Modelling Speech Biophysics

Jarmo Malinen

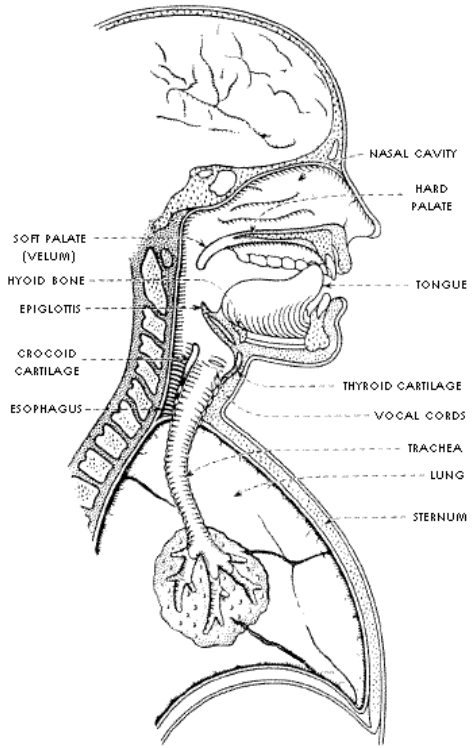
Aalto University, School of Science,
Department of Mathematics and Systems Analysis

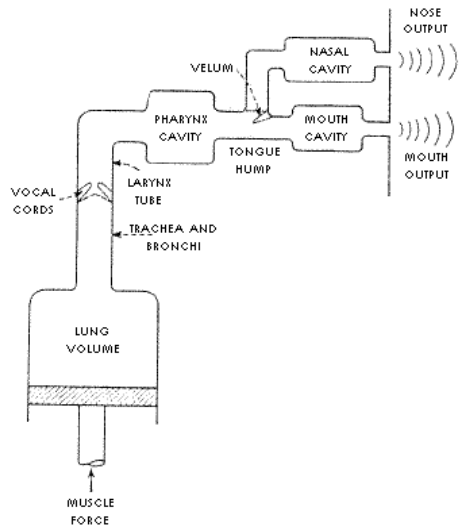
IFAC LHMNLC'15
July 4-7, 2015, Lyon

Human voice production

Simplified vowel production:

Mouth
(exterior load)
↑
Vocal tract
(filter)
↑
Vocal folds
(source)





Flanagan, J. L. (1972). Speech Analysis Synthesis and Perception, Springer-Verlag.

- Vocal tract (VT) shape changes, and there are feedbacks.
- Not all speech sounds originate in vocal folds.

General rules of the game

Modelling speech from *first principles* requires at least:

- Vibrating vocal fold tissues $\rightarrow$ elasticity and mechanics.
- Air flow $\rightarrow$ flow mechanics and aerodynamics.
- Vocal tract, subglottal tract, environment $\rightarrow$ acoustics.

Possible approaches:

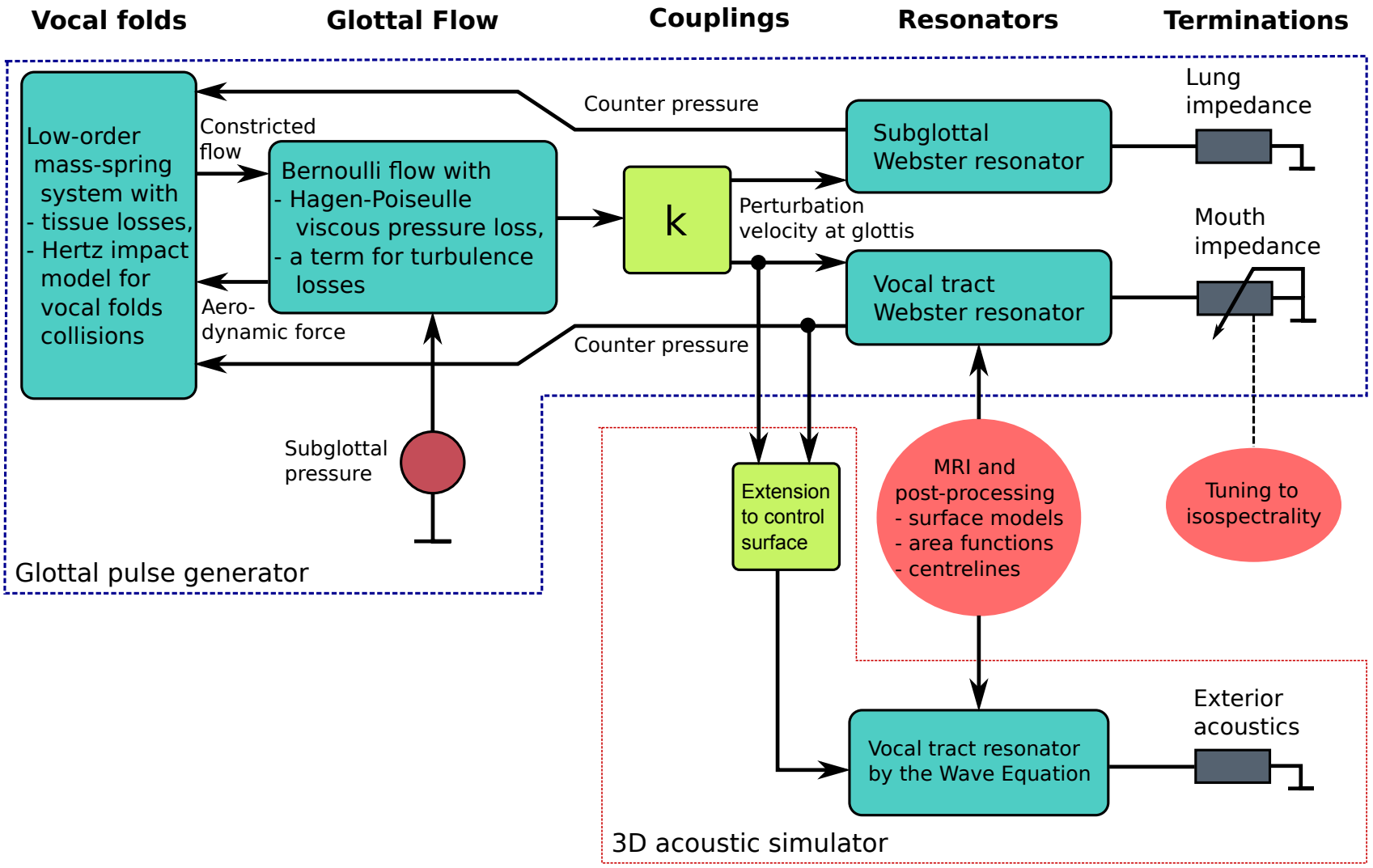
- All-out CFD solver with elastic boundaries at tissue walls, or
- a hybrid multi-physics model having a modular design.

Issues:

- Computational efficiency vs. its intended use, and
- practical aquisition of model parameter and validation data.

The main objective is to get a model that serves its intended purpose, and *to avoid Pyrrhic victories* in modelling work.

DICO – a multiphysics vowel synthesiser



Design and purpose of DICO

Specifications:

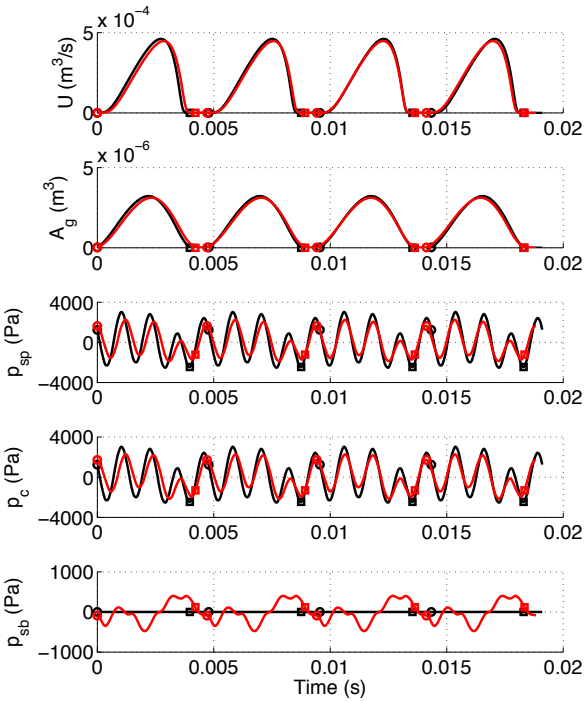
- Low order modelling of flow and vocal folds dynamics.
- Optimal tuning of model parameters in the flow and vocal folds models, based on measured signals.
- High resolution articulation geometries by MRI from a large number of test subjects and patients.
- Accurate modelling of vocal tract and environment acoustics by linear PDE's.

Applications:

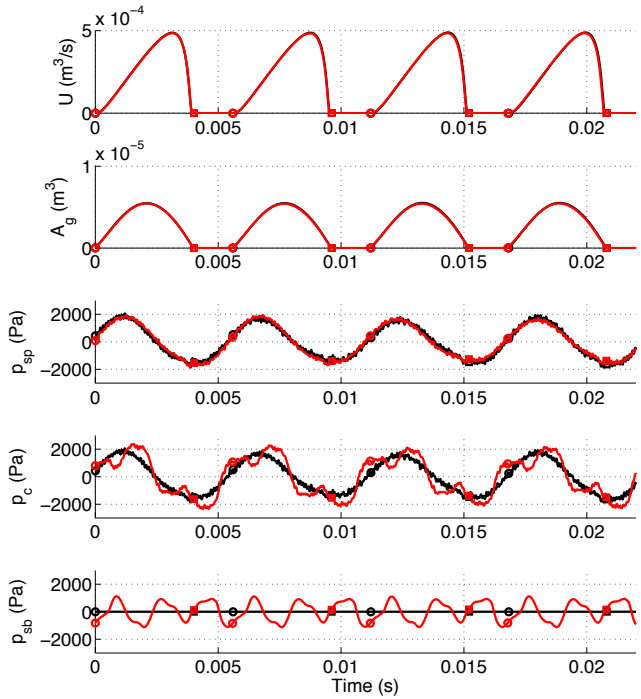
- Understanding the speech acoustic outcome of oral and maxillofacial surgery.
- As a development platform for speech processing algorithms such as Glottal Inverse Filtering (GIF).

Examples of simulated glottal signals

After parameter tuning, DICO is able to produce glottal pulse that matches experiments quite well in all vowel configurations.



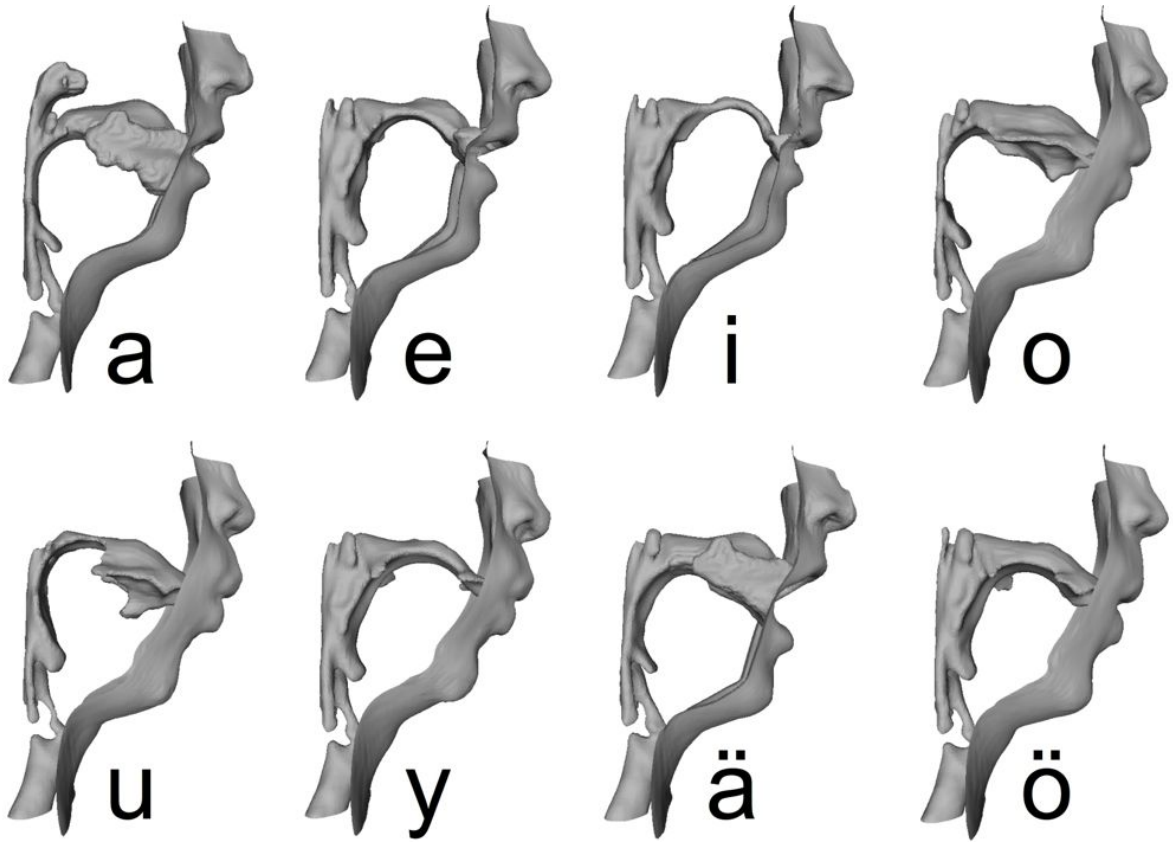
The glottal flow, opening area, and pressure signals during phonation of [ɑ] with the lowest VT resonance at $\approx 1\mathrm{kHz}$.



The same signals for [i] where the lowest resonance is at $\approx 250\mathrm{Hz}$. Note the difference in VT "ringing".

Vocal tract acoustics

Computational geometries



- PDE's of acoustics should be solved in these kinds of domains.
- The exterior space acoustics must be modelled as well.

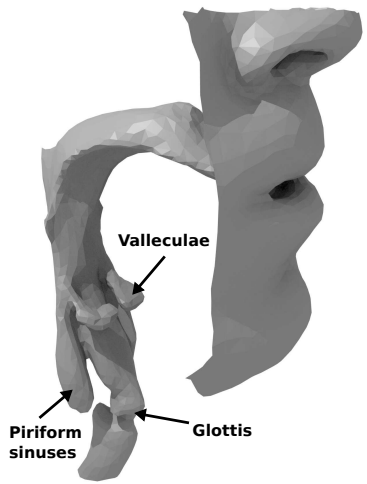
Wave equation model ("Dirichlet mouth")

Equations for the velocity potential $\phi = \phi(\mathbf{r}, t)$:

$$\begin{cases} \phi_{tt} = c^2 \Delta \phi & \text{in VT volume } \Omega \\ \phi(\mathbf{r}, t) = 0 & \text{at mouth opening } \Gamma(\ell) \\ \frac{\partial \phi}{\partial \nu}(\mathbf{r}, t) + \alpha \phi_t(\mathbf{r}, t) = 0 & \text{on VT walls } \Gamma \\ c \frac{\partial \phi}{\partial \nu}(\mathbf{r}, t) + \phi_t(\mathbf{r}, t) = 2 \sqrt{\frac{c}{\rho A(0)}}\, u(\mathbf{r}, t) & \text{at vocal folds } \Gamma(0). \end{cases}$$

This is a passive boundary node (with output omitted).

- c speed of sound
- ρ density of air
- α boundary dissipation coefficient
- ν exterior normal
- $A(0)$ area of $\Gamma(0)$



Cheaper model for tubular domains?

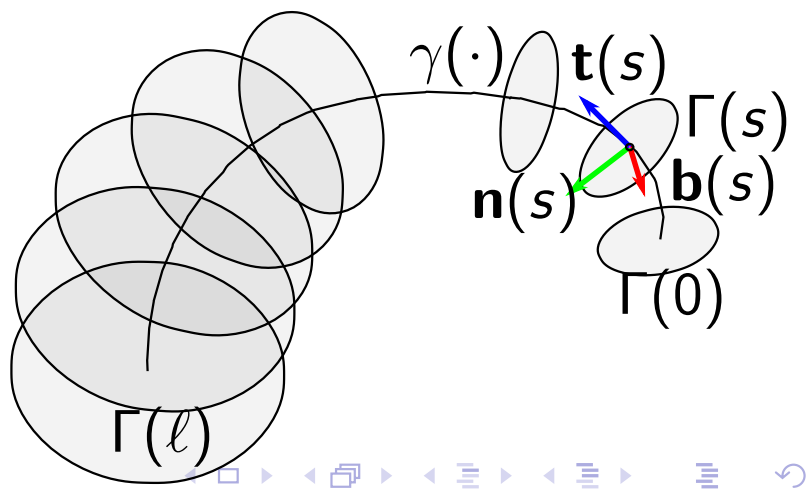
Let $\Omega \subset \mathbb{R}^3$ be a variable diameter, curved tube. Now, is there an approximate equation for the averages

$$\bar{\phi}(s,t) := \frac{1}{A(s)} \int_{\Gamma(s)} \phi \, dA \quad \text{for} \quad s \in [0, \ell]$$

of the velocity potential ϕ given by the wave equation on Ω?

YES, the generalised Webster's horn model for longitudinal dynamics!

- $\gamma(\cdot)$ centreline of Ω
- ℓ length of $\gamma(\cdot)$, i.e., Ω
- $\Gamma(s)$ slice of Ω, normal to $\gamma(\cdot)$ at s
- $A(s)$ area of $\Gamma(s)$



Webster's lossy resonator

Equations for the Webster's velocity potential $\psi = \psi(s,t)$:

$$\begin{cases} \psi_{tt} = \frac{c(s)^2}{A(s)} \frac{\partial}{\partial s}\left(A(s)\frac{\partial \psi}{\partial s}\right) - \frac{2\pi\alpha W(s) c(s)^2}{A(s)} \frac{\partial \psi}{\partial t} & \text{in vocal tract } s \in [0,\ell] \\ \psi(\ell,t) = 0 & \text{at mouth } s = \ell \\ -c\psi_s(0,t) + \psi_t(0,t) = 2\sqrt{\frac{c}{\rho A(0)}}\,\tilde{u}(t) & \text{on vocal folds } s = 0. \end{cases}$$

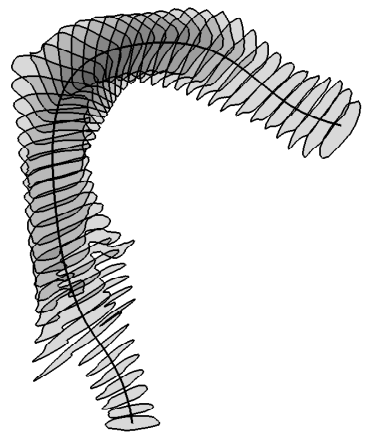
This is a passive **strong** boundary node (with output omitted).

c, ρ, α, ℓ as above

$A(s)$ area at $s \in [0,\ell]$

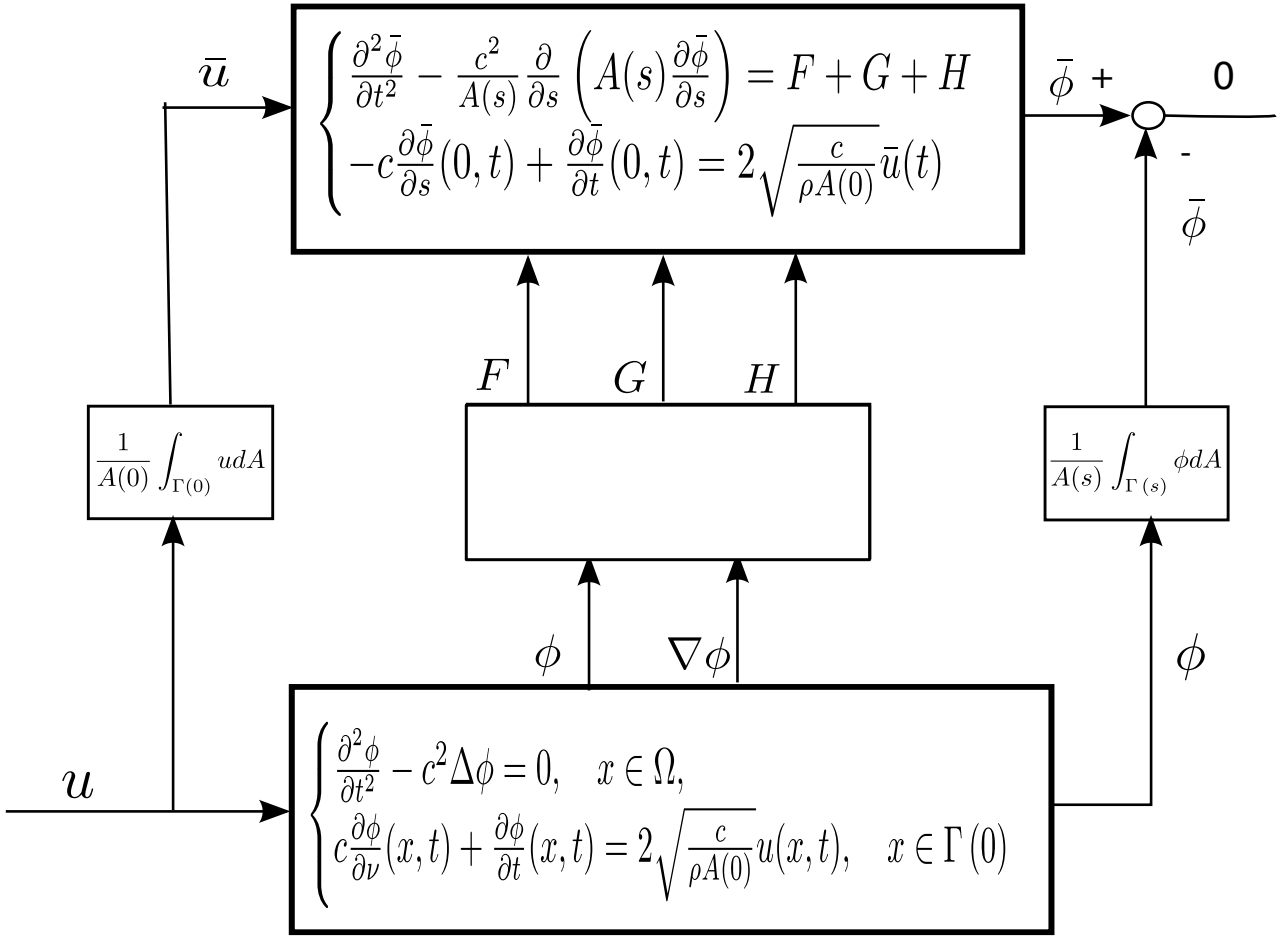
$\Sigma(s)$ curvature correction with $c(s) = c\Sigma(s)$

$W(s)$ stretching correction



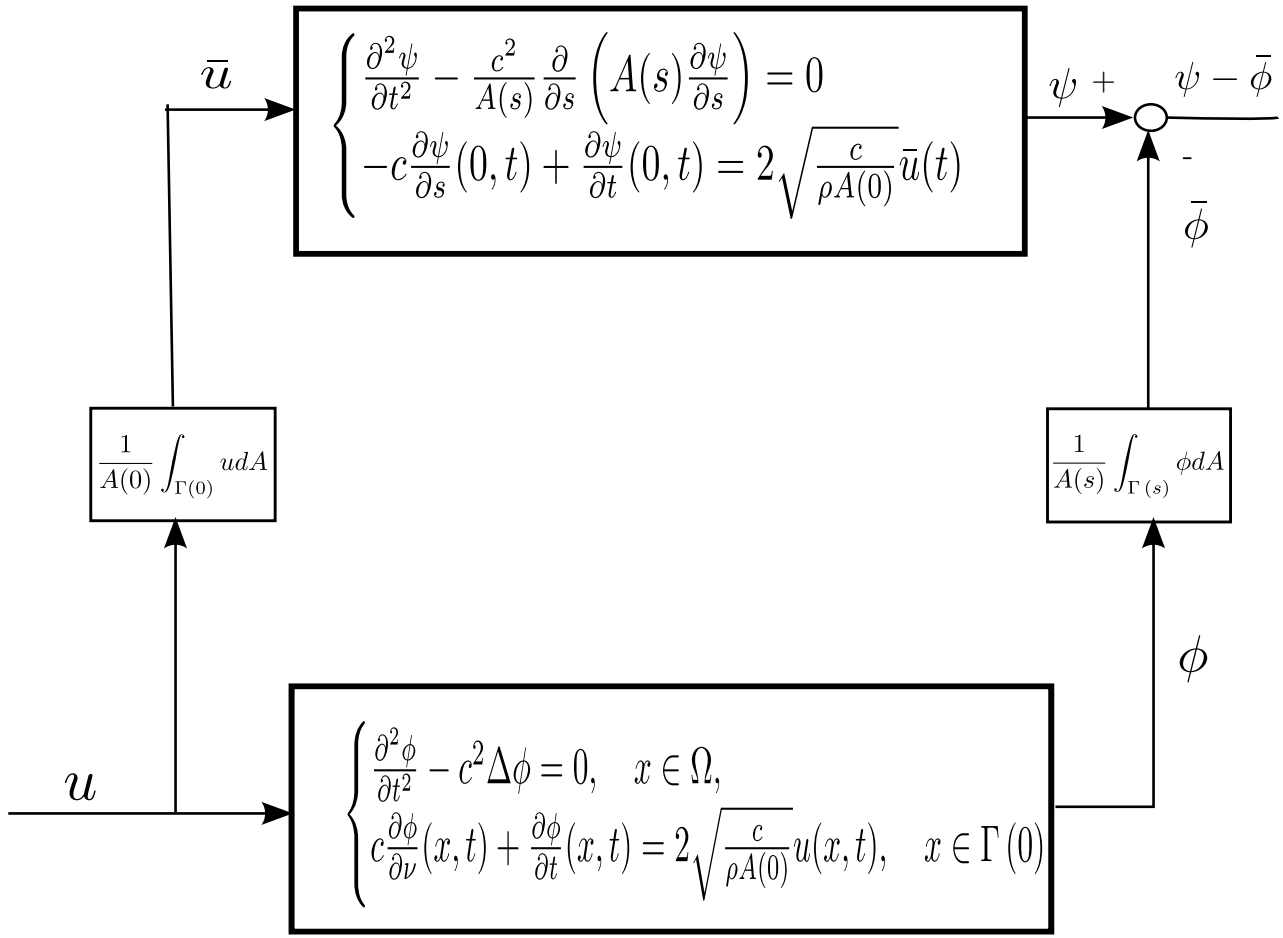
From now on, we restrict ourselves to the conservative case $\alpha = 0$.

Approximation by Webster's model? (1)



Don't worry about the formulas for functions F, G, H.

Approximation by Webster's model? (2)



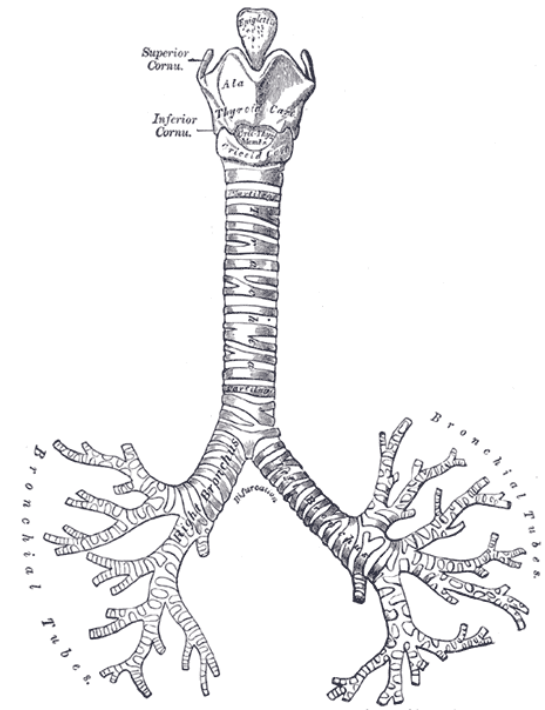
To make a long story short: $F + G + H \to 0$ as $\phi - \bar{\phi} \to 0$, giving an *a posteriori* estimate for the approximation error $\psi - \bar{\phi}$.

Subglottal acoustics

Treatment of the subglottal acoustics using Webster's model on subdividing bronchi, bronchioles, and alveoli?

Yes, it is possible.

Any finite number of passive strong boundary nodes can be coupled to a *transmission graph* that is passive and internally well-posed.



We currently use Webster's model for exponential horn for SGT, and tune its lowest resonance to the experimental value of 500 Hz.

Resonance equations

Ceteris paribus, the measured resonance structure (i.e., the formants) from vowel sounds should match the computed resonances from the model.

Wave Equation $\rightarrow$ Helmholtz equation:

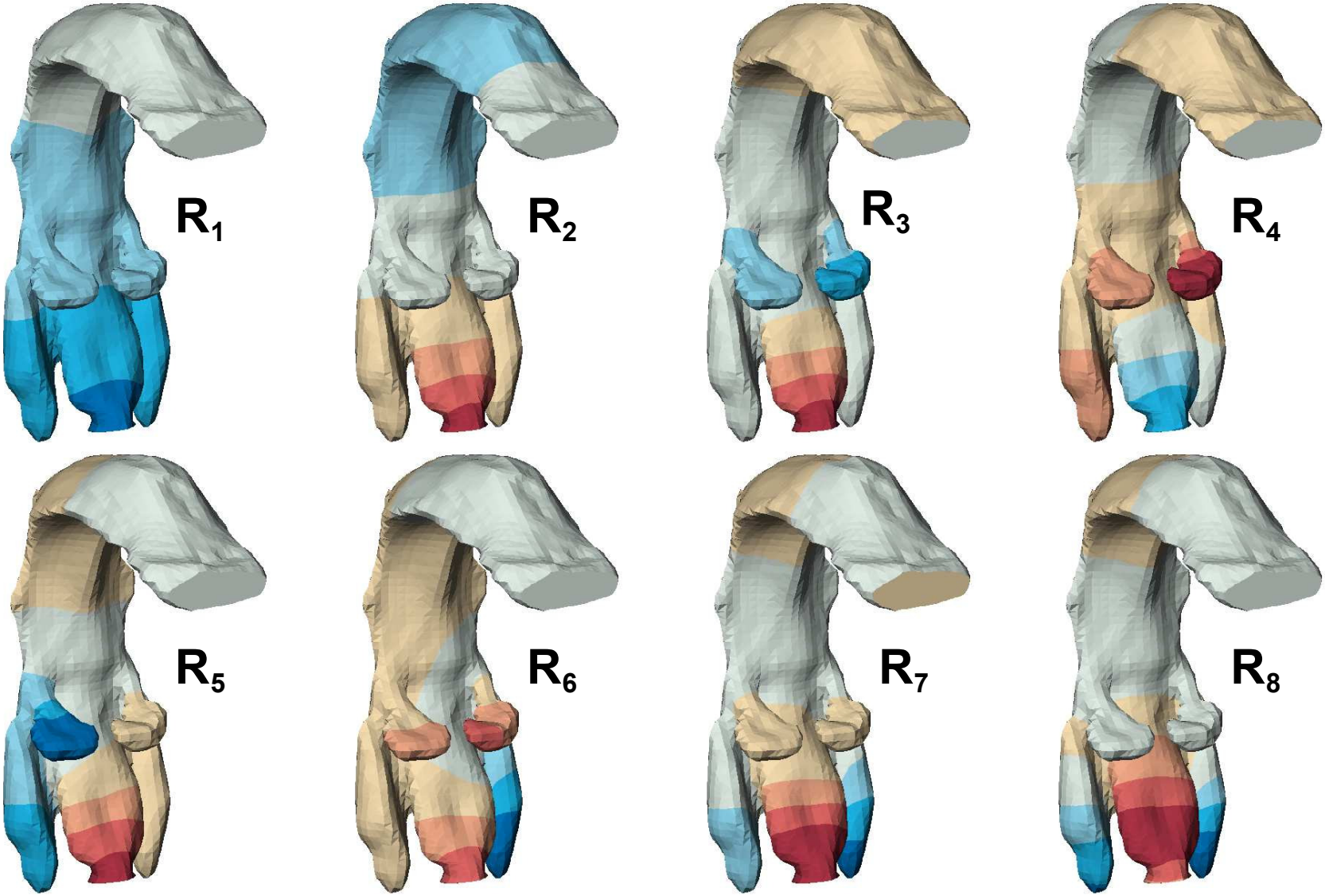
$$\lambda^2 \Phi_\lambda = c^2 \Delta \Phi_\lambda \quad \text{in VT volume } \Omega.$$

Webster's Equation $\rightarrow$ time-independent Webster:

$$\lambda^2 \psi_\lambda = \frac{c(s)^2}{A(s)} \frac{\partial}{\partial s} \left(A(s) \frac{\partial \psi_\lambda}{\partial s} \right) \qquad \text{for} \qquad s \in [0, \ell].$$

- The boundary conditions for the time-variant PDE give the corresponding boundary conditions of the resonance PDE.
- Discrete resonance frequencies: $R = \frac{1}{2\pi} Im(\lambda)$.

Helmholtz mode shapes Φ_λ for [œ]

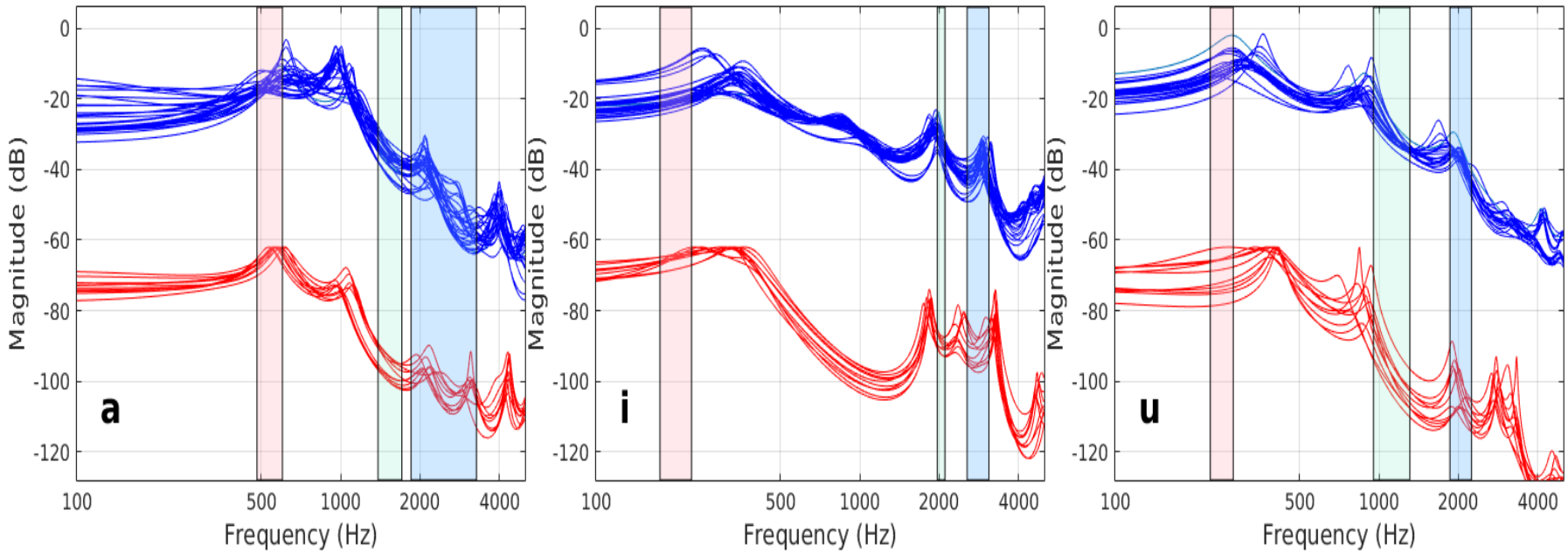


It seems a general fact that first three are purely longitudinal.

Environment acoustics

Matching measurements and computations (1)

Resonances of the quantal vowel geometries [ɑ, i, u]:

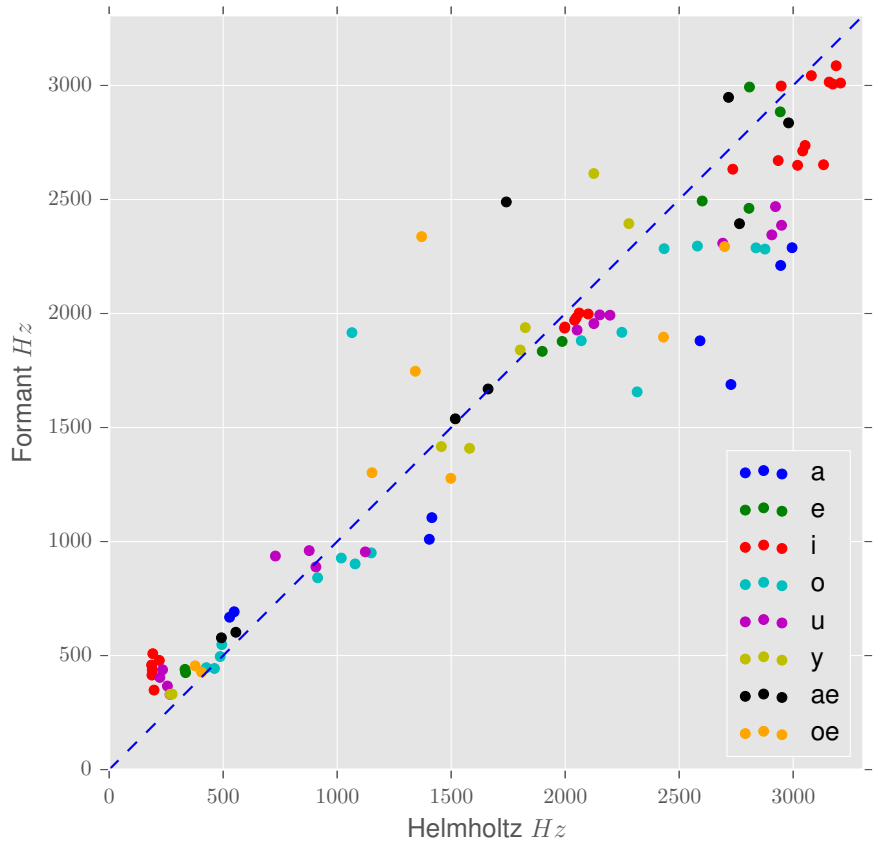


- Spectral envelopes. The red curves processed from recordings during the MRI, the blue in the anechoic chamber.
- Vertical bars contain the corresponding three lowest Helmholtz resonances computed using "Dirichlet mouth".

Matching measurements and computations (2)

Measured and computed resonances differ consistently as a function of the frequency.

The "hot frequencies" at $1\,\mathrm{kHz}$ and $2\,\mathrm{kHz}$ are due to the acoustic resonances in speech samples not present in Helmholtz spectrum of the vocal tract.



Conclusion: The "Dirichlet mouth" is far too simple model of the environment acoustics for the Helmholtz equation.

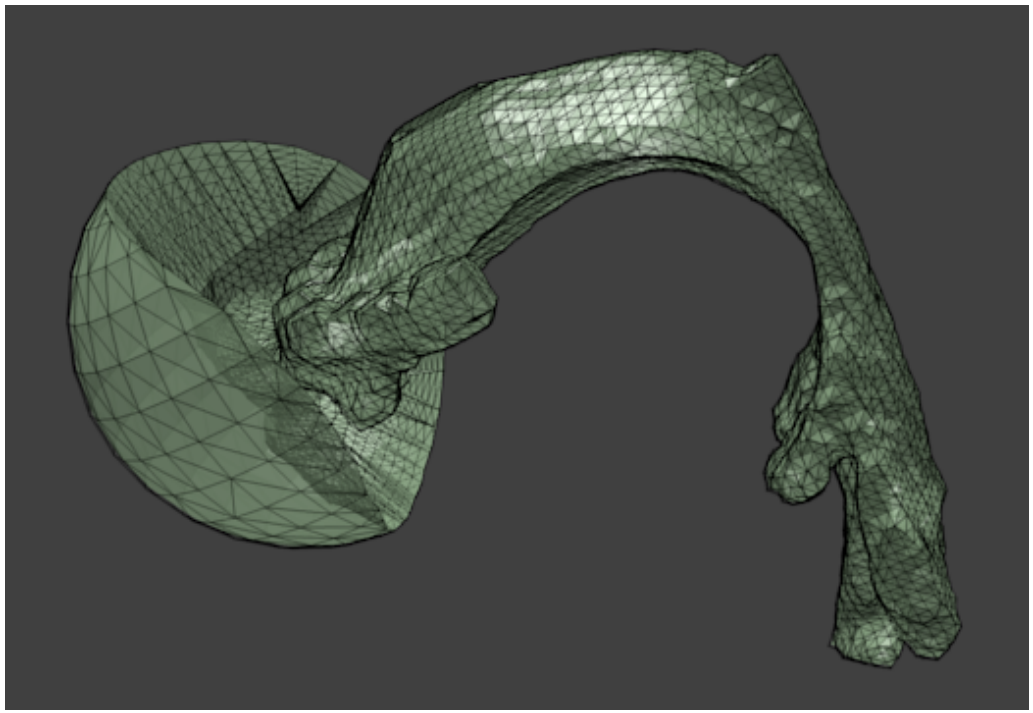
Modelling the exterior acoustics (1)

Requirements:

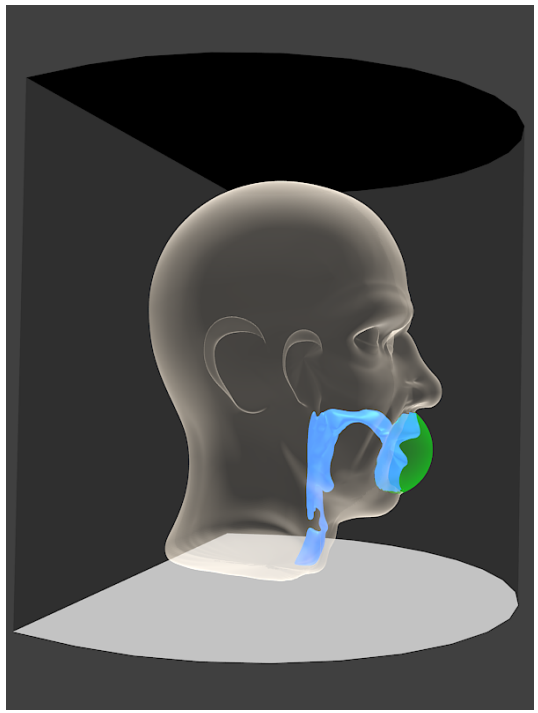
- The model must accommodate speech in open space as well as speech inside the MRI machine coils.
- The same exterior space geometries must be adaptable to vocal tracts from different test subject in different configurations $\rightarrow$ interface problems.
- Manual "fitting and stitching" is to be avoided since there are thousands of vocal tract geometries in the dataset.
- The computational burden of the exterior space should not exceed that of the vocal tract $\rightarrow$ model order reductions.

Neglecting the exterior acoustics leads to a frequency-dependent discrepancy of ≈ 2.5 semi-tones between (i) VT formant measurements from speech and (ii) Helmholtz resonances computed from VT geometries.

Exterior acoustics (2)

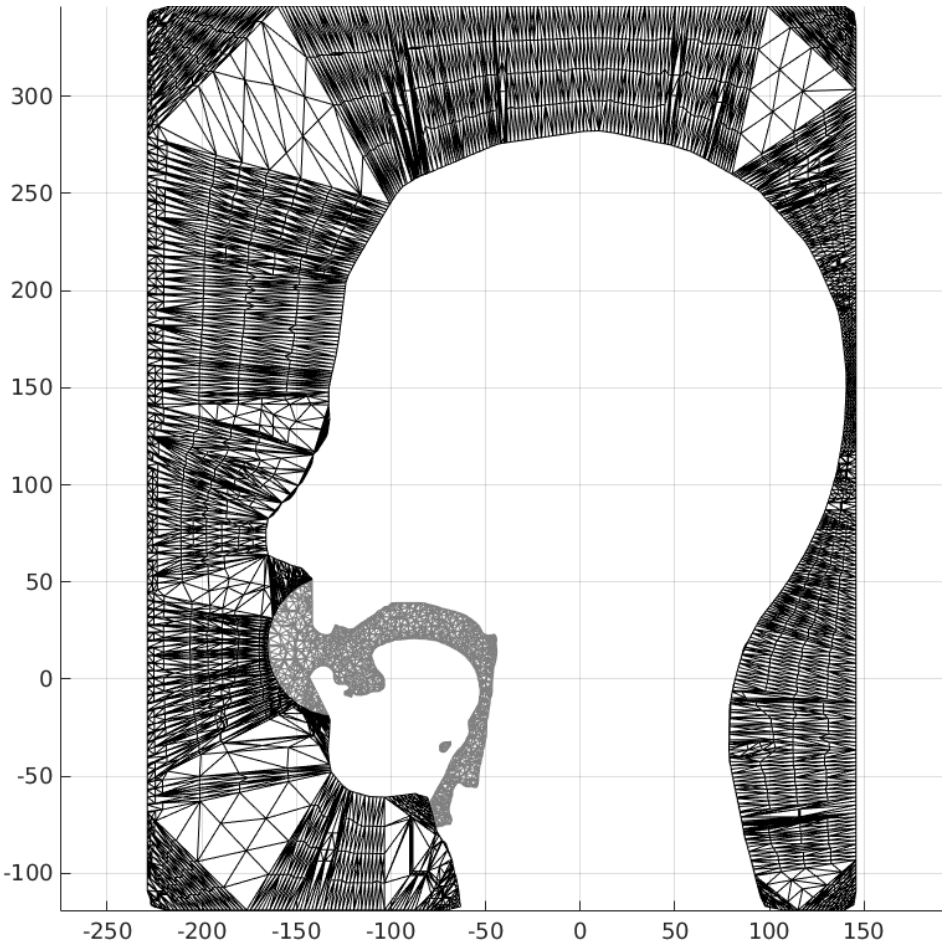


The triangulated interface manifold can be automatically attached at the mouth of any vocal tract geometry from the MRI data.



The standardised head model and tubular walls describing the interior surfaces of the MRI machine.

Exterior acoustics (3)



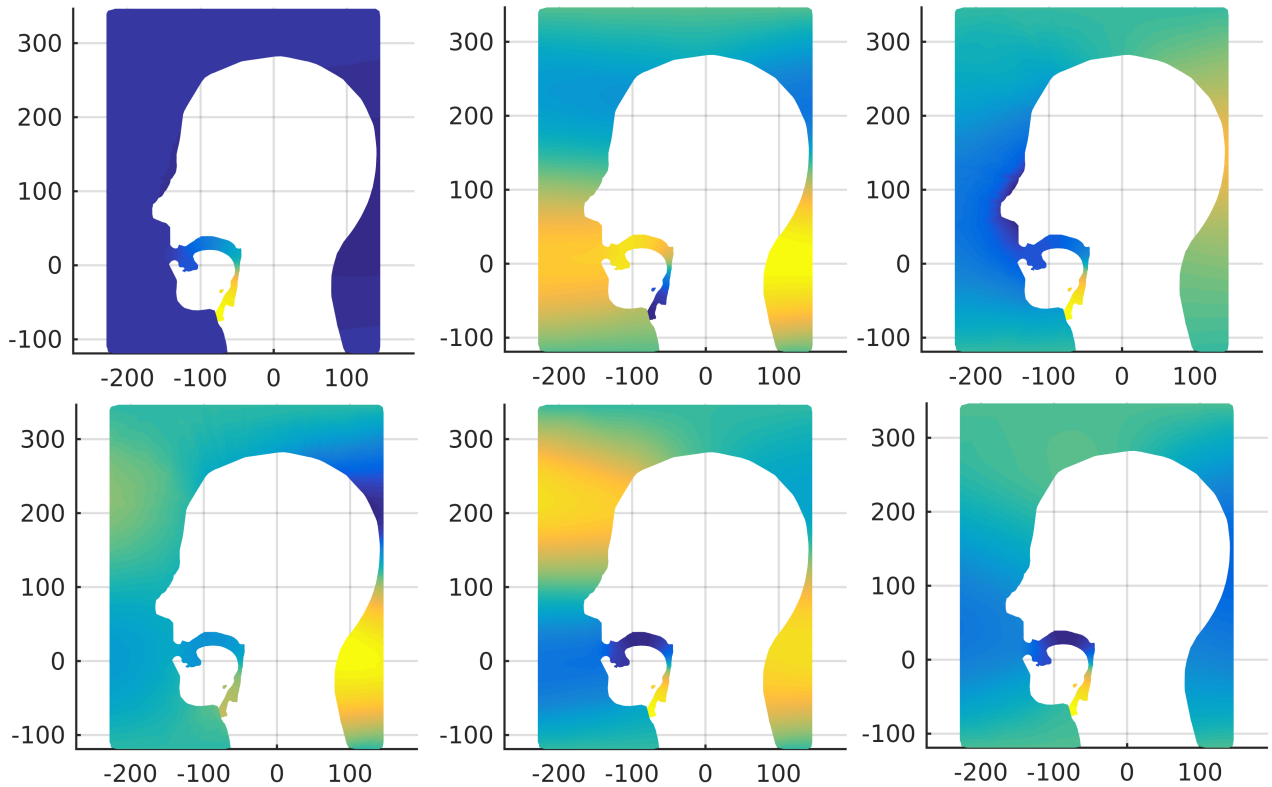
The mid-sagittal section of the triangulated acoustic space.

Note that the interior and exterior FEM meshes are not compatible at the interface manifold as would be required by a naive FEM solver.

Hence, the interior and exterior Helmholtz problems must be coupled at the interface using Nitsche's method (i.e., a type of interpolation across the mesh boundaries).

Exterior acoustics (4)

Helmholtz resonance pressure modal patterns of vowel utterance in constrained space. The tubular environment walls are hard reflectors, and the ends above and below are open.



TODO: 3D wave equation model that is cheap and accurate?

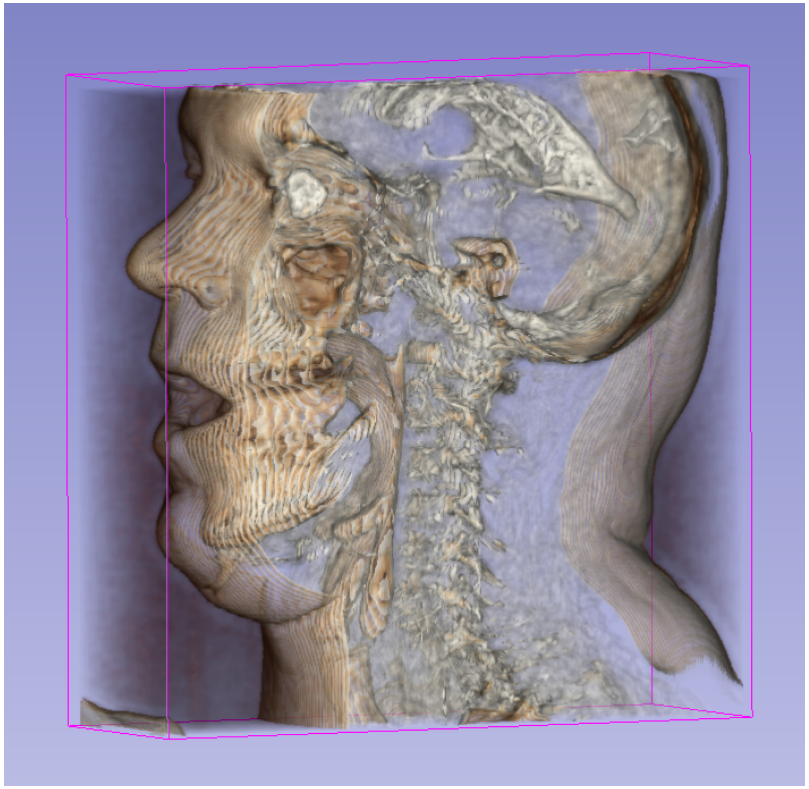
Data aquisition and processing

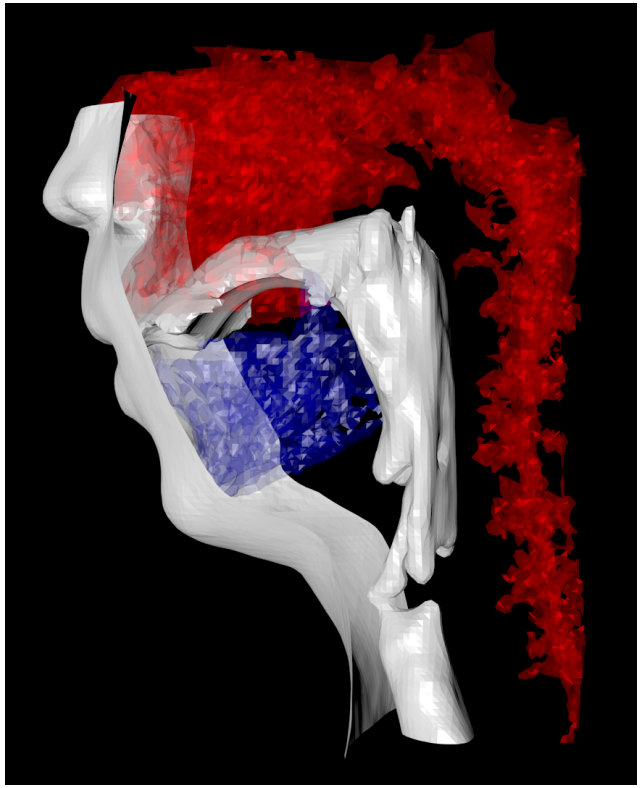
Data acquisition (1)

- Vocal tract: Simultaneous speech recording during 3D magnetic resonance imaging.
- Vocal folds: High-speed cinematography at $1 - 4\,\mathrm{kHz}$ simultaneously with electroglottogram (EGG) and speech recordings.

Data acquisition (2)

Computational geometries are processed from MRI data by custom software with minimum manual intervention.





Conclusions

- Mathematics is difficult.
- Applications require a lot of hard work.
- Applied mathematics is difficult and requires a lot of hard work.

Extra profit: The modelling work produces excellent research problems for applied mathematics.

“Opera magna”

A. Hannukainen, T. Lukkari, J. Malinen, and P. Palo.

Vowel formants from the wave equation. *J Acoust Soc Am*, 122(1):EL1–EL7, 2007.

D. Aalto, O. Aaltonen, R.-P. Happonen, J. Malinen, P. Palo, R. Parkkola, J. Saunavaara, M. Vainio,

Recording speech sound and articulation in MRI, *Proceedings of BIODEVICES 2011*, 168–173, 2011.

A. Aalto and J. Malinen.

Composition of passive boundary control systems. *Math Control Relat F*, 3(1):1–19, 2013.

T. Lukkari and J. Malinen.

Webster’s equation with curvature and dissipation. *arXiv:1204.4075 (submitted), 2013.*

D. Aalto, O. Aaltonen, R.-P. Happonen, P. Jääsaari, A. Kivelä, J. Kuortti, J. M. Luukinen, J. Malinen, T. Murtola, R. Parkkola, J. Saunavaara, and M. Vainio.

Large scale data acquisition of simultaneous MRI and speech. *Appl Acoust*, 83(1):64–75, 2014.

A. Aalto, T. Lukkari, and J. Malinen.

Acoustic wave guides as infinite-dimensional dynamical systems. *ESAIM Contr Optim Ca*, 21(2): 324-247, 2015.

T. Lukkari and J. Malinen.

A posteriori error estimates for Webster’s equation in wave propagation. *J Math Anal Appl*, 427(2):941–961, 2015.

A. Aalto, T. Murtola, J. Malinen, D. Aalto, and M. Vainio.

Modal locking between vocal fold and vocal tract oscillations: Simulations in time domain. *arXiv:1506.01395 (submitted), 2015.*

Theses

A. Kivelä.

Acoustics of the vocal tract: MR image segmentation for modelling.
Master's thesis, 2015.

A. Aalto.

Infinite Dimensional Systems: Passivity and Kalman Filter Discretization.
Doctoral dissertation, 2014.

T. Murtola.

Modelling vowel production.
Licentiate thesis, 2014.

P. Palo.

A wave equation model for vowels: Measurements for validation.
Licentiate thesis, 2011.

A. Aalto.

A low-order glottis model with nonturbulent flow and mechanically coupled acoustic load.
Master's thesis, 2009.

Thanks for your patience.
Any questions?

http://speech.math.aalto.fi
https://www.youtube.com/channel/UCDRLICfptS1TQNLkzFjC94g