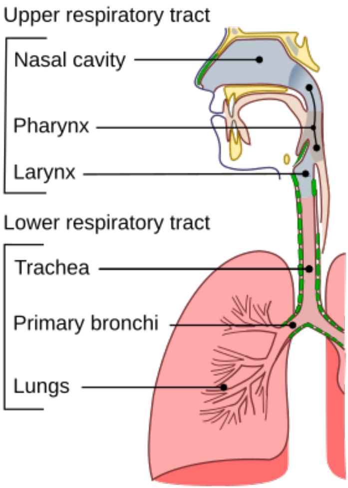
Lecture 6: Vocal Fold Physiology

Mark Hasegawa-Johnson

ECE 537: Speech Processing

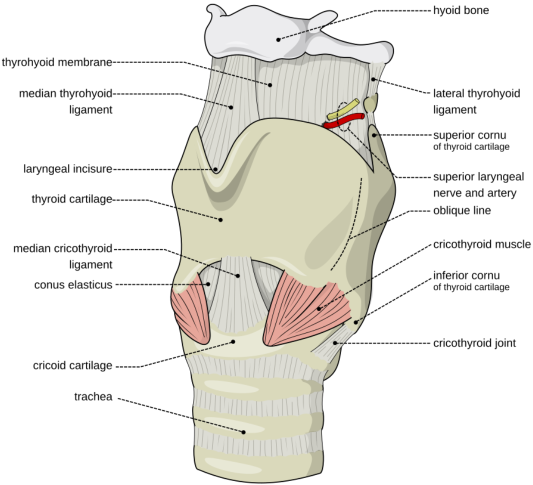
Outline

Physiology of the human airways



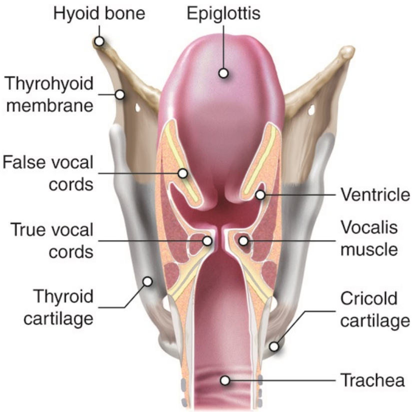
The larynx contains the vocal folds. Their function is to close when you eat, protecting your lungs from food.

Human larynx, external view



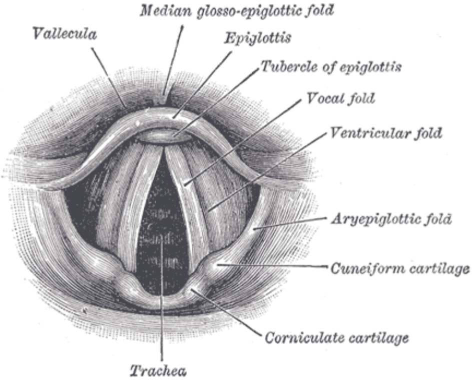
Larynx = thyroid cartilage over cricoid cartilage. Thyroid tilts up to stretch the vocal folds, down to relax the vocal folds.

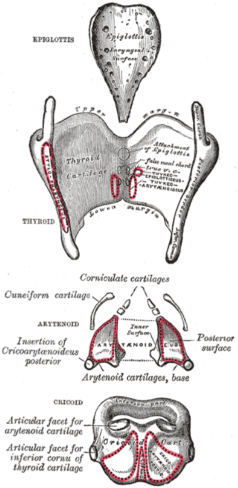
Human larynx, posterior cutaway



Vibrating portion = mucosal surface layer. Vocalis muscle is mostly too heavy to vibrate.

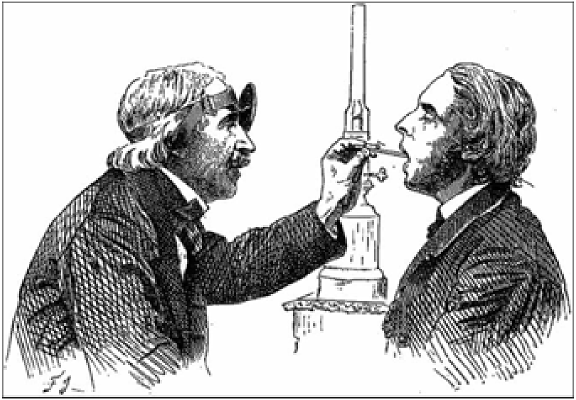
Human larynx, superior view





Vocal folds connected to thyroid in front, and to arytenoid in back. Arytenoids can be opened to produce breathy voice or whisper, or just to breathe.

Laryngoscope



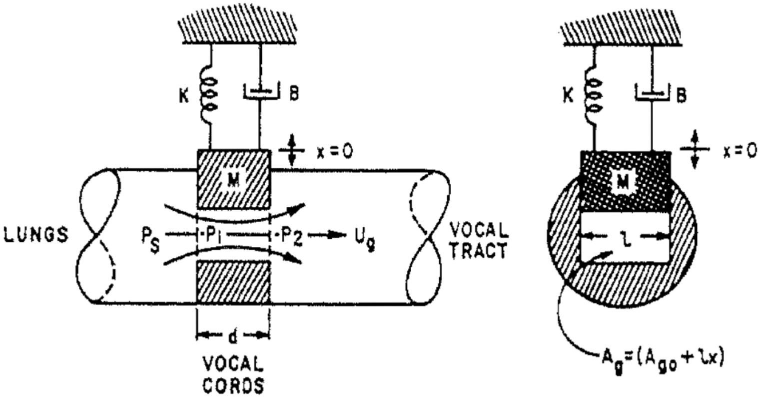
A laryngoscope is a light and a mirror or optical fiber. If the light is a strobe with period $0.99\,T_0$, you can sample one image per period, producing a 100-fold slowed video.

Laryngoscopic video

Here are full-speed and strobe-slowed videos of phonation:
https://commons.wikimedia.org/wiki/File:Stroboscopy_Normal_Female_Vocal_Cords.webm

Outline

The Flanagan-Landgraf One-Mass Model of Vocal Fold Vibration



The Flanagan-Landgraf One-Mass Model of Vocal Fold Vibration

For small displacements, vocal fold stiffness and damping are approximately linear, so

$$M\ddot{x} + B\dot{x} + Kx = F$$

Force is the average of entry and exit pressure, multiplied by length (l) and depth (d) of vocal folds

$$F = \frac{ld}{2}\left(P_1 + P_2\right)$$

Natural frequency of the vocal folds

Equations on the previous slide suggest that the vocal folds have a natural frequency:

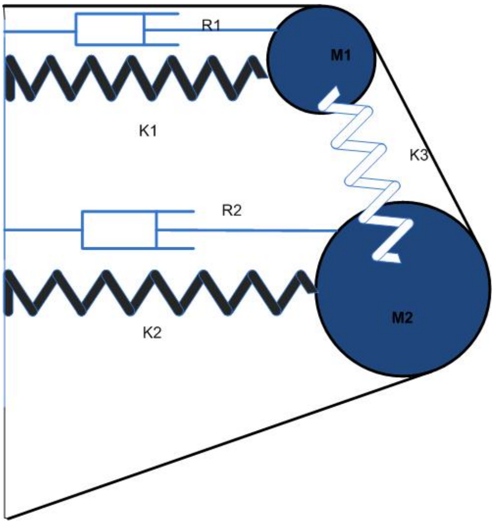
$$\omega_g = \frac{1}{\sqrt{KM}}$$

- Experimental measurements show that $\omega_g \approx \omega_0$, though ω_g is usually slightly larger.
- The difference is caused by energy loss when the vocal folds close.
- M correlates with inter-person differences in average pitch.
- K can be adjusted rapidly and fluently by rocking the thyroid cartilage

Missing from this model: Sustained voicing

- Sustained voicing requires some nonlinearity to prevent steady-state. Air rushing out of the vocal tract would be enough, if the vocal tract were more open.
- What actually sustains voicing is the flapping of the vocal folds, which is modeled by the Ishizaka-Flanagan (1972) two-mass model: two coupled oscillators inside each vocal fold.
- Other models represent partial opening and closure (the two-beam model) and detailed vibration patterns (the finite element model).

Ishizaka-Flanagan Two-Mass Model of the Vocal Folds



Outline

The acoustic terms

- Stiffness, damping, and mass of the vocal folds explain pitch frequency, but don't explain how voicing is sustained. Sustained oscillation requires some nonlinearity that prevents the vocal fold oscillation from exponentially decaying to zero.
- Ishizaka & Flanagan (1972) showed that the flapping of the vocal folds (the two-mass model) is enough to sustain voicing.
- Flanagan & Landgraf (1968) thought that acoustic mass sustained voicing. They were wrong (it would if the vocal tract were more open), but acoustic mass is important for speech production, so let's learn about it.

Acoustic flow variables

- Acoustic flow ($U(t)$), measured in liters/second or m^3/s, is air velocity ($v(t)$) times cross-sectional area (A).

$$U(t) = v(t)A$$

- Pressure ($P(t)$) is force per unit area (of a fluid on its container), measured in Pascals.
- Density (ρ) is g/cm^3 or kg/m^3.

Acoustic mass = Newton's second law of motion, $f = ma$

- The block of air in the glottis has a mass of $\rho d A_g$.
- The force on this block of air is $(P_1 - P_2)A_g$.
- The acceleration of this block of air is $\frac{1}{A_g}\dot{U}_g$.
- Putting those together, Newton's second law is:

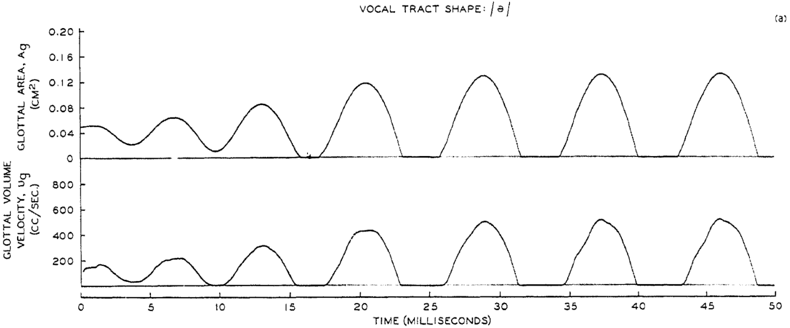
$$P_1 - P_2 = \frac{\rho d}{A_g}\dot{U}_g = L_g \dot{U}_g$$

- We define the acoustic mass $L_g = \rho d A_g^{-1}$ to get a simple equation $\Delta P = L_g \dot{U}_g + \text{losses}$

What is the nonlinearity?

- The most important nonlinearity in vocal fold vibration is the closure of the glottis.
- When the glottis closes, the downstream flow keeps moving, causing a negative pressure spike downstream from the glottis.
- "Steady-state" would be if the vocal tract reached lung pressure. That never happens.

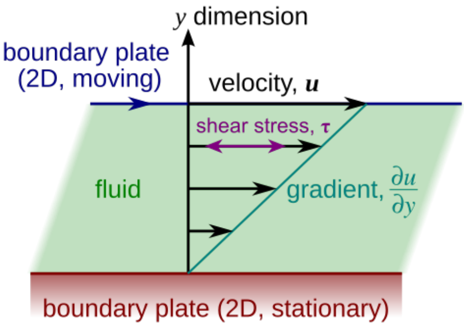
Example glottal waveforms



What about the air pressure drop across the glottis?

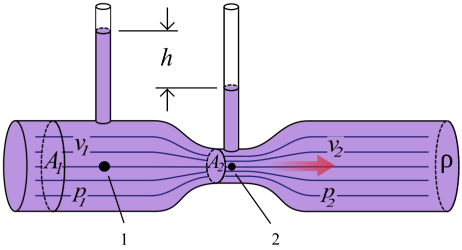
- The glottis is the narrowest constriction between lungs and lips, by a factor of ten or more.
- Subglottal air pressure $\approx$ lung pressure, supraglottal air pressure $\approx$ atmospheric pressure.
- Acoustic mass is not enough to explain this.
- Most of the air pressure is spent on viscous losses and kinetic losses.

Viscous resistance



https://en.wikipedia.org/wiki/Viscosity

Kinetic resistance: Bernoulli effect



https://en.wikipedia.org/wiki/Bernoulli's_principle

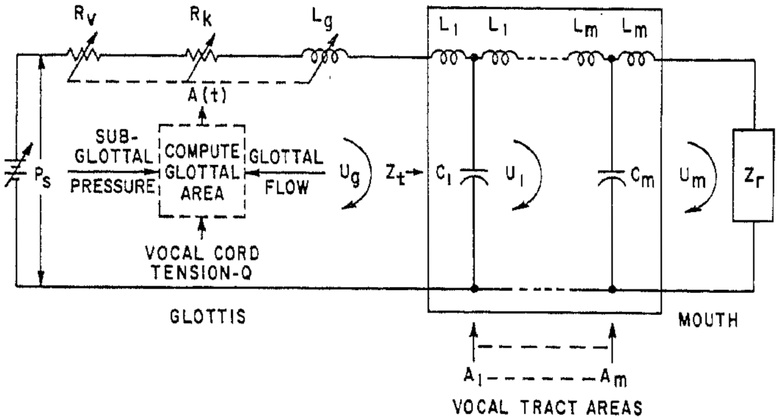
Losses

- Viscous losses: Air "sticks" to the vocal folds; its velocity is maximum only in the center.
- Kinetic losses (Bernoulli effect): In order to get the flow through a narrow contraction, every individual air molecule needs to be accelerated from 0 velocity up to U_g/A_g.

$$\begin{aligned}\Delta P &= \rho d \frac{d}{dt}\left(\frac{U_g}{A_g}\right) + 12\mu d l^2 \frac{U_g}{A_g^3} + 0.44\rho\left(\frac{U_g}{A_g}\right)^2 \\ &= L_g \dot{U}_g + (R_v + R_k) U_g\end{aligned}$$

Outline

The complete acoustic circuit model

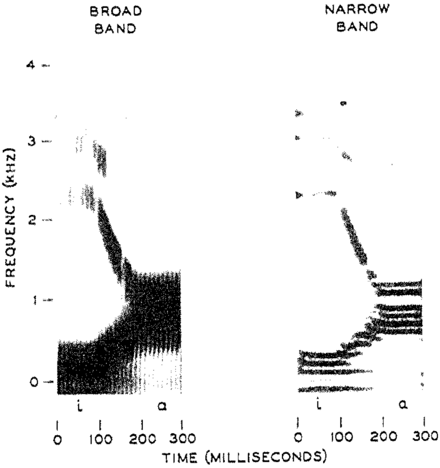


Speech synthesis in Flanagan and Landgraf, 1968

- The vocal folds are modeled using the one-mass model.
- The vocal tract is modeled using a sequence of resonators, one for each resonant frequency of the vocal tract.
- The flow coming from the mouth, U_m, produces the air pressure at the mouth, P_m, according to $P_m = Z_r U_m$, where Z_r is the radiation impedance. Z_r is mostly Newton's law, so

$$P_m(t) \approx |Z_r| \frac{dU_m}{dt}$$

Example synthesized speech



Outline

Summary

- Mass and stiffness of the vocal folds govern pitch frequency:

$$M\ddot{x} + B\dot{x} + Kx = F = \frac{ld}{2}(P_1 + P_2)$$

- The pressure drop across the glottis is caused by acoustic mass, viscous resistance (linear: R_v independent of U_g), and kinetic resistance (nonlinear: $R_k \propto |U_g|$):

$$\Delta P = L_g \dot{U}_g + (R_v + R_k) U_g$$

- Phonation is sustained because energy from the lungs doesn't build up to steady state; it is dissipated through inertia of air in the vocal tract (one-mass model) and flapping of the vocal folds (two-mass model)