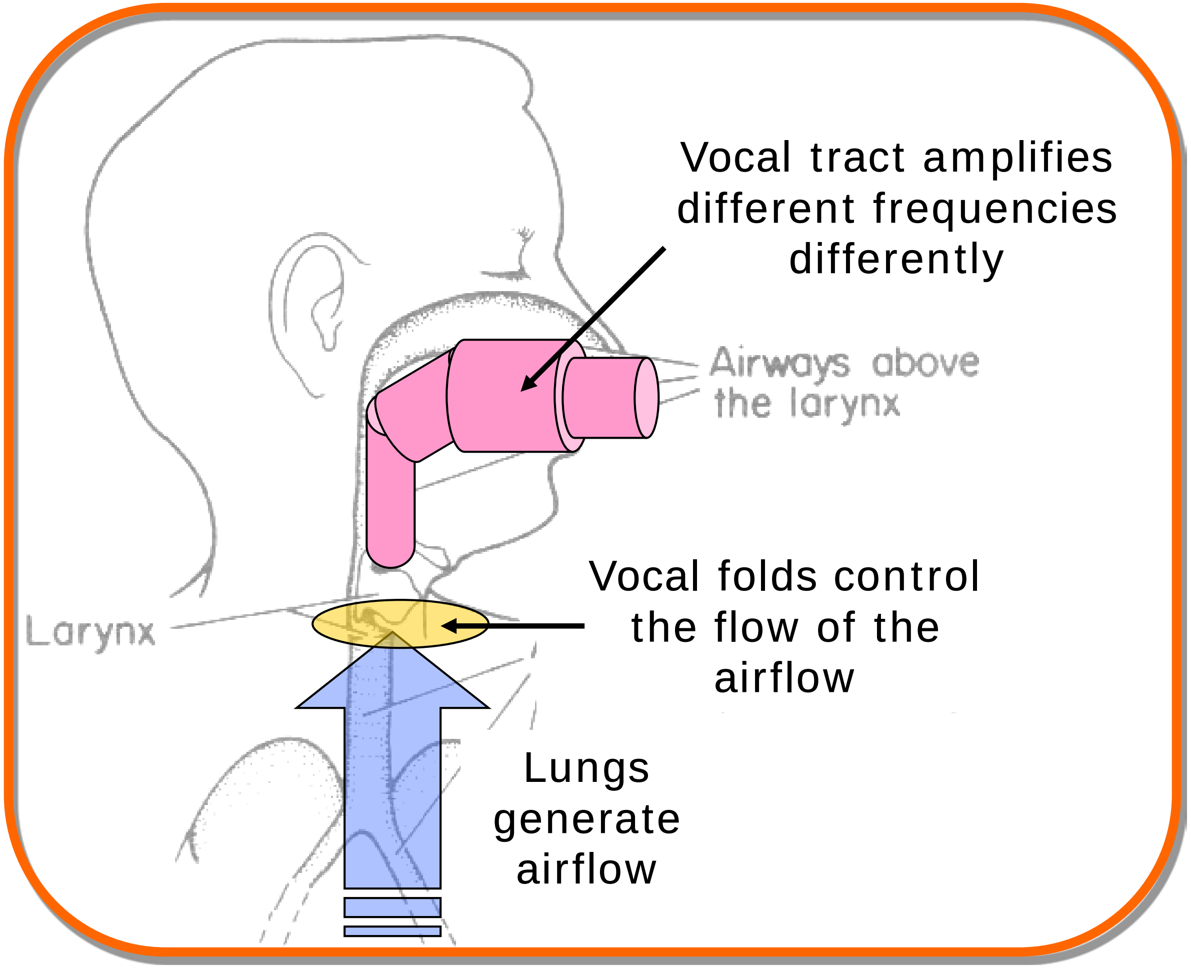
Nasal Cavity

- Over much of its length, the nasal cavity is divided into two passages. (often asymmetrical)
- total length of 11 cm
- total volume of 23 cm^3
- narrowest portion at the nostrils (1-2cm^2)
- the circumference is 3-5 times greater than that of a circular shape.
- large surface area caused acoustic losses in the nasal cavity. (characteristic of nasal sounds)

Nasal Cavity

- Cannot be controlled
- Very speaker dependent
  - good for speaker identification
  - bad for speech recognition

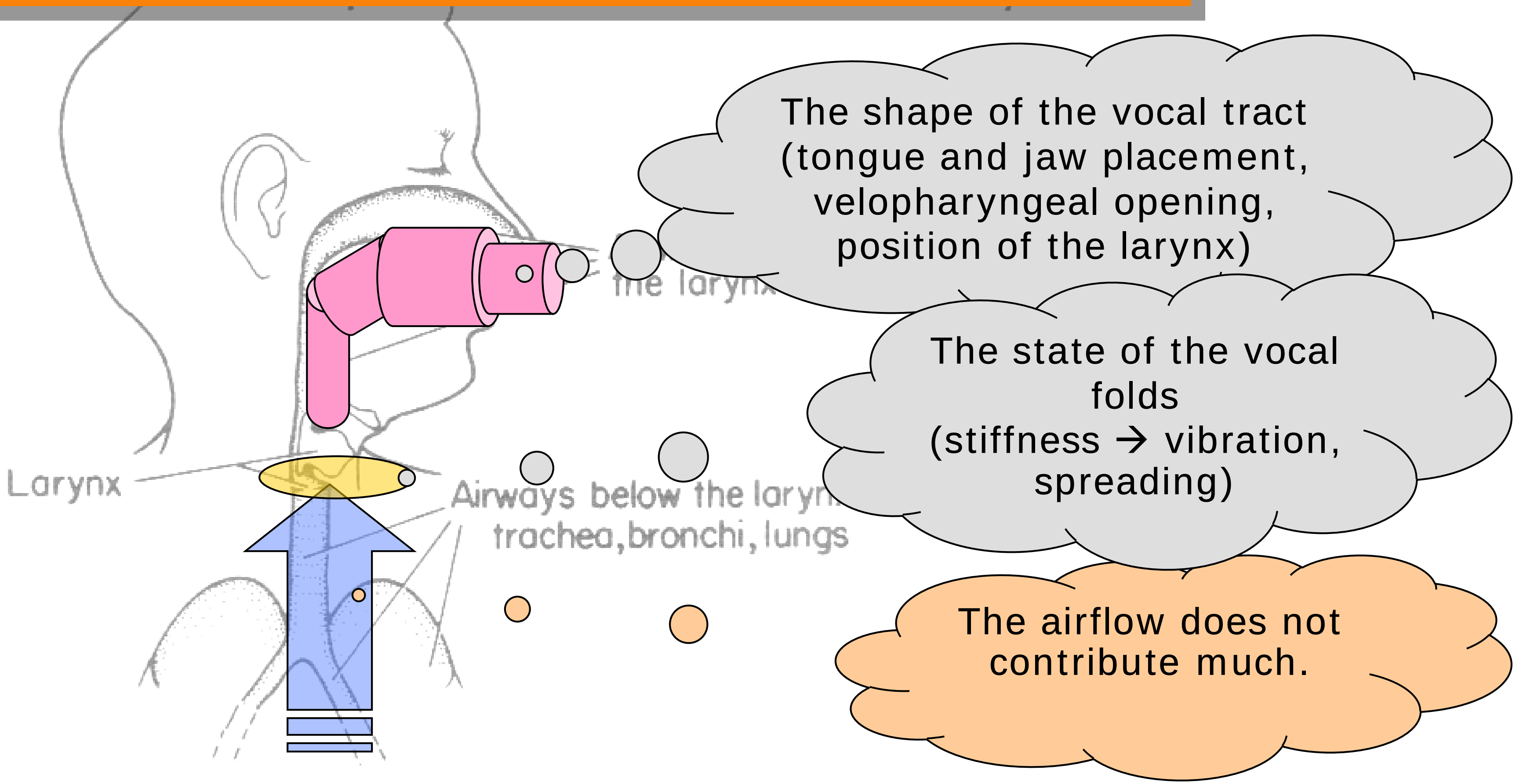
The next step...



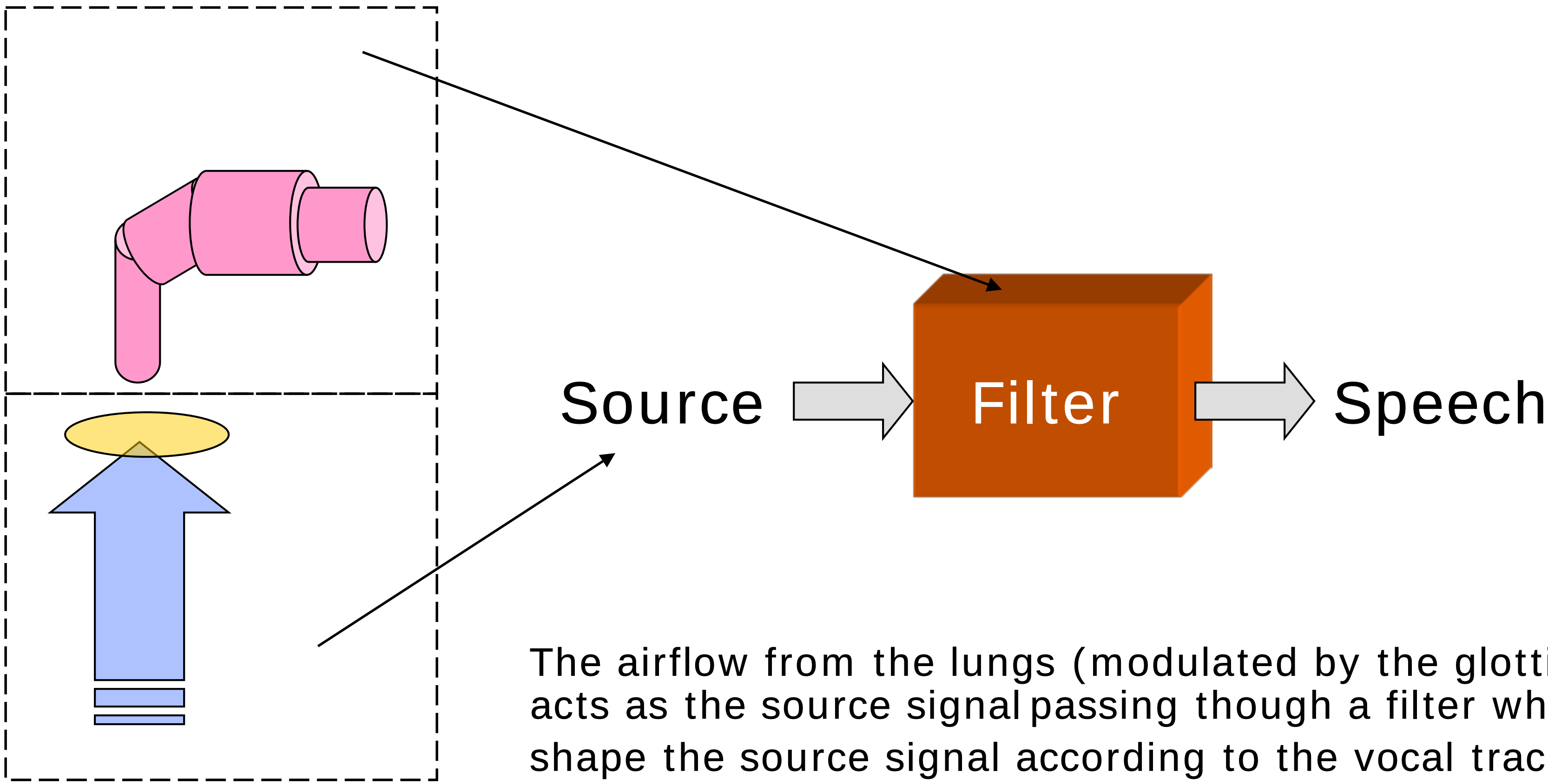
We want to build a model of this system.

Related Factors

How different speech units sound differently to us:

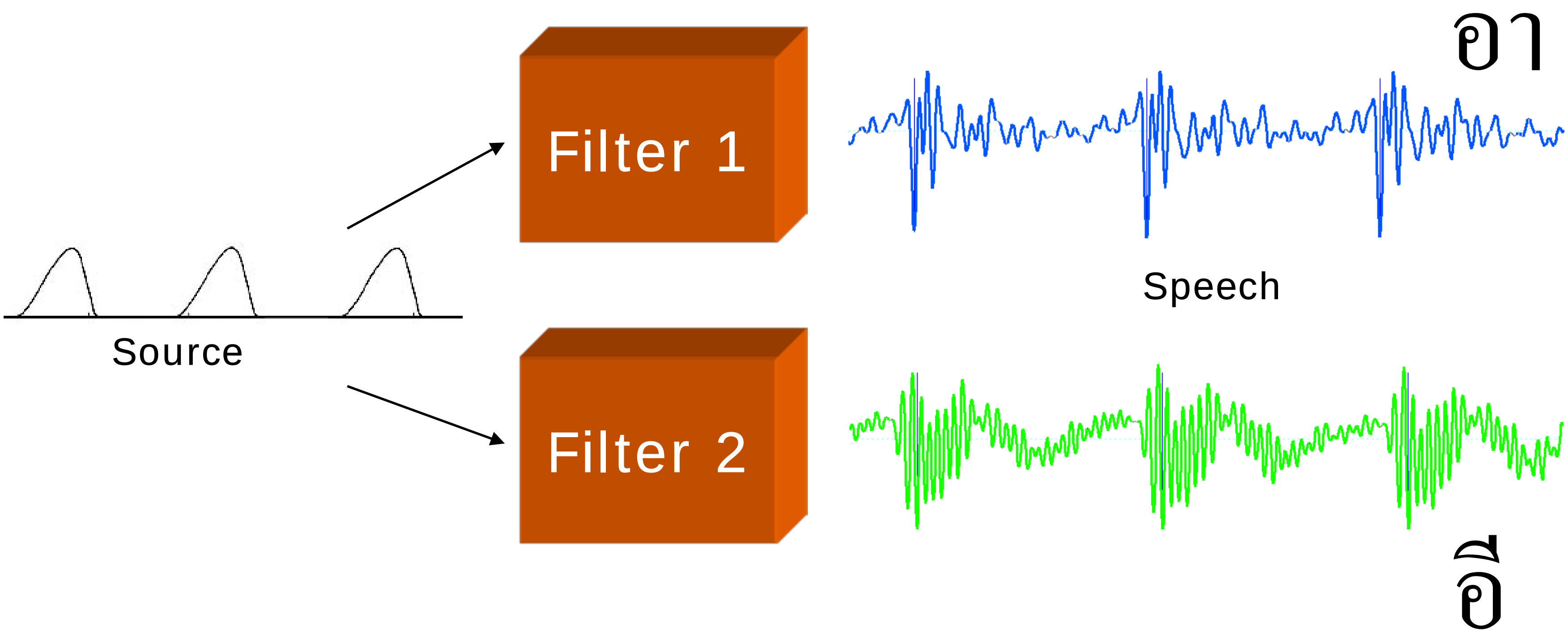


Source-filter model



The airflow from the lungs (modulated by the glottis) acts as the source signal passing though a filter which shape the source signal according to the vocal tract configuration.

Example



Filter 1 is associated with the vocal tract configuration of อา

Filter 2 is associated with the vocal tract configuration of อี

Key Points for Modeling Speech Production

- Sources:
  - Different types of sources used in producing sounds in languages.
- Filters:
  - Amplify different sound frequencies.
  - Link those frequencies with the vocal tract configuration.
  - Other factors.

Source-filter Model of Speech Production

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

$$S(j\omega) \rightarrow \boxed{T(j\omega)} \rightarrow \boxed{R(j\omega)} \rightarrow P_r(j\omega)$$

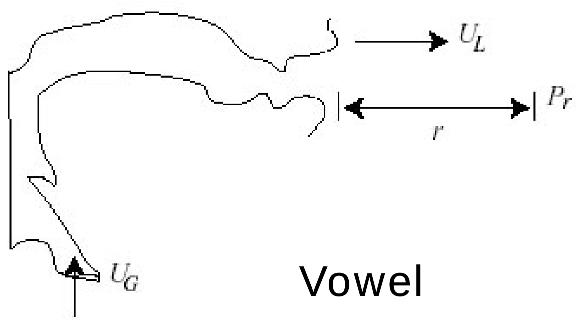
Transfer function of the radiation characteristic

Transfer function of a portion of the vocal tract

Transfer function of the source signal

The source can be:

-Glottal Source
-Noise Source



Vowel

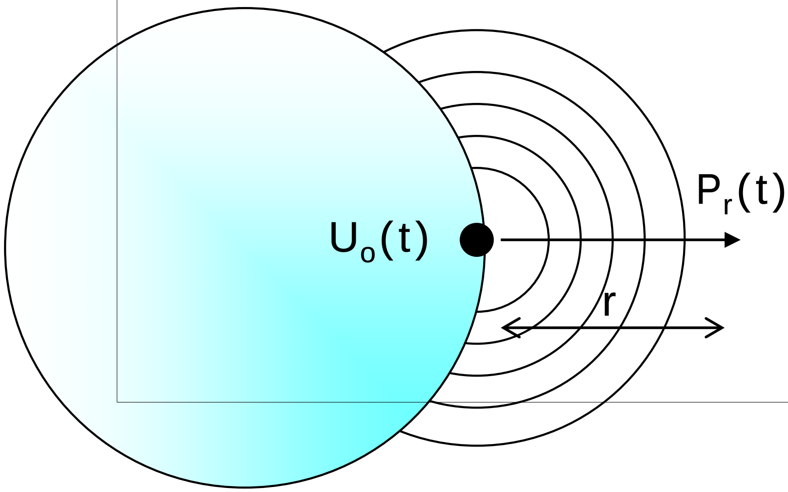
$$S(j\omega) = U_g(j\omega)$$

$$T(j\omega) = \frac{U_L(j\omega)}{U_g(j\omega)}$$

$$R(j\omega) = \frac{P_r(j\omega)}{U_L(j\omega)}$$

Radiation Characteristic

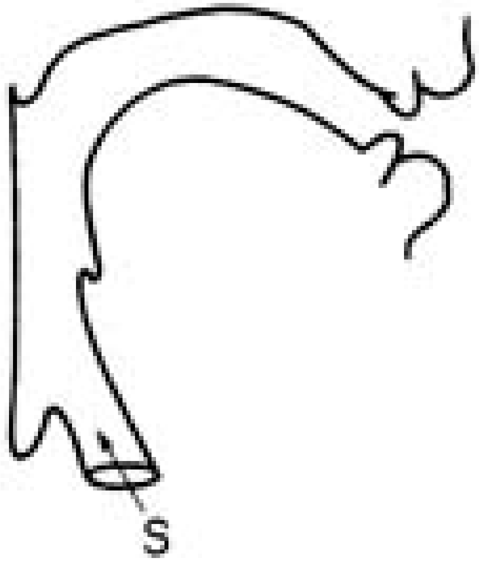
- The sound pressure at distance r from the lips, $P_r(t)$, is linearly related to the volume velocity $U_o(t)$ at the lips.
- For up to 4000Hz, the mouth opening can be regarded as a simple source of strength $U_o(t)$

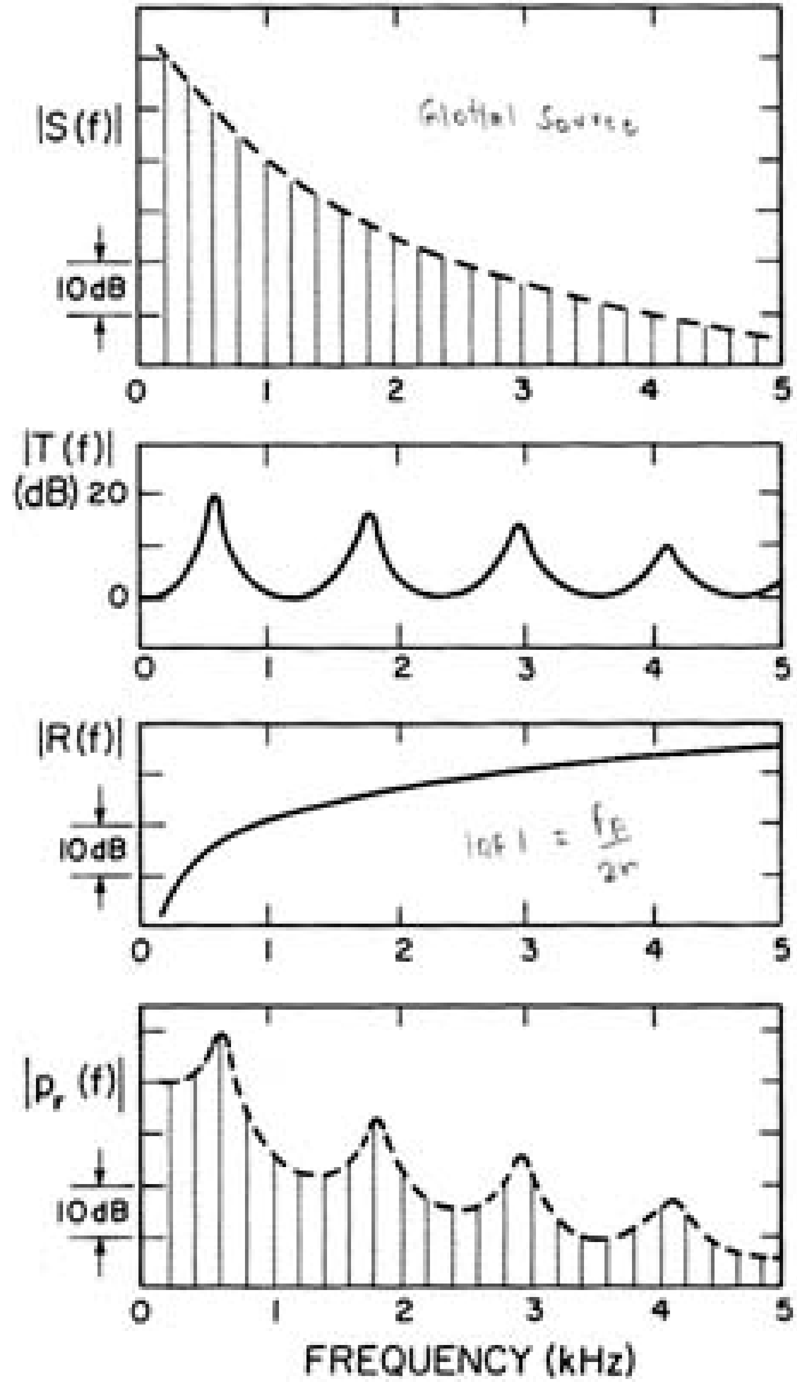


$$R(j\omega) = \frac{P_r(j\omega)}{U_o(j\omega)}$$

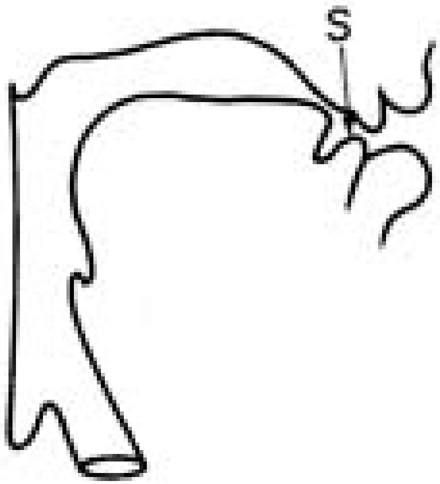
$$= \frac{j\omega}{4\pi r} e^{-j\frac{\omega}{c}}$$

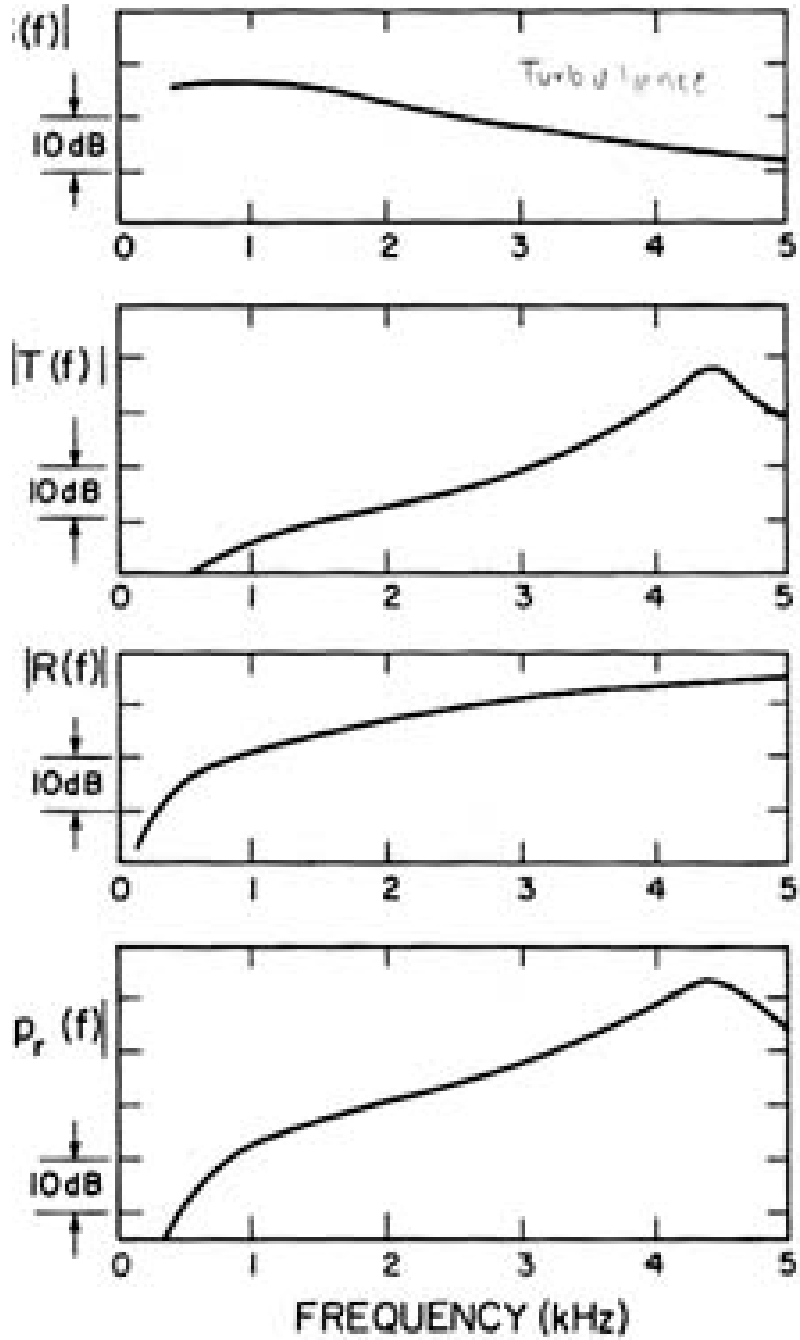
$$\boxed{\left|R(f)\right| = \frac{f\rho}{2r}}$$





picture from Stevens 1999

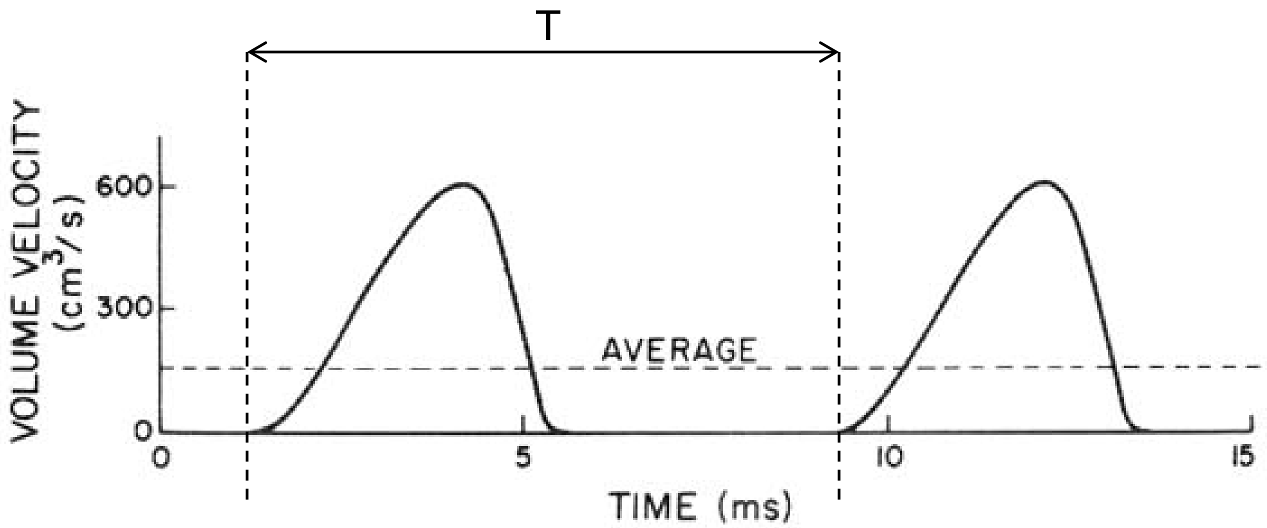




picture from Stevens 1999

Glottal source

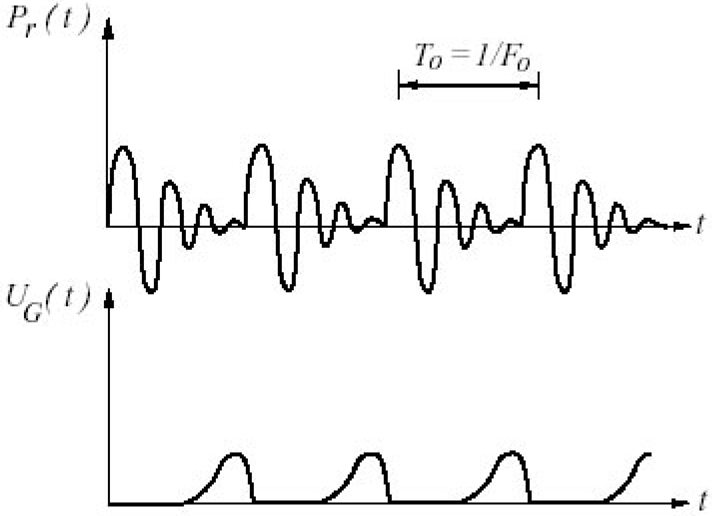
- Volumn Velocity Source at Glottis U_g
- generated from the vibration of the vocal folds
- periodic

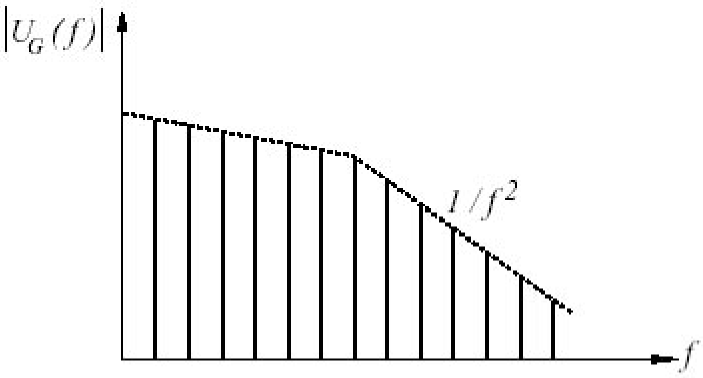


typical shape of the glottal waveform

picture MIT OCW

Glottal source



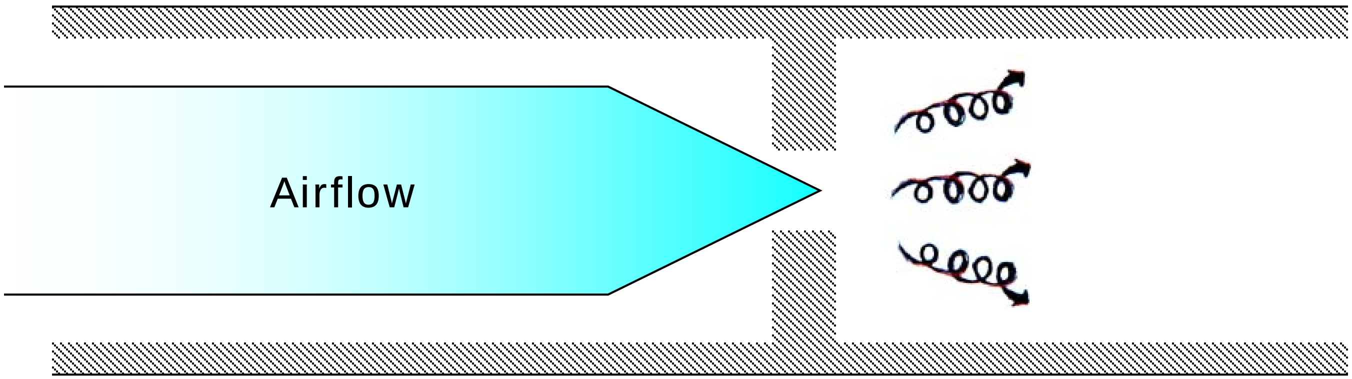


	F_o ave (Hz)	F_o min (Hz)	F_o max (Hz)
Men	125	80	200
Women	225	150	350
Children	300	200	500

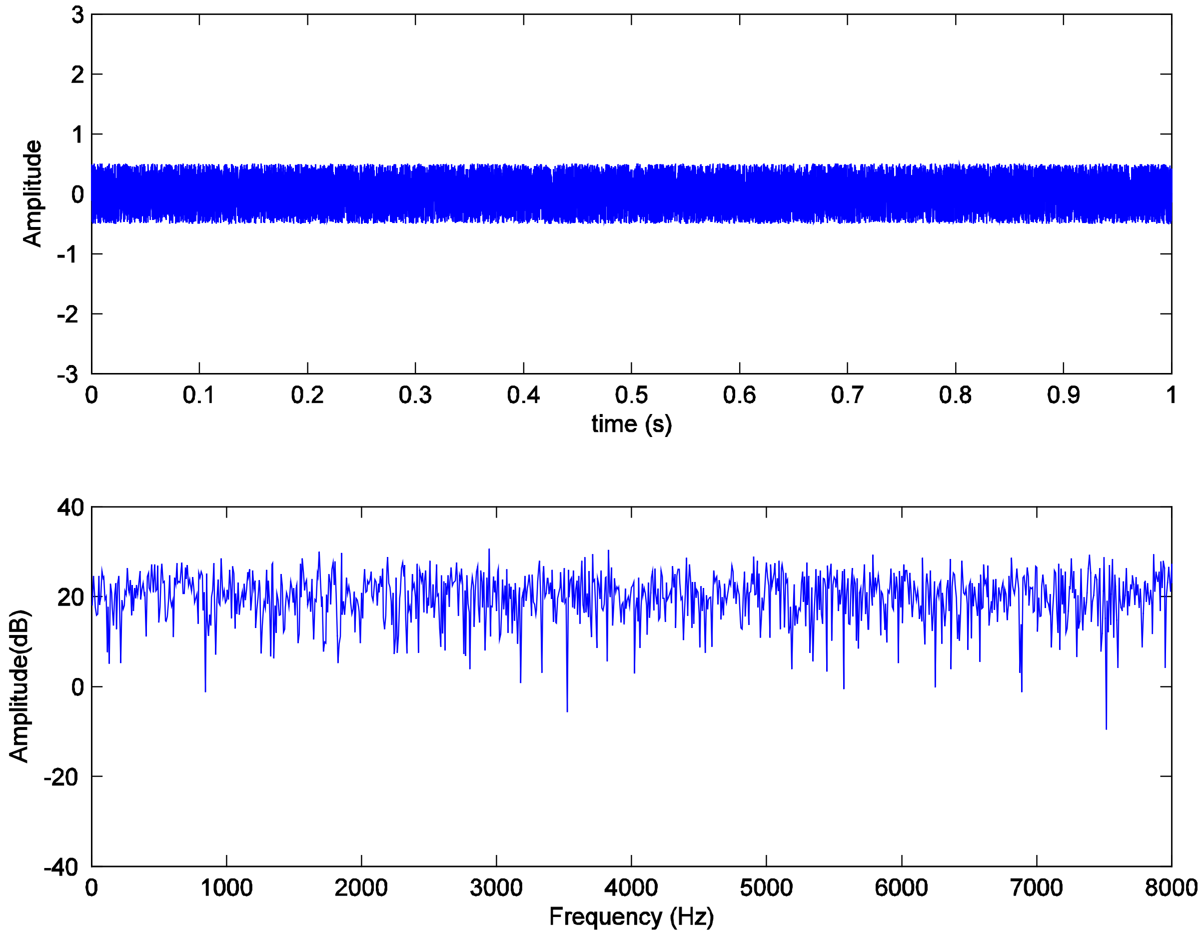
picture from MIT OCW

Noise Source

- Noise source
  - generated from air turbulence at some constrictions in the vocal tract
  - white noise



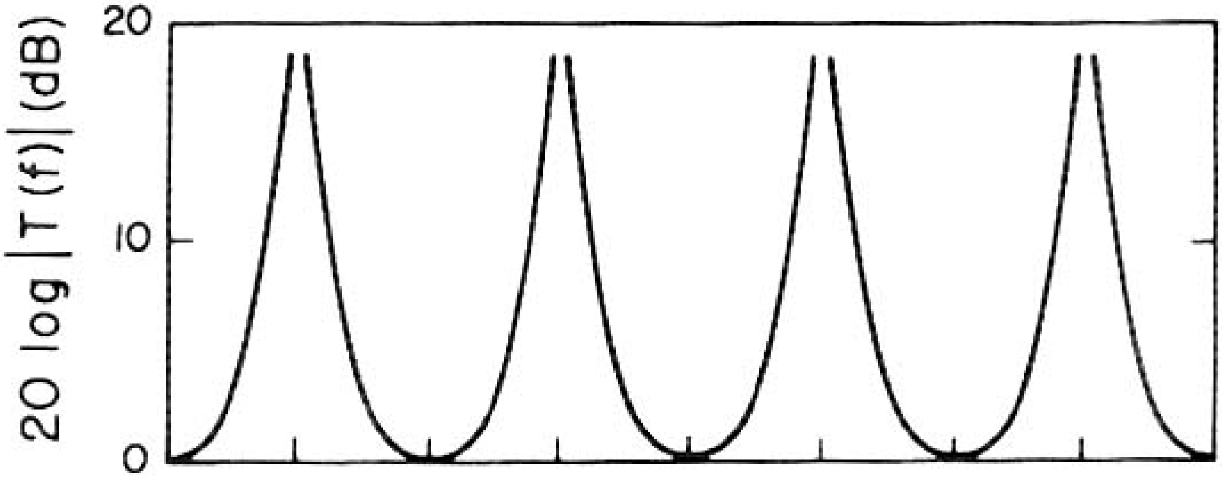
Noise Source



Waveform and spectrum of the white noise

Vocal tract transfer function

- $T(j\omega) = U_l(j\omega)/U_s(j\omega)$
- usually characterized by several peaks corresponding to resonances of the cavity of the vocal tract



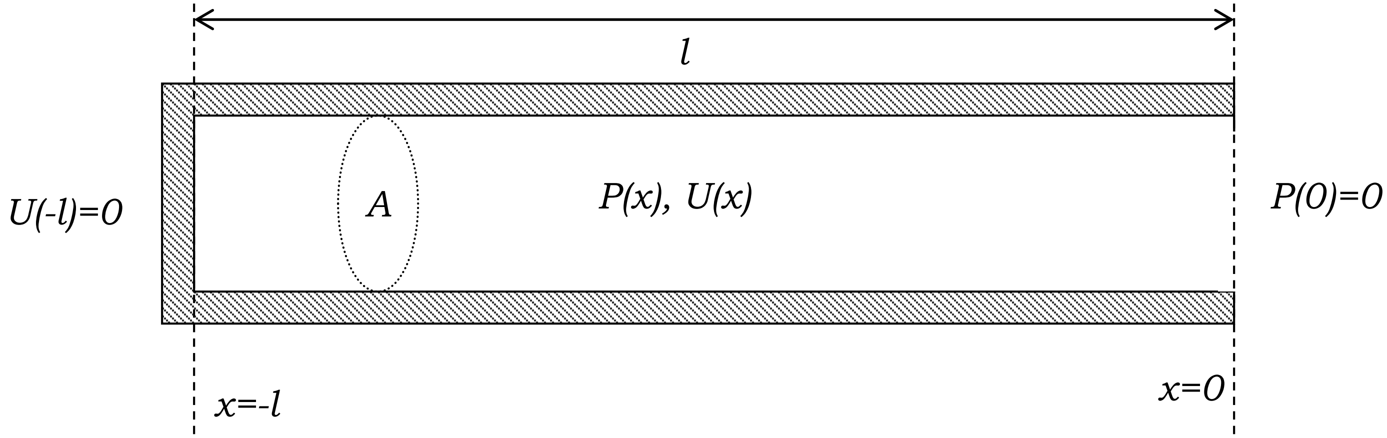
picture from Stevens 1999

A Little Summary

- So, now we know that there are two kinds of source.
- Whichever the source, it will pass through a filter which is a frequency-dependent amplifier.
- The filter will amplify frequency components of the source. This results in the speech waveform that comes out of the mouth (and reach a microphone, if there is one.)

NOW: We would like to know that, given a vocal tract configuration, how the frequency-dependent amplifier is going to be like.

Natural Frequencies of a Uniform Tube



Acoustic equations

$$\frac{\partial P(x,t)}{\partial x} = -\frac{\rho}{A}\frac{\partial U(x,t)}{\partial t} \quad (1)$$

$$\frac{\partial U(x,t)}{\partial x} = -\frac{A}{\gamma P_0}\frac{\partial P(x,t)}{\partial t} \quad (2)$$

ρ : density of air (0.00114 gm/cm^3)

γ : ratio of specific heat at constant pressure to specific heat at constant volume (1.4 for air)

P_0 : ambience pressure level (cm H_2O)

c : velocity of sound (35,400 cm/s)

Natural Frequencies of a Uniform Tube

Assume:

$$\left.\begin{aligned} P(x,t) &= P(x)e^{j\omega t} \\ U(x,t) &= U(x)e^{j\omega t} \end{aligned}\right\}$$

Air particles at position x oscillates with angle frequency ω

From 1 and 2 $\Rightarrow$ $\boxed{\frac{\partial^2 P(x)}{\partial x} + k^2 P(x) = 0}$ 3 $\quad k = \frac{\omega}{c}, c = \sqrt{\frac{\gamma P_0}{\rho}}$

- The solution of 3 is either in the form of *sin* or *cos*
- The requirements for natural frequencies are $p(0)=0$ and $U(-l)=0$

$\Rightarrow$ $P(0)=0 \rightarrow \boxed{P(x) = P_m \sin(kx)}$

Natural Frequencies of a Uniform Tube

- From 1 and $P(x,t) = P(x)e^{j\omega t}$ → $U(x) \propto \dfrac{\partial P(x)}{\partial x}$

$$\frac{\partial P(x)}{\partial x} = kP_m \cos(kx)$$

$$\therefore U(-l) = C\cos(-kl) = C\cos(kl) = 0$$

$$\cos(kl) = 0 \Leftrightarrow kl = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$$

So, formant frequencies are

$$\boxed{f = \frac{c}{4l}, \frac{3c}{4l}, \frac{5c}{4l}, \ldots}$$

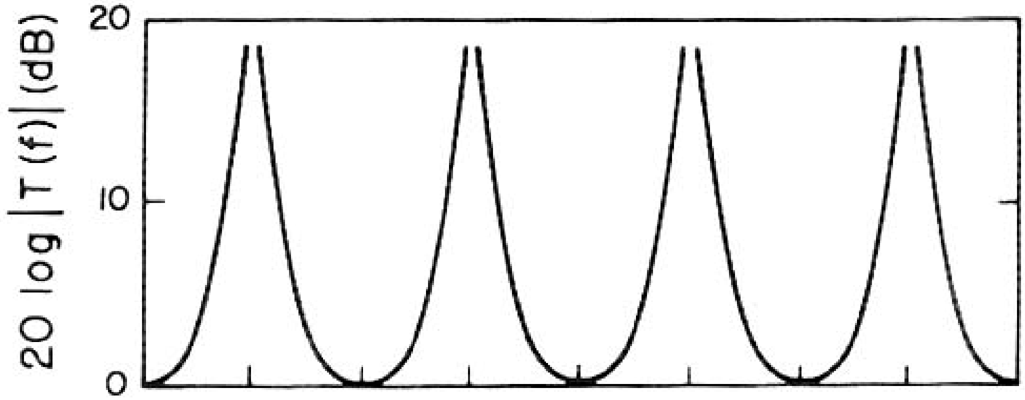
Transfer Function

$$U(-l,t) = C\cos(kl)e^{j\omega t}$$

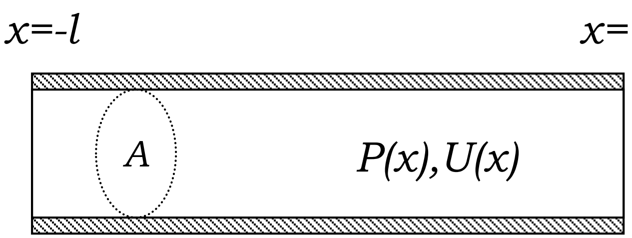
$$U_{-l}(j\omega) = C\cos(kl)F(e^{j\omega t})$$

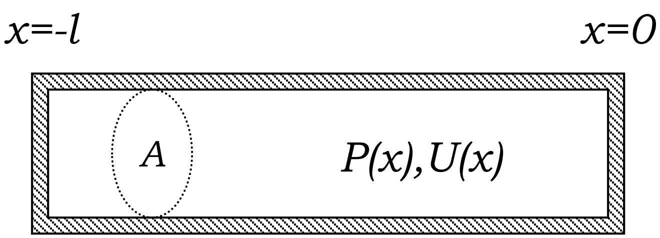
$$U_0(j\omega) = C\cos(0)F(e^{j\omega t})$$

$$T(j\omega) = \frac{U_0(j\omega)}{U_{-l}(j\omega)} = \frac{1}{\cos(kl)}$$



Uniform Tube open/closed at both ends





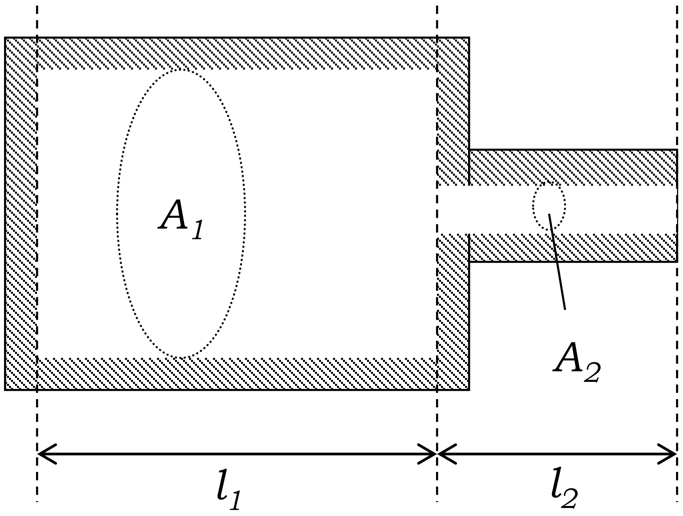
$P(0) = 0,\ U(0) = U_m$
$P(-l) = 0,\ U(-l) = U_m$

$P(0) = P_m,\ U(0) = 0$
$P(-l) = P_m,\ U(-l) = 0$

Formant frequencies are

$$\boxed{f = 0, \frac{c}{2l}, \frac{c}{l}, \frac{3c}{2l}, \ldots}$$

Helmholtz Resonator



$$f_1 = \frac{1}{2\pi\sqrt{M_A C_A}}$$

$$M_A = \frac{\rho l_2}{A_2}, C_A = \frac{A_1 l_1}{\rho c^2}$$

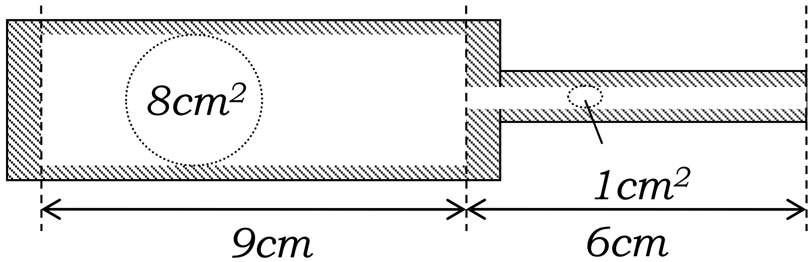
M_A : Acoustic Mass
C_A : Acoustic Compliance

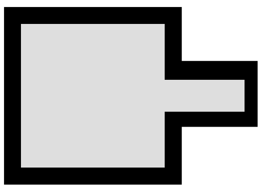
$$\boxed{f_1 = \frac{c}{2\pi}\sqrt{\frac{A_2}{A_1 l_1 l_2}}}$$

Vocal Tract Modeling by Concatenated Tubes

- A number of vocal tract shapes, both vowels and consonants, can be approximated by two or more resonators or uniform tubes of different cross-sectional areas connected together.
- neglect coupling between components
- combination of natural frequencies for all of the components (+Helmholtz structure)

Concatenated Tubes





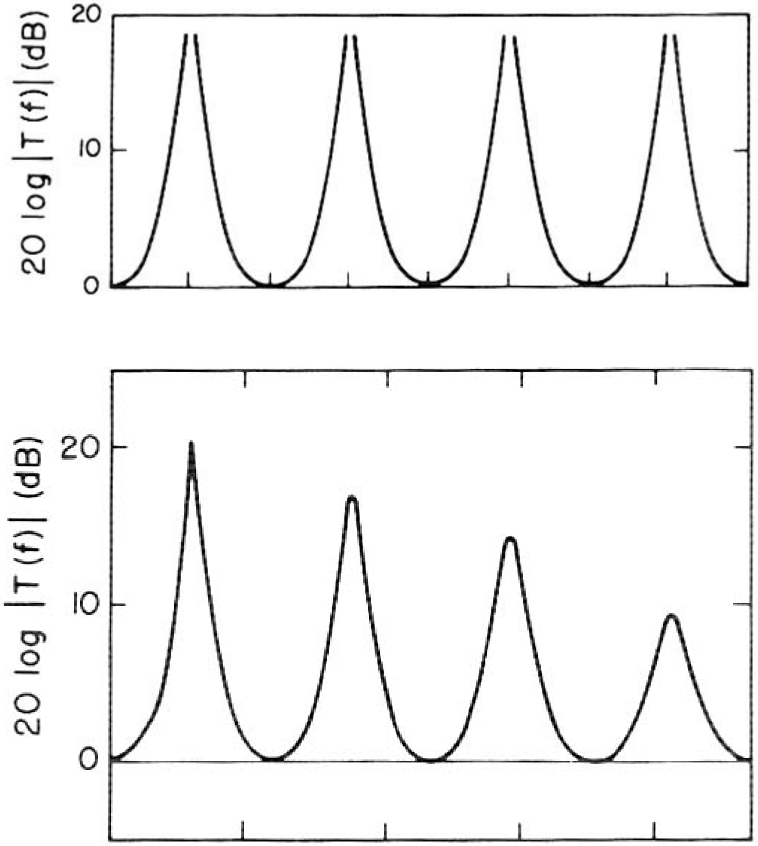
$$f = \frac{354}{2\pi}\sqrt{\frac{1\times10^{-4}}{(8\times10^{-4})(9\times10^{-2})(6\times10^{-2})}}$$
$$f = 271Hz$$

$$f_1 = \frac{c}{2l} = \frac{354}{2\times9\times10^{-2}} = 1967Hz$$
$$f_2 = \frac{c}{l} = 1967\times2 = 3933Hz$$
$$f_3 = \ldots$$

$$f_1 = \frac{c}{2l} = \frac{354}{2\times6\times10^{-2}} = 2950Hz$$
$$f_2 = \frac{c}{l} = 2590\times2 = 5180Hz$$
$$f_3 = \ldots$$

formant frequencies = 271Hz, 1967Hz, 2950Hz

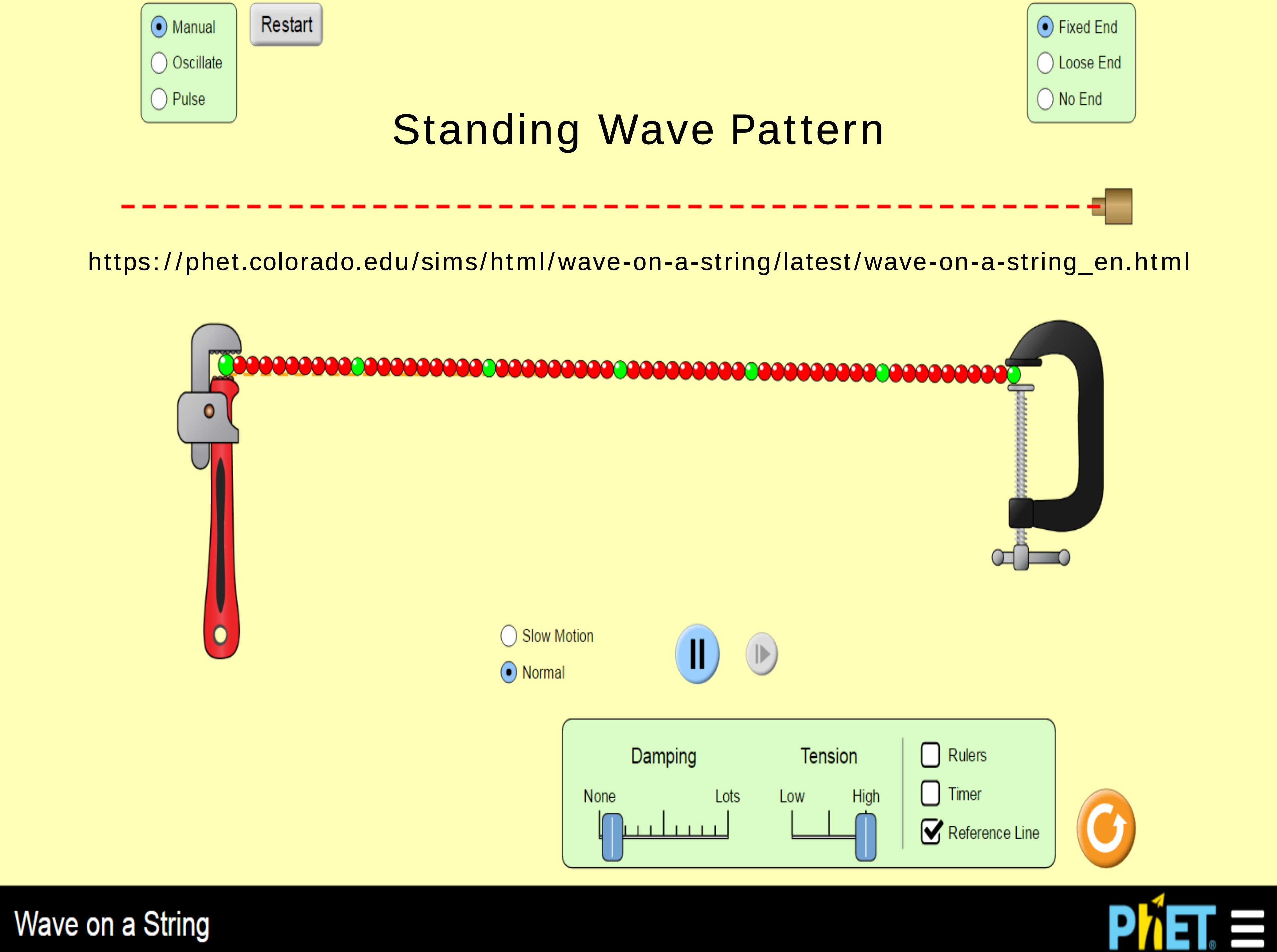
Loss in the Vocal Tract



picture from Stevens 1999

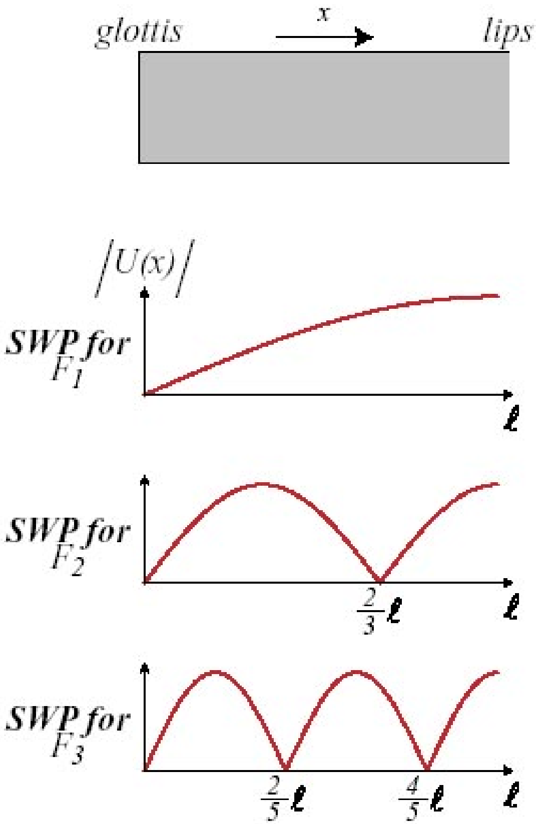
Standing Wave Pattern

https://phet.colorado.edu/sims/html/wave-on-a-string/latest/wave-on-a-string_en.html



Standing Wave Pattern

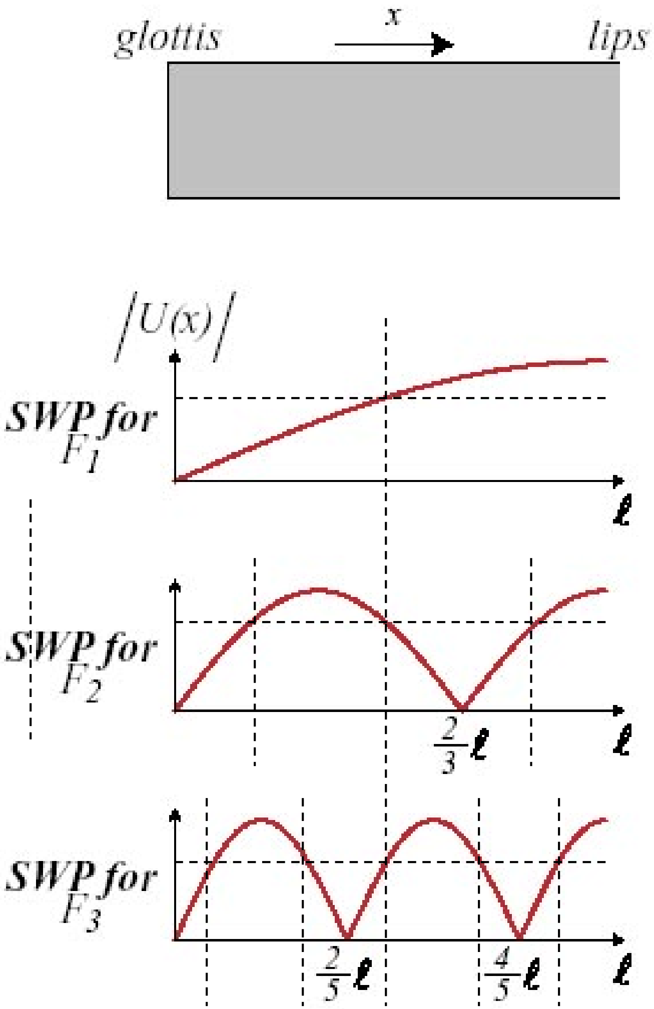
A uniform tube open at one end and closed at the other is often referred to as a “quarter wavelength resonator”



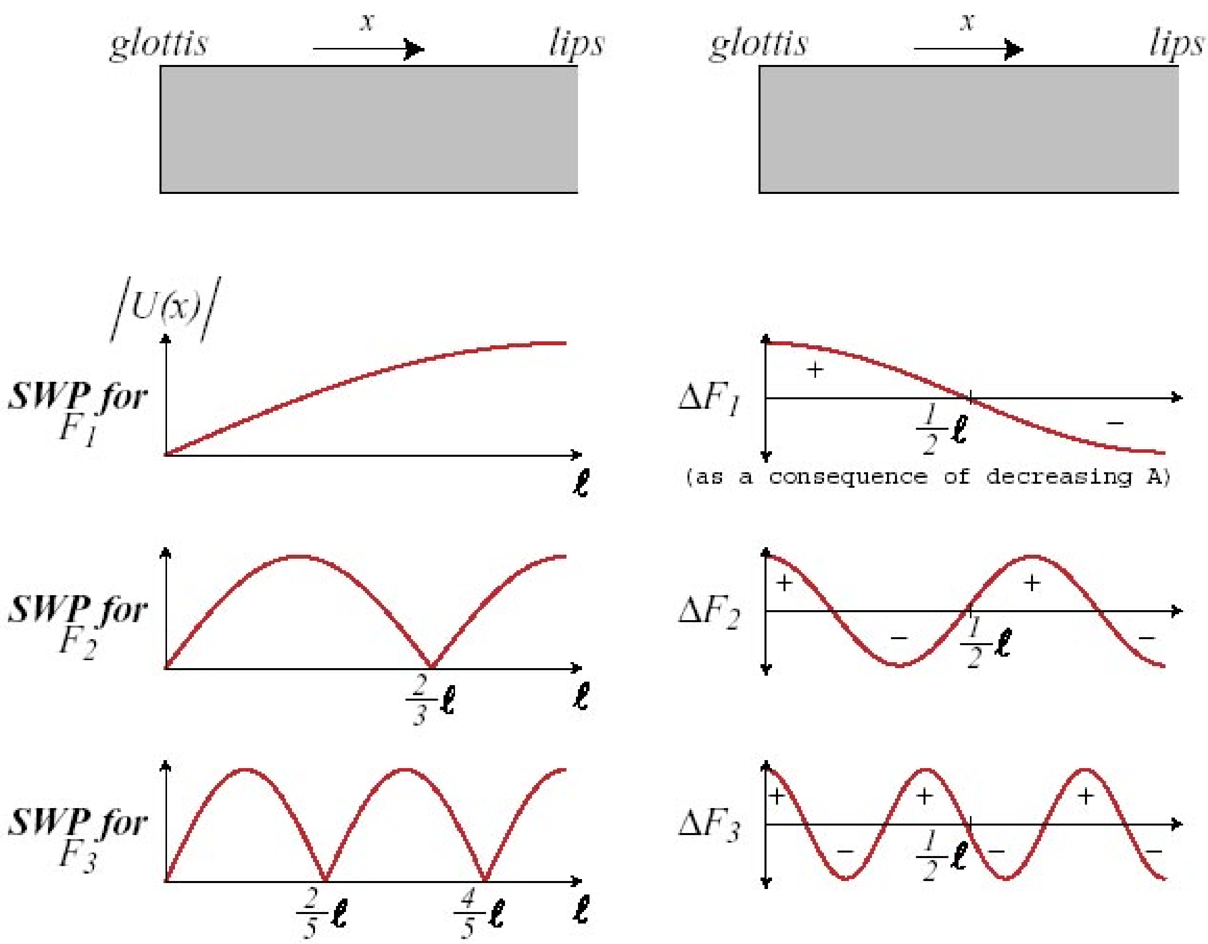
picture from MIT OCW

Perturbation Theory

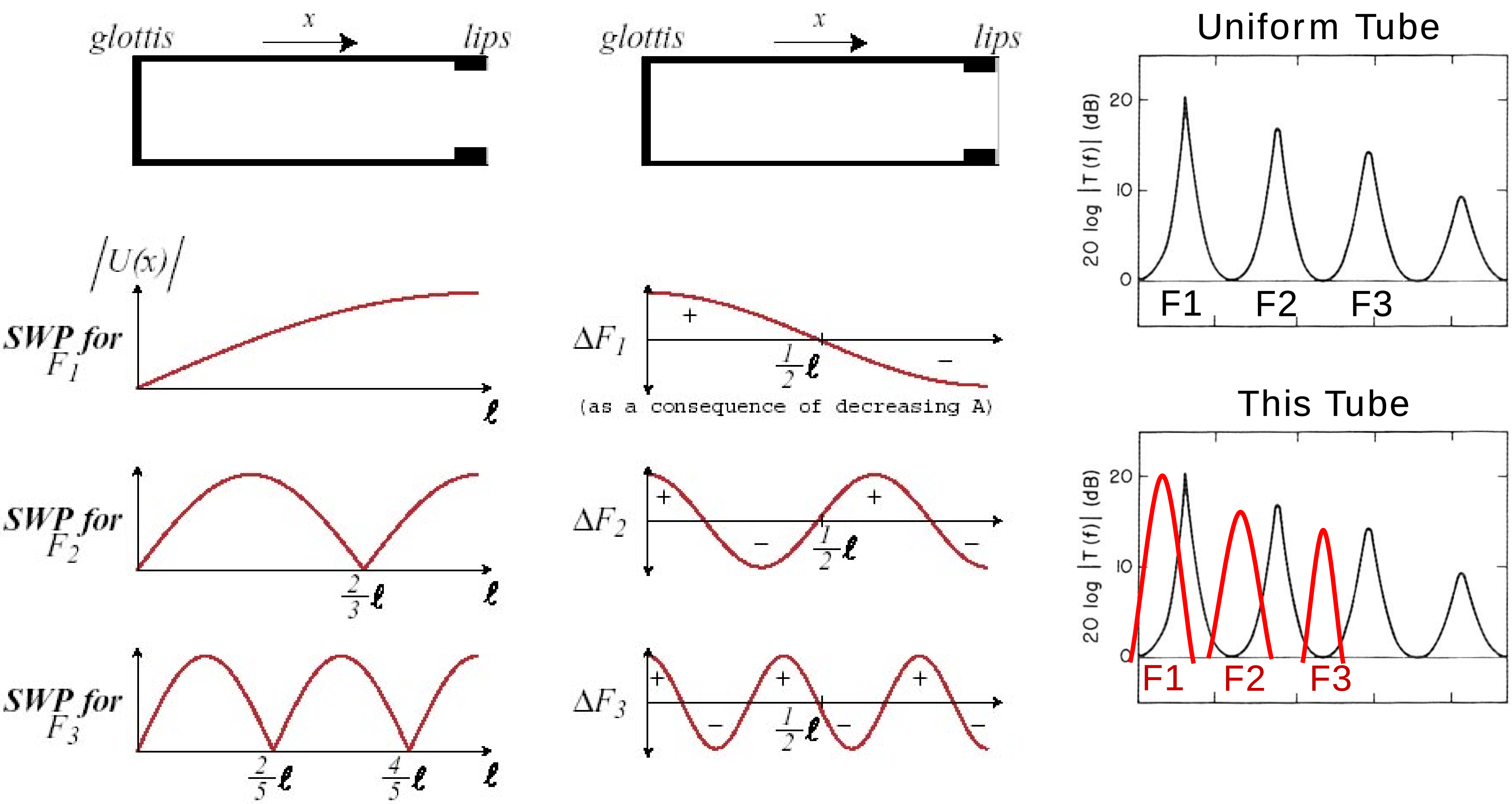
“ Narrowing
the tube where
U(x) is HIGH
in the Standing Wave Pattern
associated with a Formant Frequency
Decreases
the Frequency of the Formant ”



Perturbation Theory



Perturbation Theory



References

- Stevens, Kenneth. *Acoustic Phonetics*. Cambridge, MA: MIT Press, 1999. ISBN: 0-262-19404-X.

Classes of Sound

- classify by degree of constriction
  - Consonant
    - obstruction in the path of air flow
      - complete closure
      - narrow constriction
    - abrupt change in vocal tract configuration
    - abrupt change in the signal
  - Semivowel
    - between vowel and consonant
    - non-abrupt change / slightly constricted
  - Vowel
    - relatively no obstruction of the air flow
    - smooth gradual change in the signal

IPA

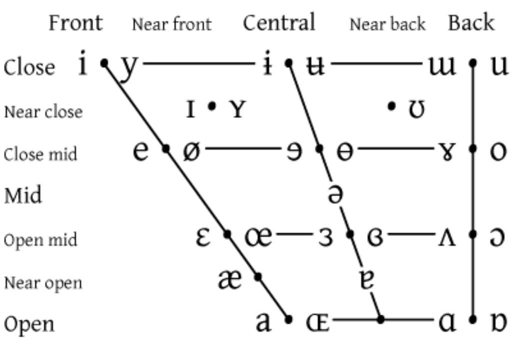
THE INTERNATIONAL PHONETIC ALPHABET (2005)

CONSONANTS (PULMONIC)

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Epi-glottal	Glottal
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ			
Plosive	p b	ȹ ȸ		t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʡ	ʔ
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	ʜ ʢ	h ɦ
Approximant		ʋ		ɹ		ɻ	j	ɰ				
Trill	ʙ			r					ʀ		ᴙ	
Tap, Flap		ⱱ		ɾ		ɽ						
Lateral fricative				ɬ ɮ		ꞎ	𝼆	𝼄				
Lateral approximant				l		ɭ	ʎ	ʟ				
Lateral flap				ɺ		𝼈						

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured ɦ.
Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

VOWELS



Vowels at right & left of bullets are rounded & unrounded.

THE INTERNATIONAL PHO

How to articulate

Main Articulator

ON NTS (PULMONIC)

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Retrof
Nasal	m	ɱ		n		ɳ
Plosive	p b	ȹ ȸ		t d		ʈ ɖ
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ
Approximant		ʋ		ɹ		ɻ
Trill	ʙ			r		
Tap, Flap		ⱱ		ɾ		ɽ
Lateral fricative				ɬ ɮ		ꞎ
Lateral approximant				l		ɭ

A Note on Phonetic Symbols

- For our convenience in this class (avoiding problems with fonts), we will not use the standard International Phonetic Alphabets (IPA) as the default representation of sounds.
- We will use a set of notation presented in the next slides to represent Thai sounds.
- Note that notations used for some English sounds discussed in this lecture are not standard. (not IPA)

Thai Syllable Structure

$$C_i + V + C_f + \#$$

Vowel

Tone

Initial Consonant Sound:
Consonant
Consonant Cluster

Final Consonant Sound:
Consonant

Sounds in Thai

p	ปาก
t	เต้น, กฎิ
c	จะ
k	ก่อน
z	อาน
ph	พบ, ภัย, ผ่าน
th	ทิ้ง, ธง, เฒ่า, ฐาน, มณโฑ
ch	ชอบ, เฌอ
kh	คน, เขิน, ฆ่า
b	บอก
d	ด้าน, ชฎา

m	ไม่
n	นาน, เณร
ng	เงิน
l	เล่น, กีฬา
r	รอ, ฤทัย
f	ฝน, ฟัน
s	สาย, ศิลา,รักษา, ซ่อน
h	โหน, เฮฮา
w	ว่า
j	ย้อน, หญิง

pr	ประสาน
phr	พราน
tr	เตรียม
kr	กราบ
khr	คร่ำ
pl	ปลา
phl	พลาด
thr	จันทรา
kl	เกลอ
khl	เคลื่อน
kw	กวาง
khw	ขวา

br	เบรน
bl	บลู
fr	ฟราย
fl	เฟลม
dr	ดรากอน

Sounds in Thai

a	อะ
aa	อา
i	อิ
ii	อี
v	อึ
vv	อื
U	อุ
uu	อู
e	เอะ
ee	เอ
x	แอะ
xx	แอ

o	โอะ
oo	โอ
@	เอาะ
@@	ออ
q	เออะ
qq	เออ

ia	เอียะ
iia	เอีย
va	เอือะ
vva	เอือ
ua	อัวะ
uua	อัว

V

p^	พบ
t^	เกร็ด
k^	ปาก
n^	หาร
m^	ลม
ng^	ฟาง
j^	ยาย
w^	กาว

f^	กราฟ
l^	แอล
s^	เอส
ch^	คลัช

C_f

Thai Tones

Defined based on Fundamental Frequency (Pitch)

0	คา	Mid tone
1	ข่า	Low tone
2	ข้า, ค่า	Falling tone
3	ค้า	High tone
4	ขา	Rising tone

Sounds in Thai

กนกวรรณกร

วชิรญาณวโรรส

ฮกเฮงล้ง

แป๊ะเจี๊ยะ

แอสเซมบลี

โปรเทคเตอร์