# Vocal Fold Modeling

Martin Hagmüller

hagmueller@inw.tugraz.at

Signal Processing and Speech Communication Laboratory
Institute for Communication and Wave Propagation
Graz University of Technology

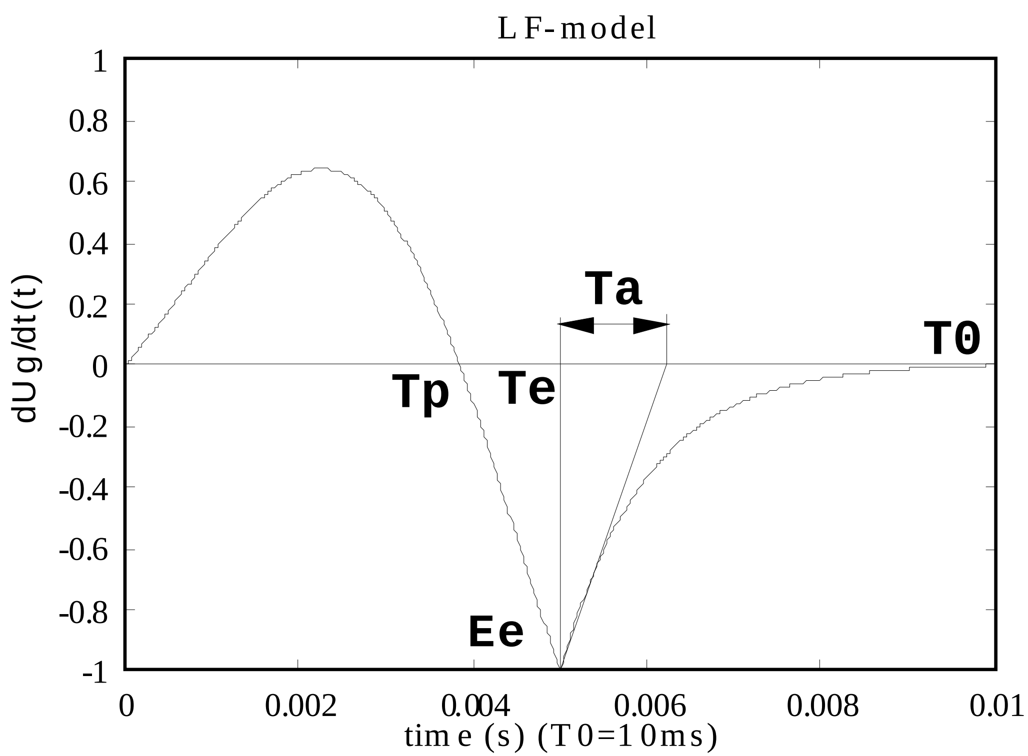
TUG

# Overview

- Motivation
- Anatomy and Physiology
- Vocal Fold Models
- Conclusion

# Motivation (1)

- Waveform Model
  - Parametric description of waveform
    e.g. Liljencrants-Fant model
    flow derivative with four parameters



# Motivation (1)

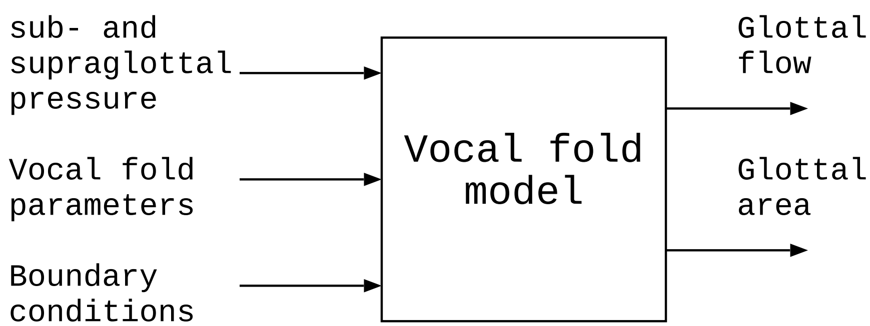
- Waveform Model
  - Parametric description of waveform
    e.g. Liljencrants-Fant model
    flow derivative with four parameters
- Physiologically motivated voice generation
  - Approximation of human voice generation
  - Flexible tool to generate different voices

# Motivation (1)

- Waveform Model
  - Parametric description of waveform
    e.g. Liljencrants-Fant model
    flow derivative with four parameters
- Physiologically motivated voice generation
  - Approximation of human voice generation
  - Flexible tool to generate different voices
- Study of physiology of voice
  - Better understanding of physiology
  - Study of voice source dynamics

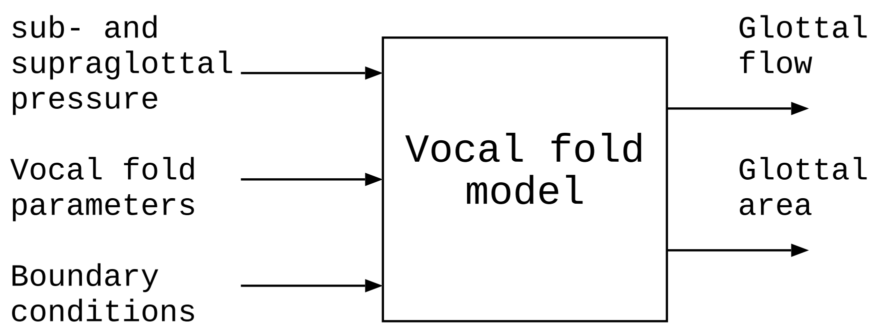
# Motivation (2)

- Voice Source
  - Voice Quality (Individuality of Voice)
  - Prosody (Speech Melody)



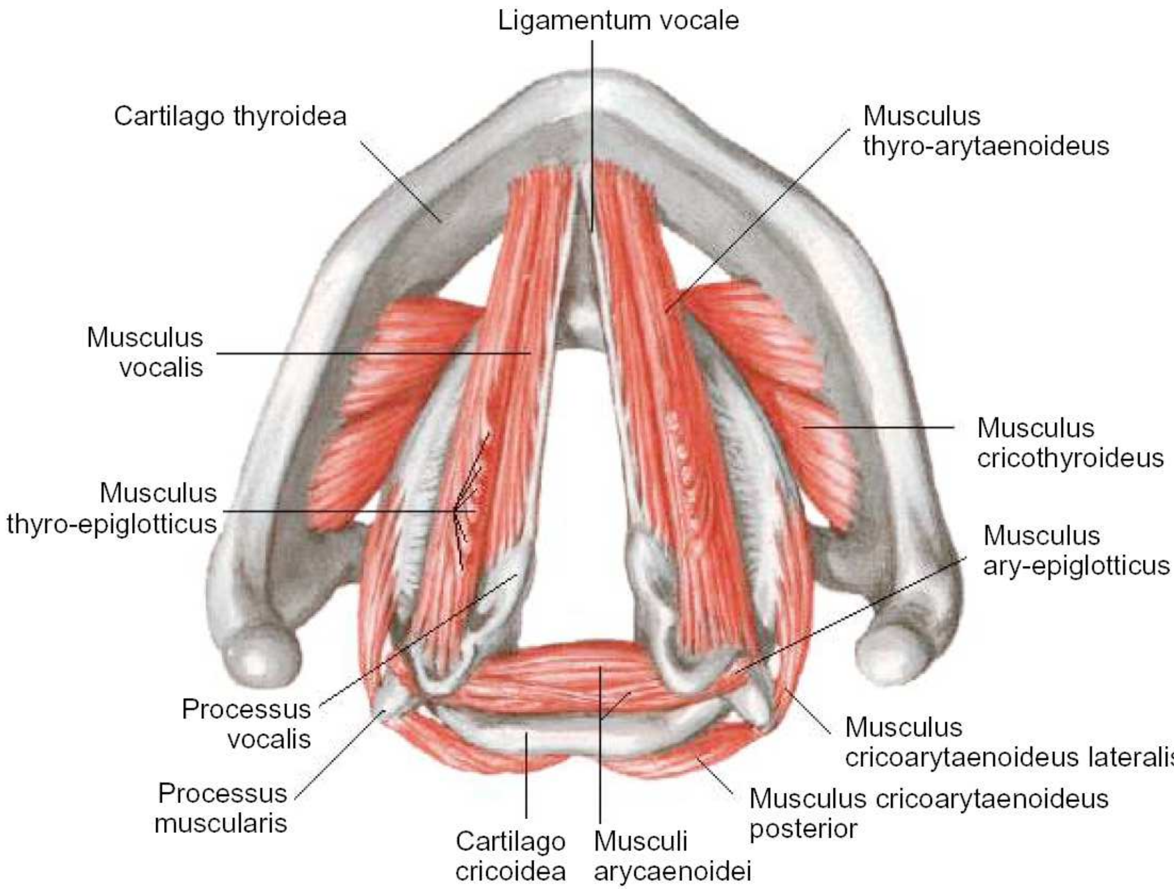
# Motivation (2)

- Voice Source
  - Voice Quality (Individuality of Voice)
  - Prosody (Speech Melody)

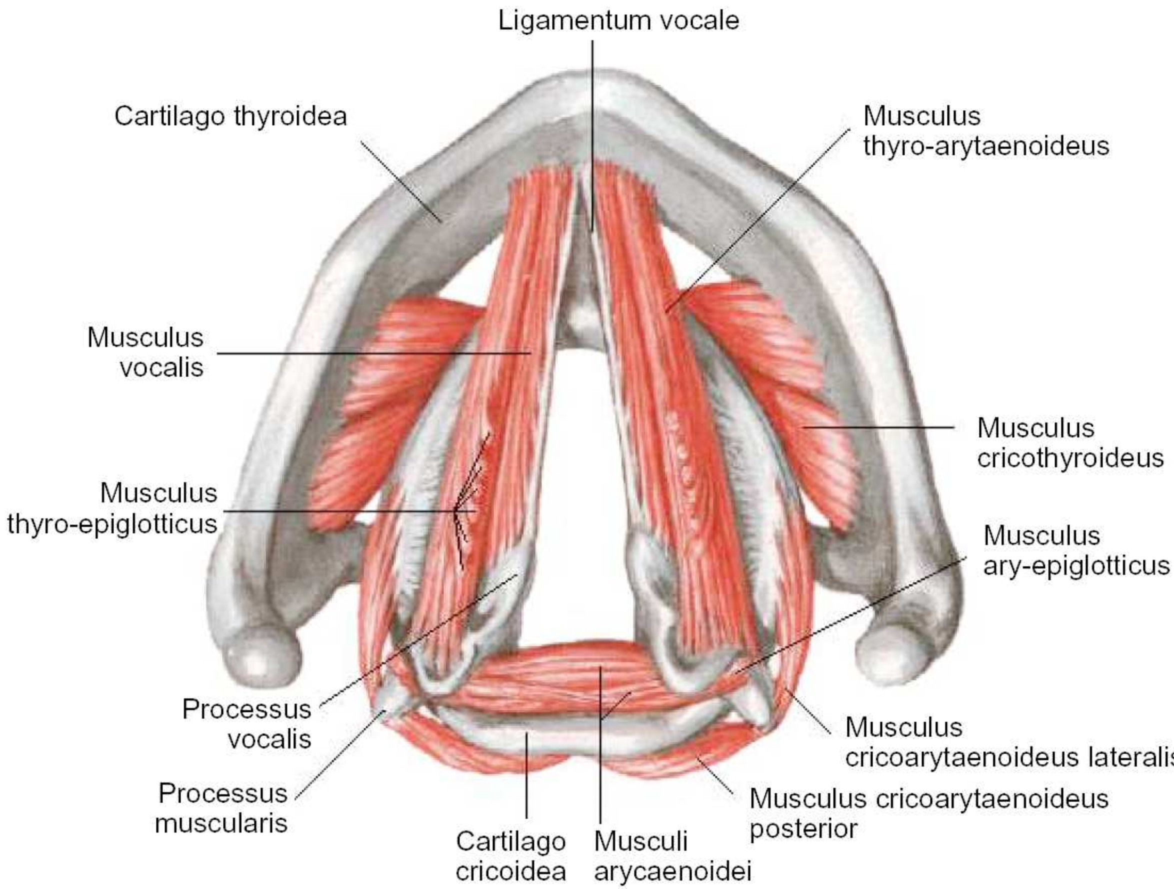



- Disadvantages
  - Many parameters
  - No intuitive mapping to perception
  - Hard to control
  - Based on measured data → hard to obtain

# Anatomy & Physiology (1)



# Anatomy & Physiology (1)

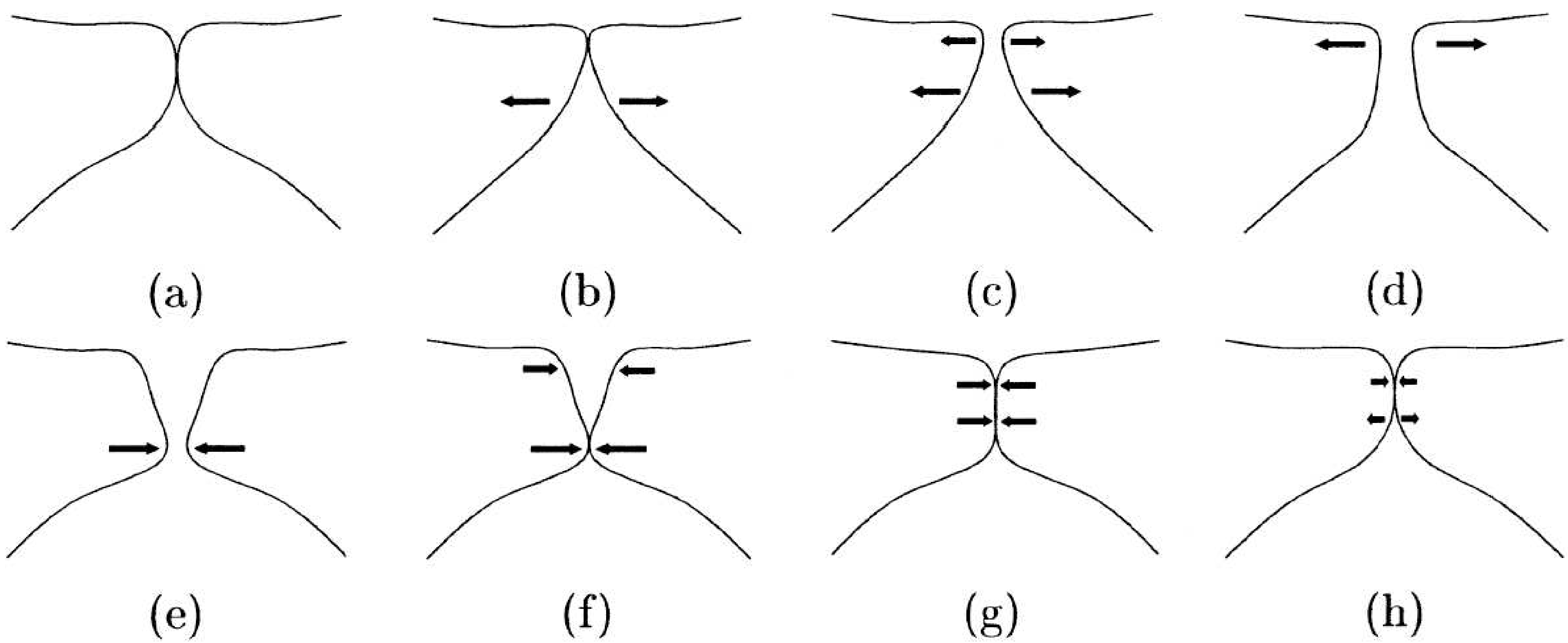



play movie

# Anatomy & Physiology (1a)

- Self oscillating system
- Steady airflow from lungs is converted into series of flow pulses by opening & closing airspace between vocal folds *(glottis)*
- ”Myoelastic-aerodynamic theory of voice production (Van den Berg, 1958)
  $\longrightarrow$ Oscillation due to Bernoulli effect

# Anatomy & Physiology (2)

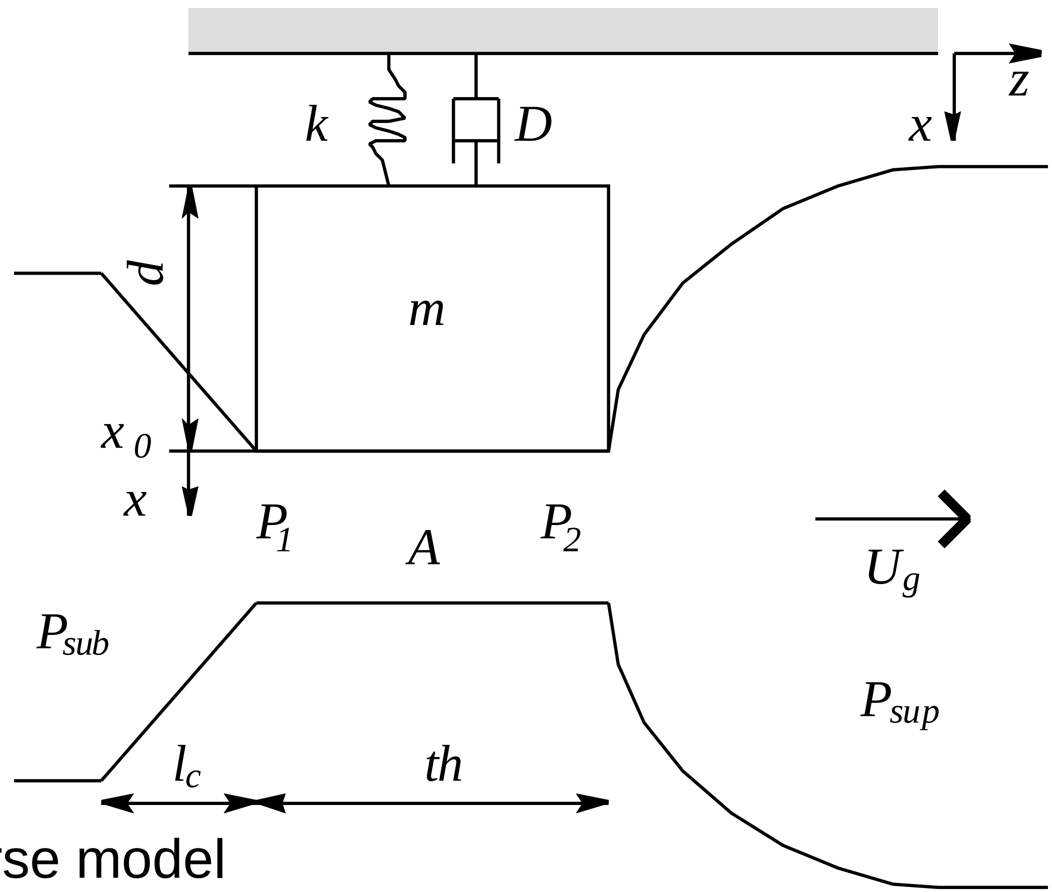



- Upper and lower portion display phase difference
  $\rightarrow$ *Mucosal wave*

# Vocal Fold Models

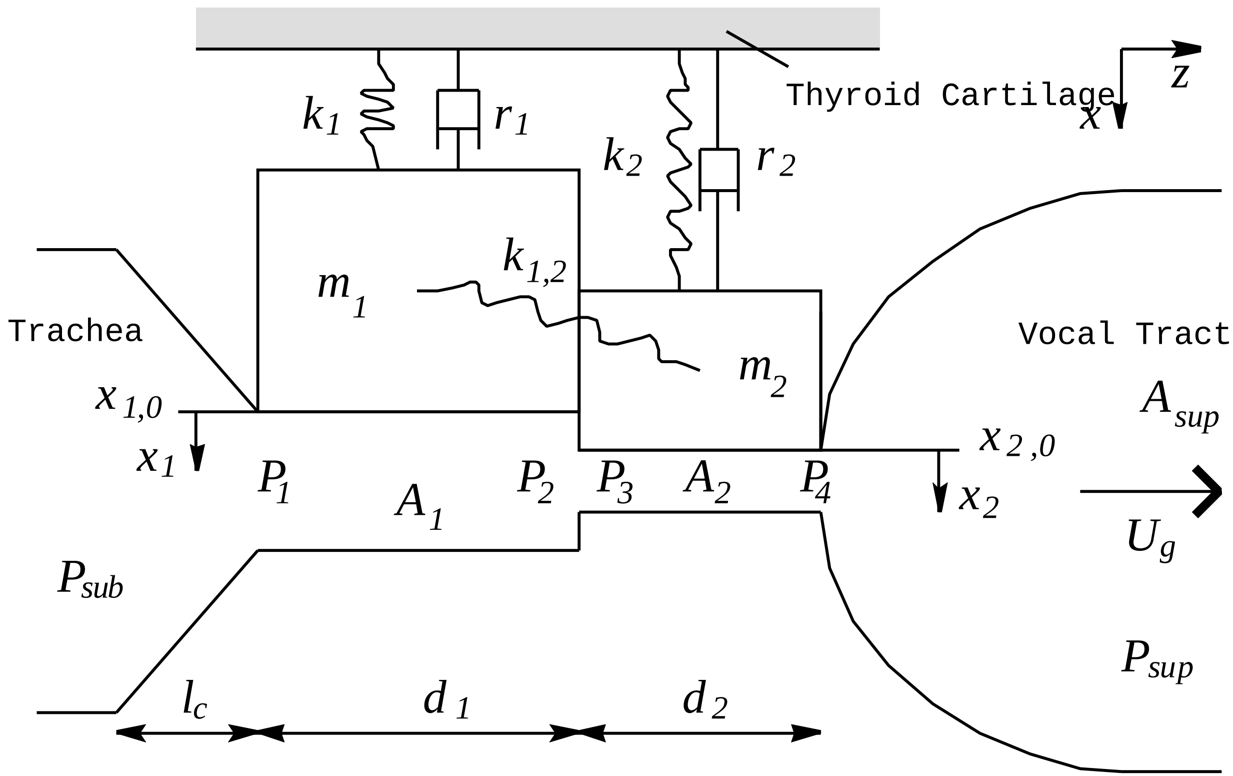
- Lumped elements models (n-mass models)
  - one mass model (Flanagan & Landgraf, 1968)
  - two mass model (Ishizaka & Flanagan, 1972)
  - three mass model (Story & Titze, 1994)
- Physically informed model
- Finite elements model

# One Mass Model

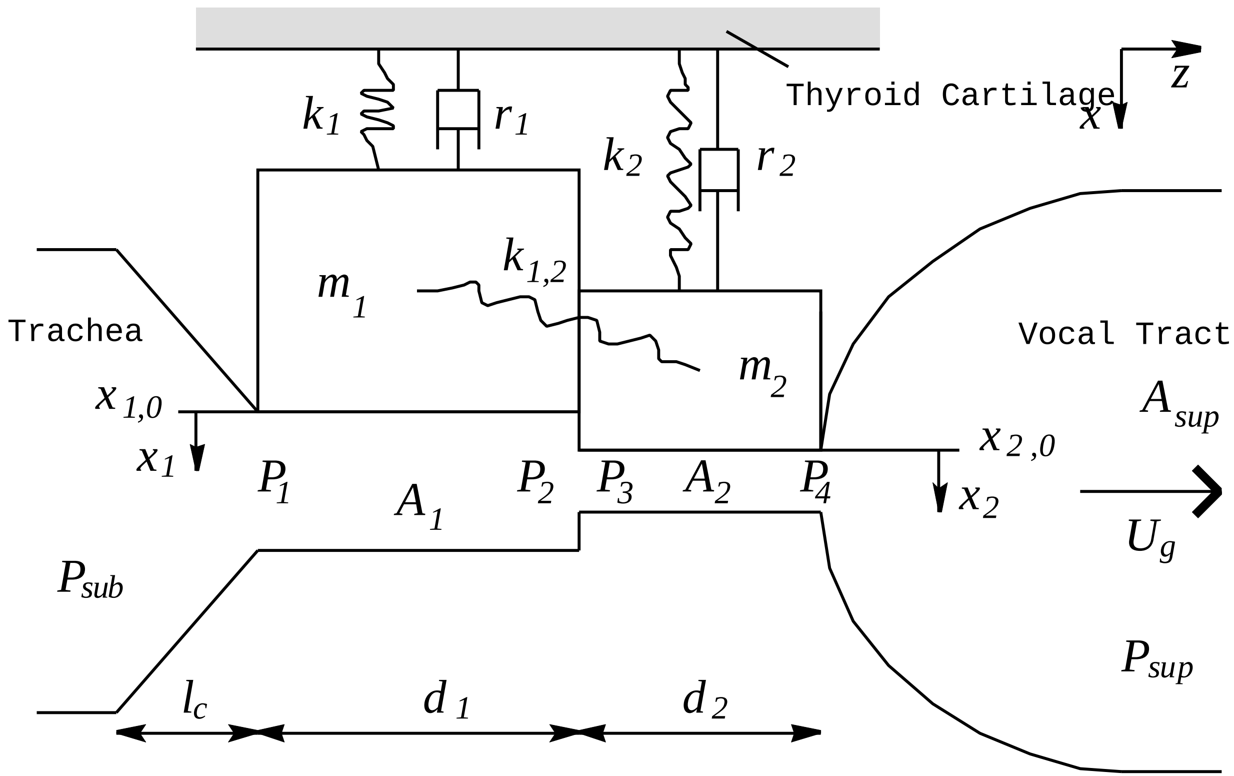



- Very coarse model
- Vocal fold lumped into one mass.
  Only lateral displacement
- Modelling of basic principles of self-sustained oscillation

# Two Mass Model - Mechanics

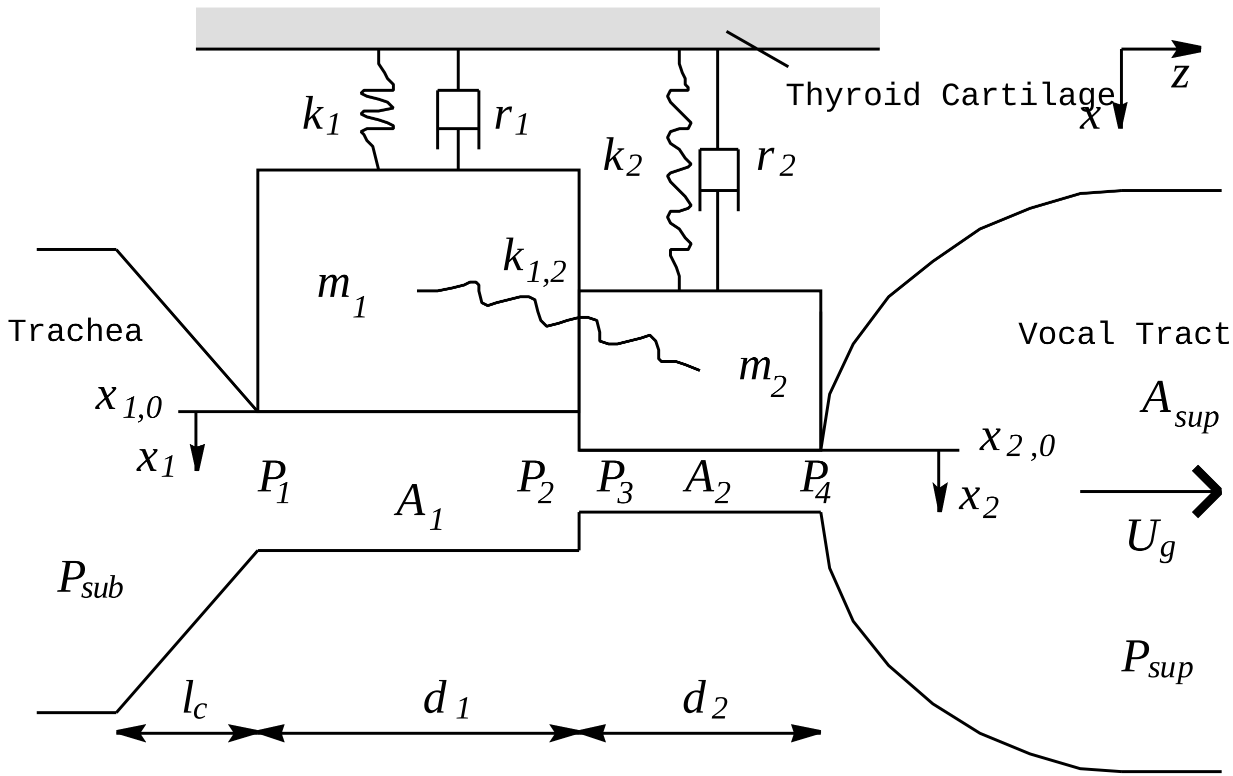


# Two Mass Model - Mechanics




- Includes representation of mucosal wave
- Reasonable agreement with physiological data
- Very popular

# Two Mass Model - Mechanics

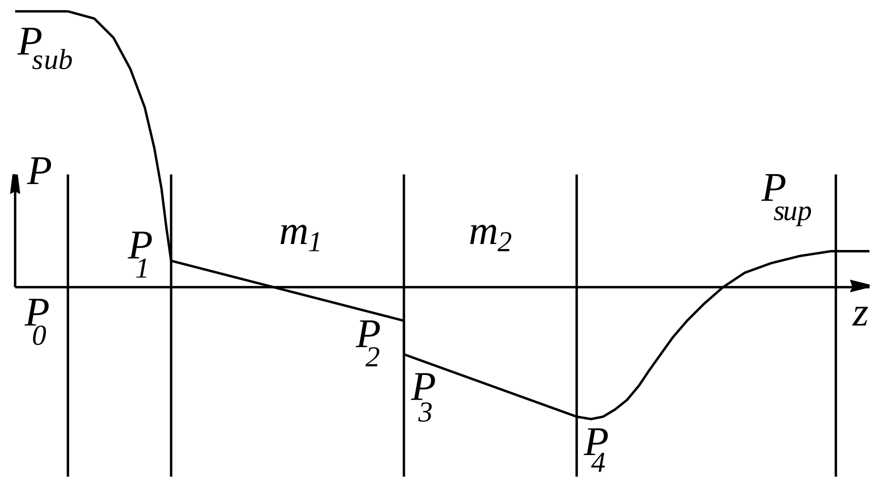



Equations for the mechanical system:

$$
\begin{aligned}
m_1\ddot{x}_1(t) + r_1\dot{x}_1(t) + k_1(x_1)[x_1(t) - x_{01}] + k_{12}[x_1(t) - x_2(t)] &= l_g d_1 p_{m1}(t) \\
m_2\ddot{x}_2(t) + r_2\dot{x}_2(t) + k_2(x_2)[x_2(t) - x_{02}] - k_{12}[x_1(t) - x_2(t)] &= l_g d_2 p_{m2}(t)
\end{aligned}
$$

# Two Mass Model - Pressure

Pressure distribution along glottis:
(assumes continuity of the fbw)



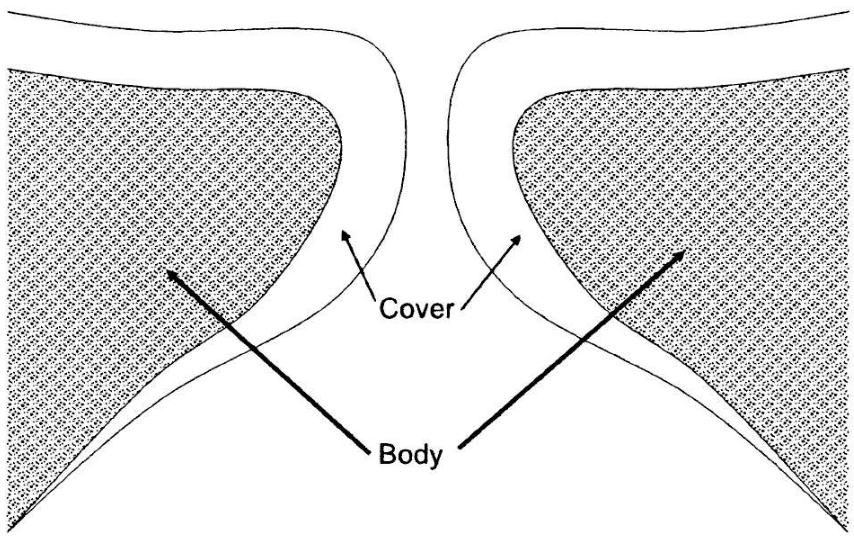

$$
\begin{aligned}
p_{sub} - p_1(t) &= 1.37 \frac{1}{2} \rho_{air} \frac{u(t)^2}{A_1(t)^2} \\
p_1(t) - p_2(t) &= 12 \nu d_1 \frac{l_g^2 u(t)}{A_1(t)^3} \\
p_2(t) - p_3(t) &= \frac{1}{2} \rho_{air} u(t)^2 \left( \frac{1}{A_2(t)^2} - \frac{1}{A_1(t)^2} \right) \\
p_3(t) - p_4(t) &= 12 \nu d_2 \frac{l_g^2 u(t)}{A_2(t)^3} \\
p_4(t) - p_{sup}(t) &= \frac{1}{2} \rho_{air} \frac{u(t)^2}{A_2(t)^2} \left[ 2 \frac{A_2(t)}{A_{sup}} \left( 1 - \frac{A_2(t)}{A_{sup}} \right) \right]
\end{aligned}
$$

$\nu$ . . . air shear viscosity coefficient

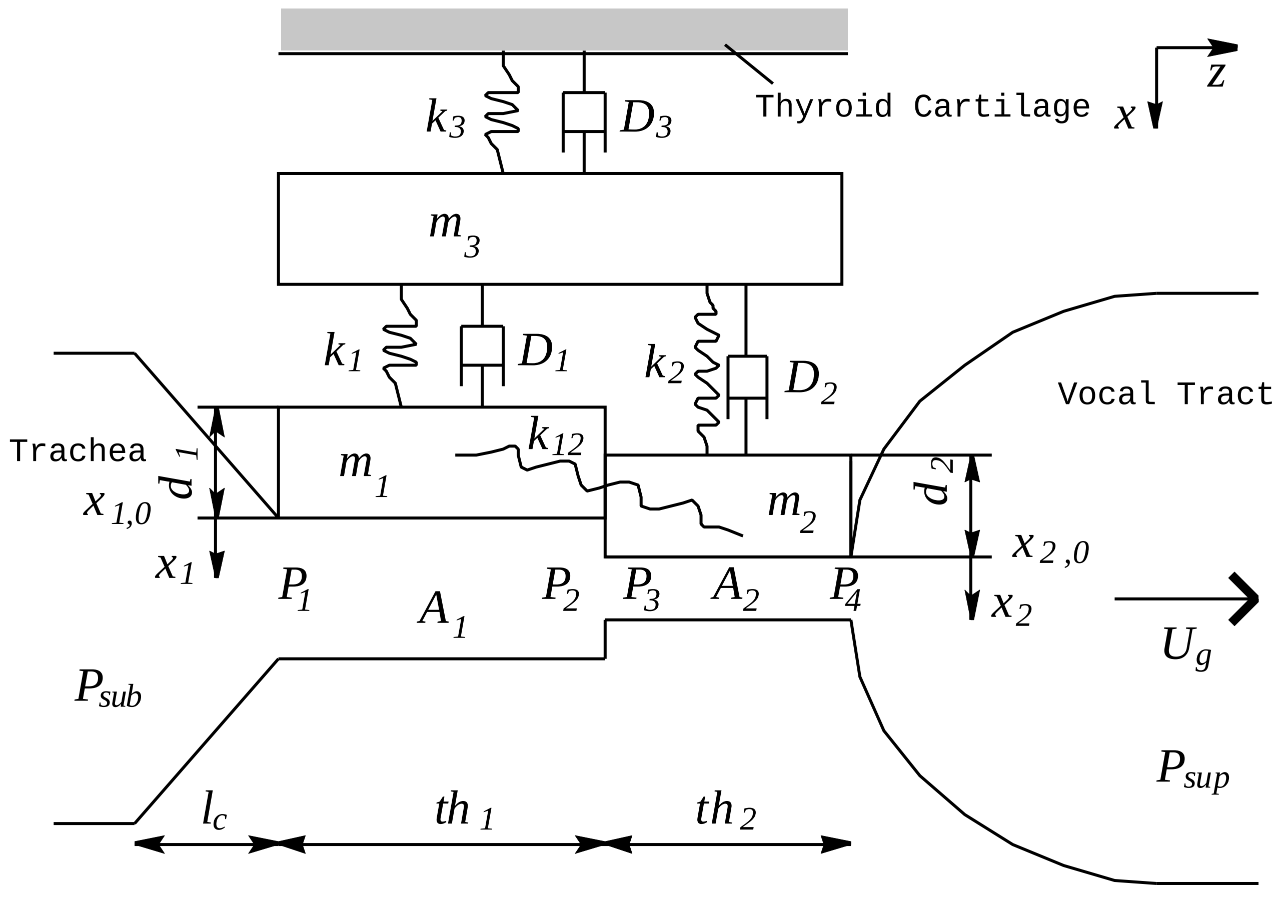
$\rho_{air}$ . . . air density

# Body – Cover Concept

- Vocal folds are divided into 2 tissue layers
  - Cover: pliable, non-contractible mucosa tissue
  - Body: muscle fiber & ligamentous tissue

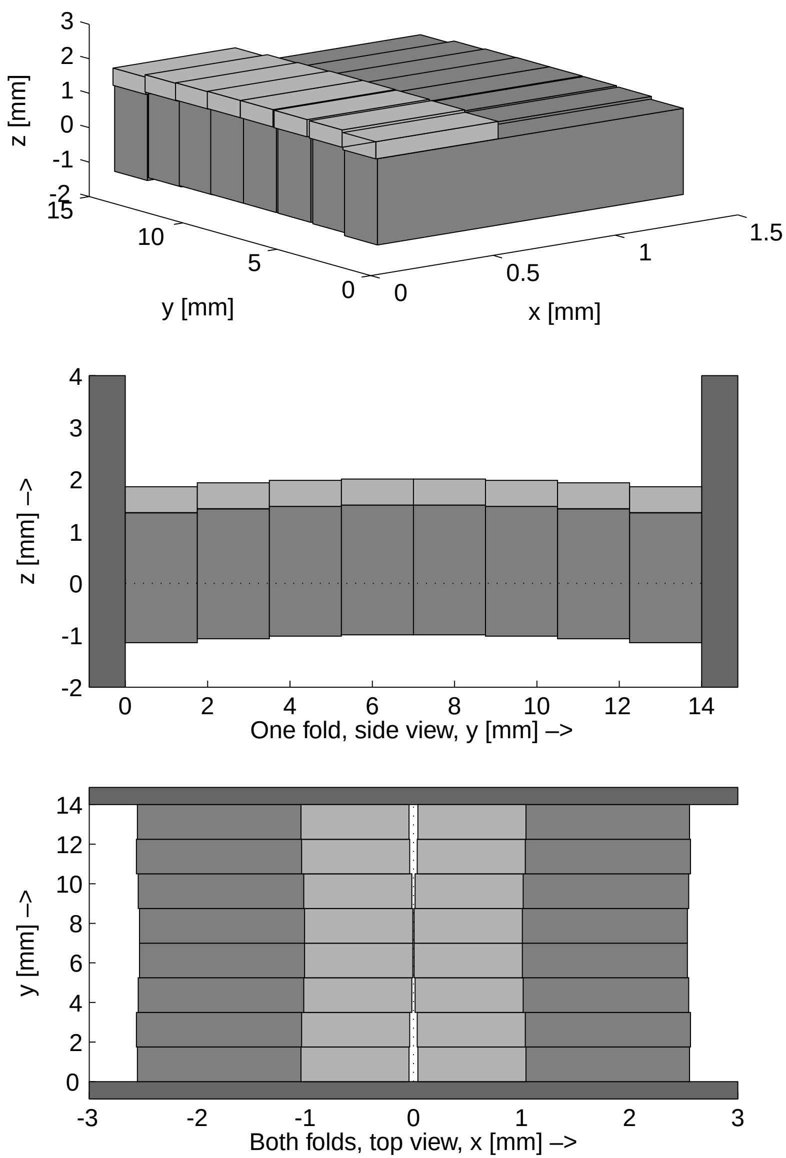


# Three Mass Model




- $m_3$: Body mass
- $m_1, m_2$: Cover masses

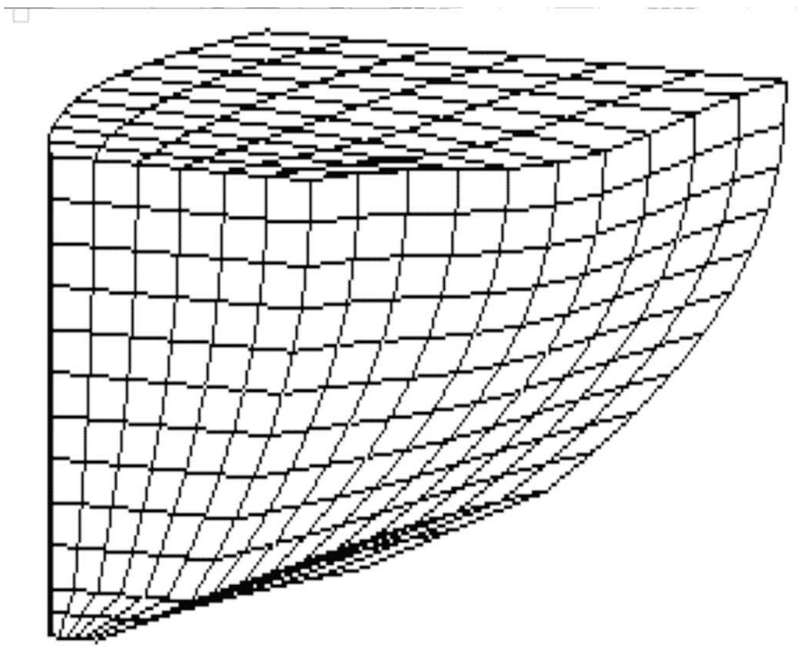
# Multiple Mass Model




+ Possible modification of properties in longitudinal dimension
- Increased complexity
- High calculation times

# Finite Elements Model

Alipour & Titze 1996



- Does not assume lumped elements
- Describes geometry & phys. properties of vocal folds with huge number of very small elements

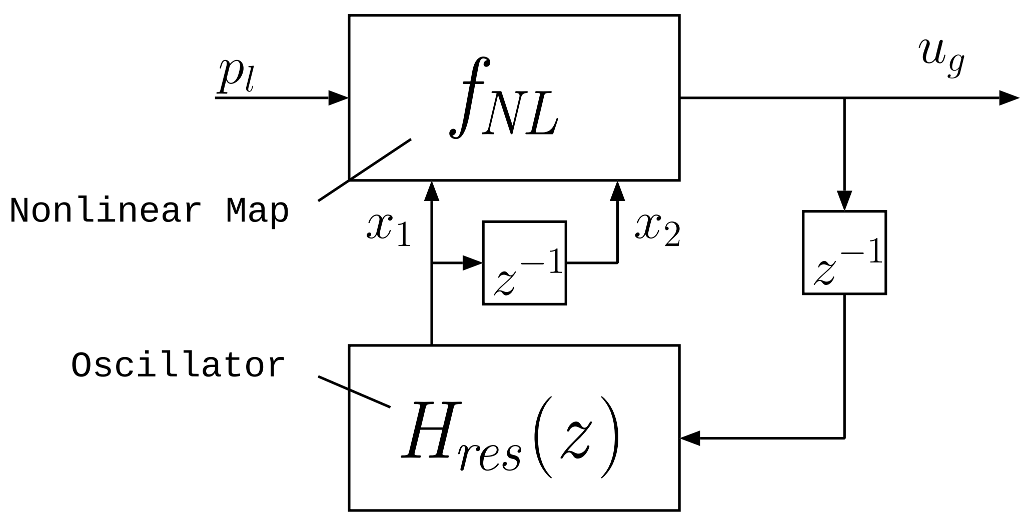
+ Very good solution if all properties are known
- Calculation times very high
- Real-time models not possible (yet)
- huge number of parameters (hard to determine)

# Physically-informed model

- Physical model not for entire system
  - Mechanical Model of Oscillator

    (calculation of displacement, fundamental frequency)
  - Nonlinear Map for calulation of glottal flow

    (interaction of vocal fold motion $x_i$, flow $u_g$ and pressure $p_l$)



# Conclusion

- Large Variety of Vocal Fold models
- Trade off between Accuracy & Complexity
  reliable data & level of detail
- Rather study of physiology than speech synthesis

# Small dictionary

| | |
|---|---|
| mucosa | Schleimhaut |
| tissue | Gewebe |
| ligament | Band |
| cartilage | Knorpel |
| trachea | Luftröhre |
| inertia | Massenträgheit |
| prosody | Sprachmelodie |
| lumped elements | konzentrierte Elemente |

# References

[1] J. Flanagan and L. Landgraf. Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics*, 16(1):57–64, March 1968.

[2] K. Ishizaka and J. L. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Systems Techn. J.*, 51(6):1233–1268, 1972.

[3] Brad H. Story and Ingo R. Titze. Voice simulation with a body-cover model of the vocal folds. *J. Acoust. Soc. Am.*, 97(2):1249–1260, 1995.

# THANK YOU