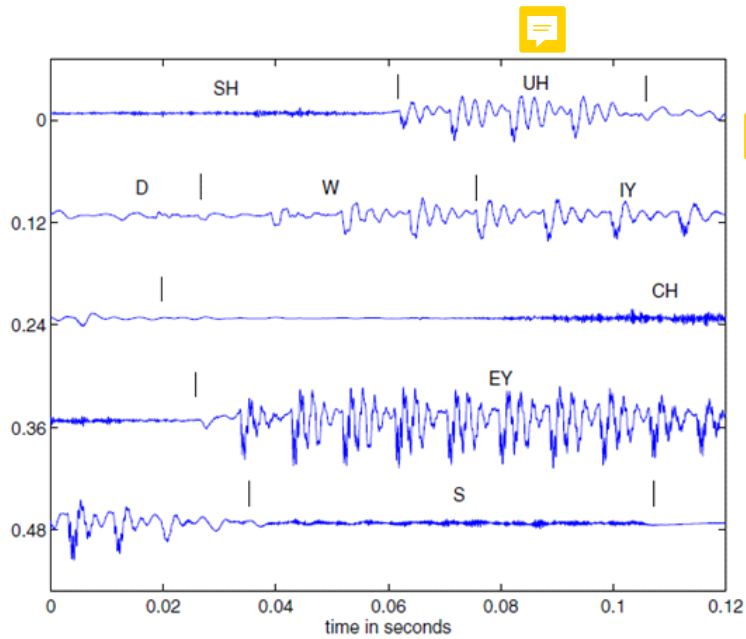




## Speech Production

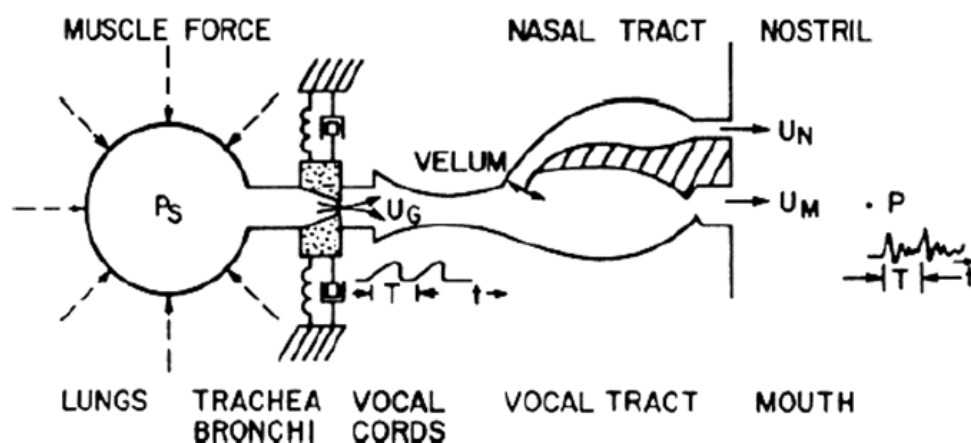
Utterance: "Should we chase"

Acoustic waveform



Production of speech:

- Respiration  $\Leftarrow$  **Lungs**
- Phonation  $\Leftarrow$  **Vocal cords**
- Articulation  $\Leftarrow$  **Vocal tract**



**Schematic model of the vocal tract system**

Ref: Fig 2.1 in *Foundations and Trends in Signal Processing*, Vol. 1, Nos. 1–2 (2007) 1–194 © 2007 L. R. Rabiner and R. W. Schafer

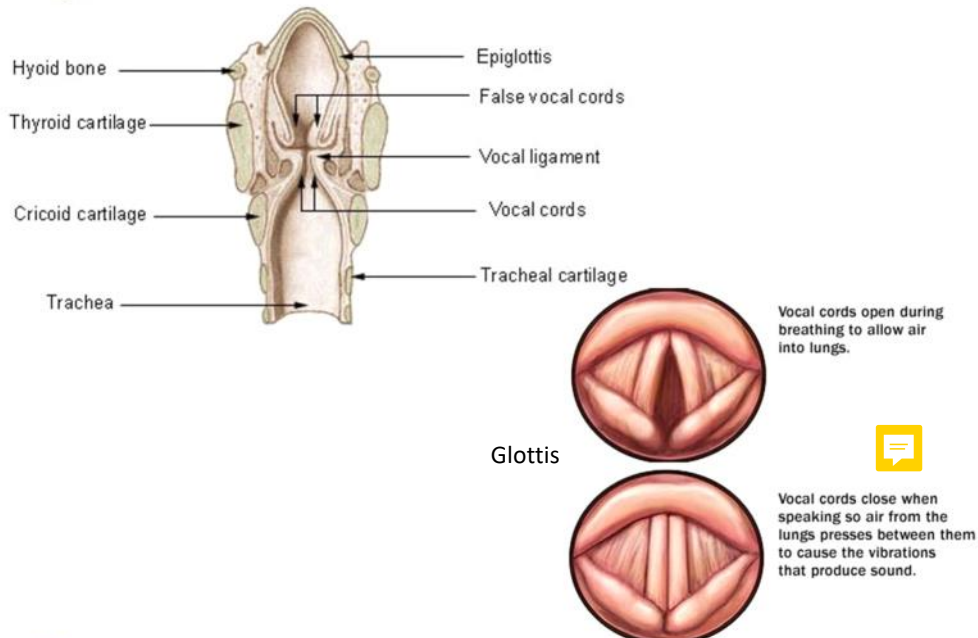
- **Respiration**: the air flow for speech production (lungs).
- **Phonation**: generation of basic sound by vibration of vocal cords (glottis). The otherwise smooth airflow is disturbed, causing sound.
- **Articulation**: changing the spectrum of sound (vocal tract). It gives rise to different types of sound. The variation is generated by adjusting nature & shape of mouth cavity.

## Respiration

- Simple but important part of speech production. Respiration provides the **air-flow** and pressure source required for speech production. The lungs primarily serve breathing: inspiration, expiration.
- Most languages sounds are formed during **expiration** (“egressive” sounds).
- Total lung capacity is 4-5 litre. The volume velocity of air leaving the lungs is about 0.2 lt/sec during sustained sounds.
- Increased air-flow rate  $\Rightarrow$  increase in **sound amplitude**

## Phonation

### Larynx



Department of Electrical Engineering, IIT Bombay

6

Anatomical views of Larynx and vocal folds <[www.mayoclinic.com](http://www.mayoclinic.com)>

### Vocal folds: anatomy and physiology

Pair of **elastic structures** of tendon, muscles and mucous membrane situated in the larynx. The variable opening between the folds is the “**glottis**”.  
In normal breathing, cords are parted to allow free passage of air.

The vocal cords functions chiefly in two modes:

1. **With phonation**: opening-closing periodic motion => periodic waveform
2. **Without phonation**: vocal folds are kept slightly parted => aperiodic (noisy) waveform

Observing vocal fold motion:

- video photography (see track9)
- electro-glottography

Phonation (vocal cords vibration) is an involuntary muscle action. It occurs when

(a) the vocal cords are elastic and close together, and

(b) there is sufficient difference between sub-glottal and supra-glottal pressure

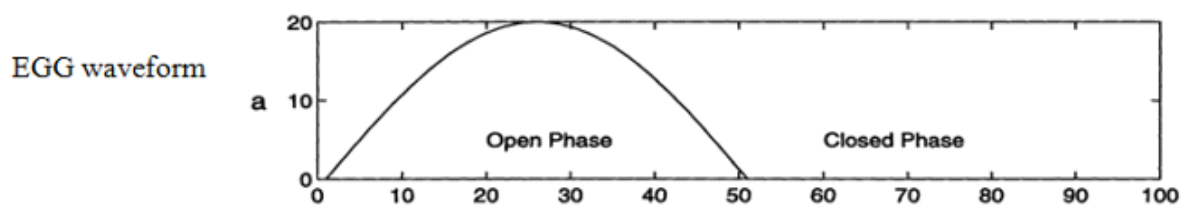
The aerodynamics.....

### Electroglottograph (EGG)

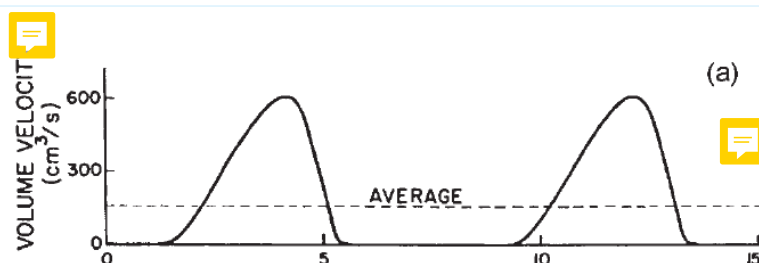
Impedance is monitored via high-frequency current between electrodes across throat.

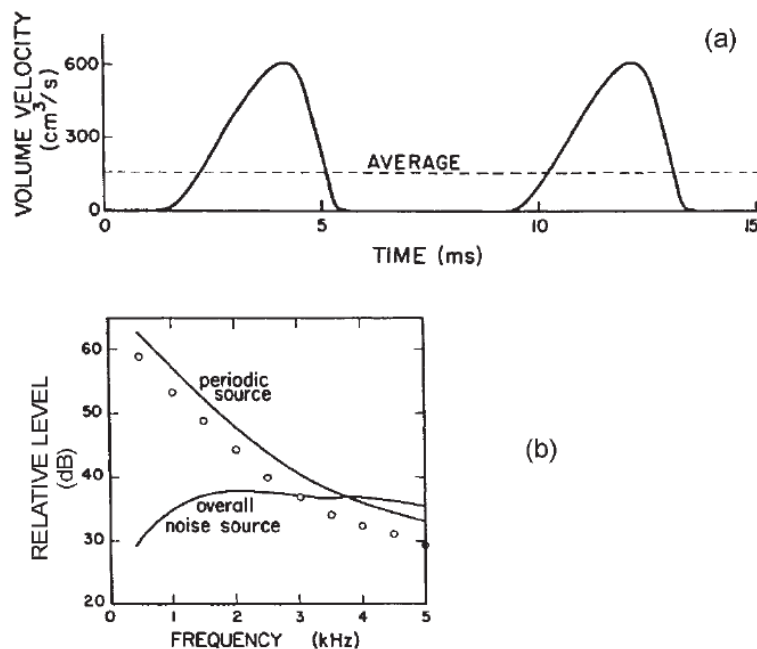
EGG is based on the principle that tissue is a moderate conductor whereas air is poor. A high frequency current is passed between electrodes positioned on either side of thyroid cartilage and electrical impedance is monitored => area of opening vs time.

Show EGG waveform (correlate of glottal opening).



But more typically, we show glottal vol. Velocity (cc/sec vs time). Not directly obtained from the glottal opening due to source-tract interaction (loading) effects. Rothenberg flow mask is used to measure flow at mouth opening and then formants are removed by inverse filtering.





**Fig. 1** Illustrating various acoustic aspects of the glottal sound source for speech production. (a) Schematic representation of the glottal phonation source, showing the volume velocity versus time for a typical male voice; (b) Comparison of the envelope of the spectrum of the phonation source and the spectrum that would be obtained for the waveform of the aspiration noise source at the glottis. The open circles are the relative levels of every fourth harmonic, assigning a fundamental frequency of 125 Hz (From [3]).

Glottal flow signal can be approximated by 2-poles near dc.

K. N. Stevens, "On the quantal nature of speech," J. Phonet., 17, 3-46 (1989).

## Rate of Vibration of the vocal cords

The average rate is inversely proportional to the length of the vocal folds.  
This length is correlated with neck circumference

**Voluntary control:** By means of muscle contractions, the vocal folds can be varied in **length (tension)**, **thickness** and **position configuration**.

Folds are relaxed (short) and thick -> **low pitch**  
Folds are tense (long) and thin -> **high pitch**

Male: 80 - 160 Hz  
Female: 160 - 320 Hz

Glottal pulses are not truly periodic but exhibit **jitter** and shimmer due to neurologic, biomechanical and aerodynamic disturbances.

Jitter: period to period variations in duration; normally < 1%

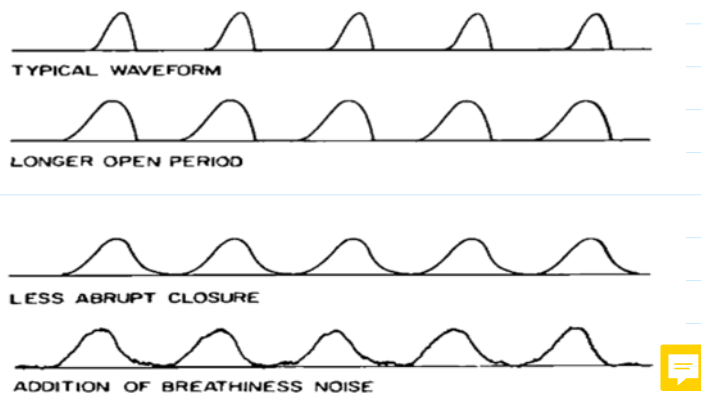
Shimmer: period to period variations in amplitude; normally < 6%


Not normally directly perceptible but add to naturalness of the voice.

High jitter-shimmer => **roughness**

**Voice quality** is altered by modifying glottal vibration pattern.  
Voice quality changes can be non-phonemic or phonemic.



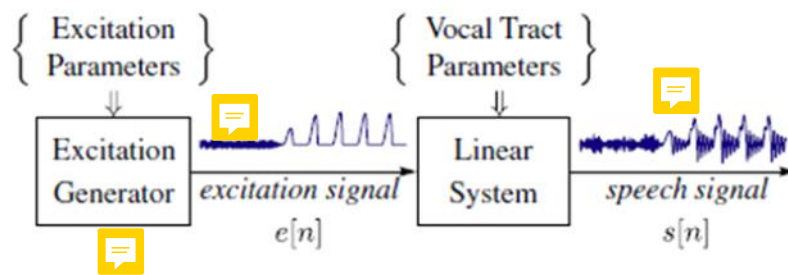


Types of Phonation :  **non-phonemic**; speaker-dependent or controlled

- **Normal** : or modal quality; can change with changing speed of glottal closure
- **Breathy / Whisper** :incomplete closure with posterior portion of the glottis always open; the airflow has periodic + noisy component; extent of breathiness depends on proportion of time vocal folds are open.
- **Creaky/Hoarse**: folds are closed with a small part vibrating with irregular period.
- **Falsetto**: folds are thin and don't close completely; only central part vibrates with high rate.



Pathological voices are rough, hoarse and quantified by measures of aperiodicity including breath noise



**Source/system model for a speech signal**

Ref: Fig 2.2 in *Foundations and Trends in Signal Processing*, Vol. 1, Nos. 1–2 (2007) 1–194 © 2007 L. R. Rabiner and R. W. Schafer

Electronic Larynx

Other source of sound in glottis: Aspiration noise

### "Phonemic" voice quality

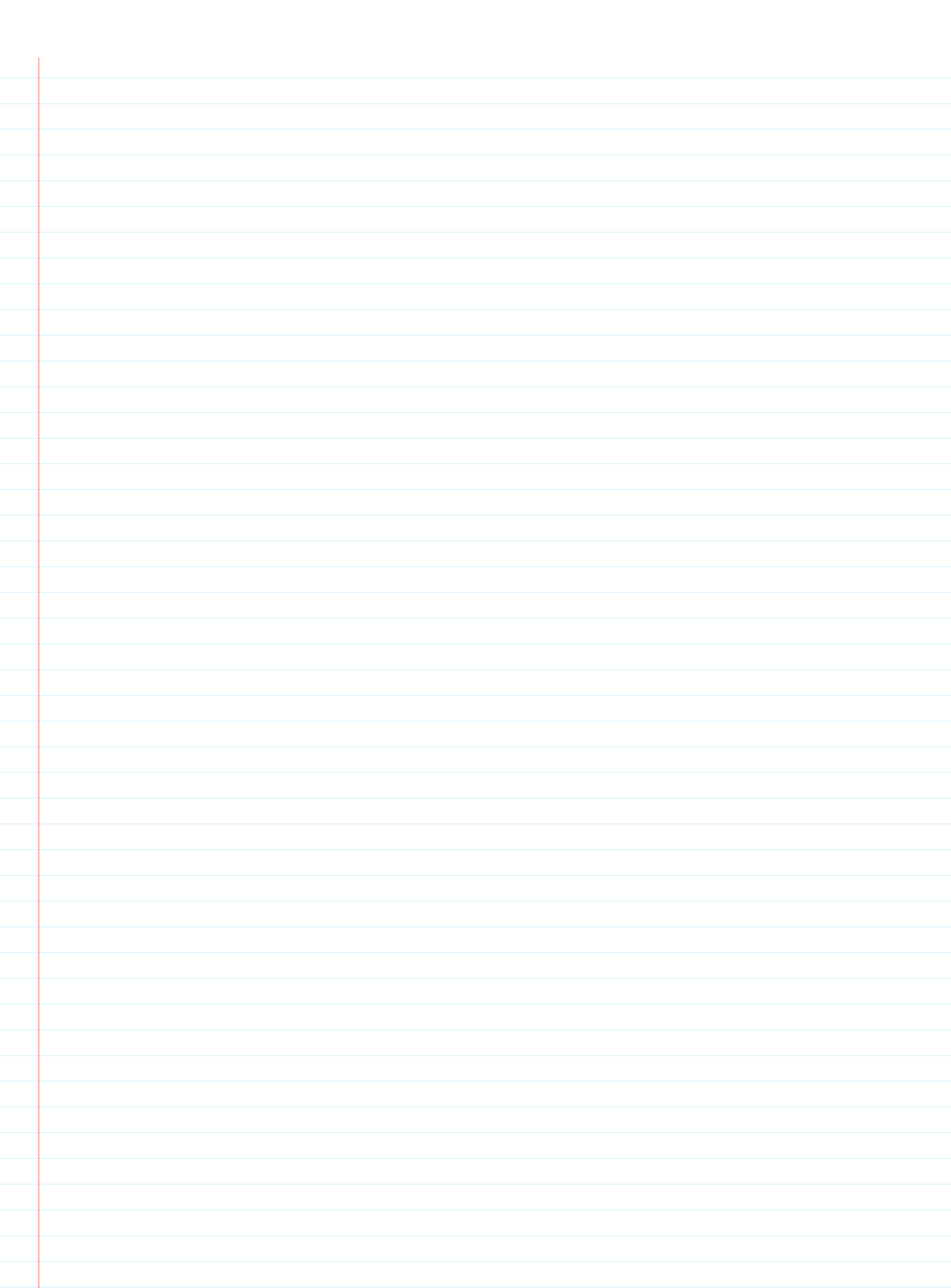
We can divide all speech sounds based on whether produced with vocal folds vibration or without (held open with narrow constriction) into the categories

- Voiced sounds
- Unvoiced sounds

	Vowels	Fricatives	Plosives
Voiced	normal	$z, j, v$	$b, d, g$
Unvoiced	whispered	$s, sh, f$	$p, t, k$

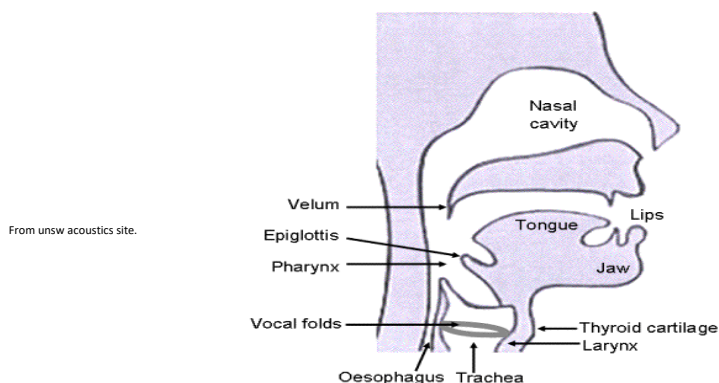






## Articulation

The sound produced at the larynx **passes through the vocal tract** which alters the sound quality based on the selected positions of the articulators (**tongue, jaw, lips, velum**) changing the shape of the vocal tract "resonator".



To appreciate the role of the vocal tract, **change your mouth shape while phonating at constant pitch and amplitude.**

We can now see how we can **independently** control the larynx (source) and vocal tract articulators (filter) for different sounds.

## Vocal tract acoustics

Tube model for vocal tract:

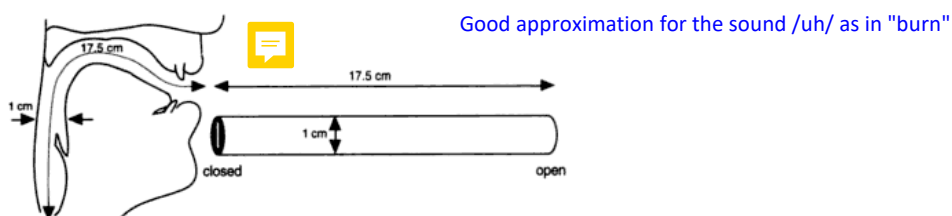
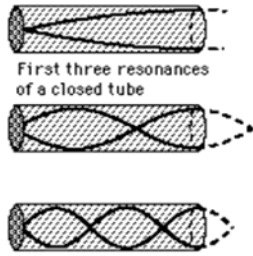


Fig. 8.2. A schematic diagram of a neutral vocal tract in the position for the vowel [ə] on the left, and a simplified version of that shape as a tube closed at one end on the right.

From: Ladefoged, Acoustic Phonetics

We can use the known **expressions for resonances of a tube** of given length and end (open/closed) conditions.

(These known expressions come from solving the Newton's 2nd law for sound propagation in the body to arrive at the constant of proportionality in the Simple Harmonic Motion differential eqn).



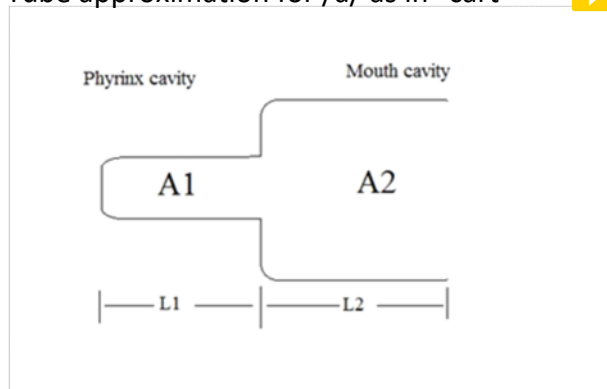
$$L = (2n - 1) \frac{\lambda}{4}$$

$$\lambda \cdot f = C (\text{Speed of sound})$$

For  $L=17.5$  cm,  $C= 340$  m/s  $\Rightarrow f = 500, 1500, 2500 \dots$  Hz



Tube approximation for /a/ as in "cart"



For  $L1 = L2 = 8.75$  cm  $\Rightarrow f = 1000, 3000, 5000 \dots$  Hz

In reality, there are perturbations in above values due to **the coupling** between the tubes. E.g. /a/ tubes' resonances at 1000 are really at 900, 1100 Hz.

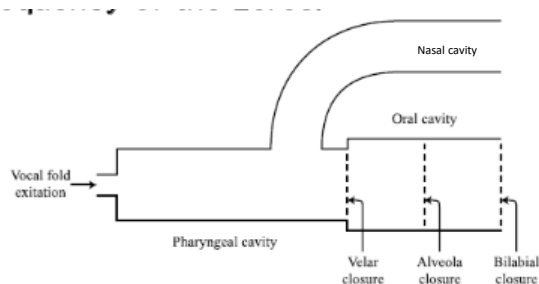


Other vowels; Role of tongue, lips.

Tongue position and height creates the vocal tract cavities. Rounding of lips changes length.

Nasal sounds: Branched resonator

Nasal consonants:



Closure of oral cavity + radiation of sound through nasal cavity.

Oral cavity acts as a side-branch resonator, introducing **zeros** (anti-resonances) based on its length.

[Figure from: UCL phonetics website]



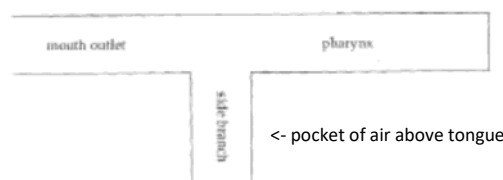
Nasalised vowels:

Both oral and nasal cavities are open and coupled but oral is more open.

Thus nasal cavity acts like an **anti-resonator**.

Laterals, fricatives

Laterals (**l, r**) have a side-cavity that introduces anti-resonances.



<- main cavity curves around tongue

<- pocket of air above tongue

Screen clipping taken: 7/26/2013, 8:38 PM

Unvoiced consonants:




There is a turbulent flow of air through a constriction within the vocal tract. This constriction creates a friction noise source that excites primarily the portion of the vocal tract in front of it. Depending on the place of the constriction we have different sounds: sh, s, f.

Effect of losses in the vocal tract:

Resonances and anti-resonances have zero bandwidth. But in practice, there are losses in the speech production system such as:

yielding (not rigid) walls that vibrate at low frequencies,

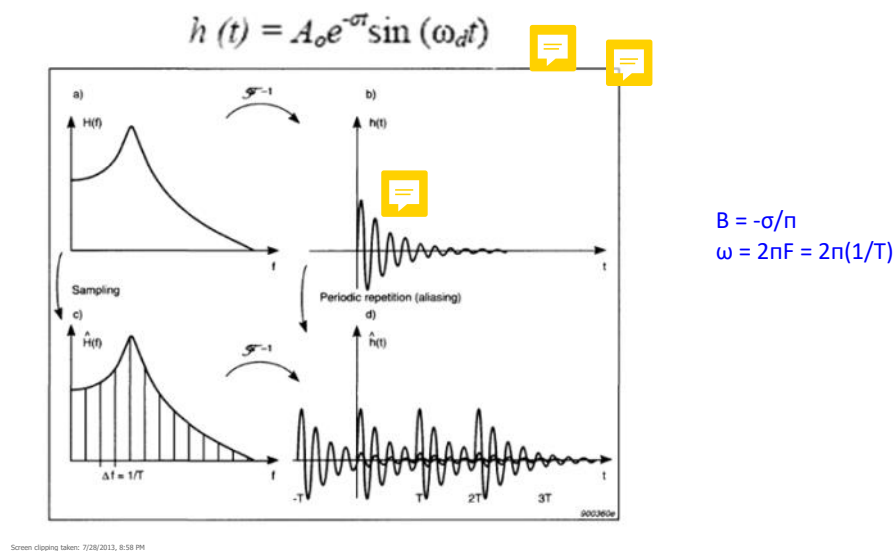
viscous friction between the air and walls and heat conduction through walls,

large yielding surface area of nasal cavity, 

sound radiation at the lips.



Damped resonator: spectrum, waveform



Digital resonator

$$H(z) = \frac{1}{(1 - re^{j\theta} z^{-1})(1 - re^{-j\theta} z^{-1})} \quad \dots (3.15)$$

For given formant frequency  $F_i$  Hz and bandwidth  $B_i$  Hz, we have for sampling period  $T$ :

$$\theta_i = 2\pi \cdot F_i \cdot T$$

$$r_i = e^{-\pi B_i T}$$

#### Lip radiation:

The lips form a small opening so that diffraction (bending) of large wavelengths (low frequencies) takes place while high frequencies are directed in front => lip radiation is modeled by high-pass filter.

#### Source-filter model of speech production

Also applies to musical instruments...

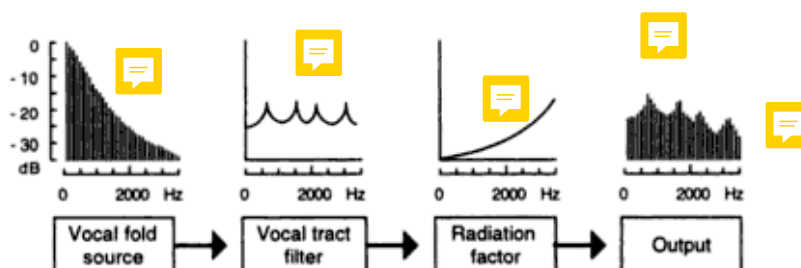
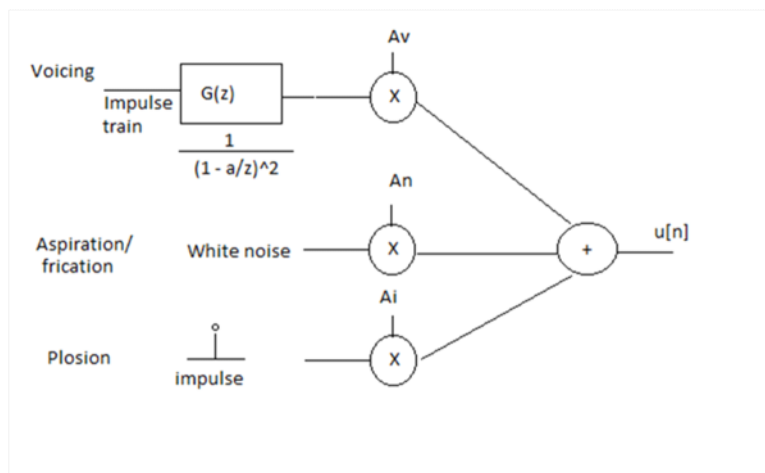


Fig. 7.7. A source-filter view of the production of a vowel.

For consonant phones:



<---- Voicing and manner

**Acoustic phonetics:** the differentiation of sounds on an acoustic basis. The acoustics are more evident spectrally rather than in the time domain.