System Theory of Speech

Antti Hannukainen, **Jarmo Malinen**, Antti Ojalammi

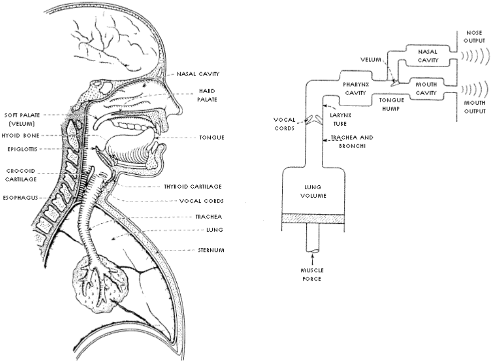
Aalto University, School of Science,
Department of Mathematics and Systems Analysis

Lorentz Center
May 23-26, Leiden

Human voice production

Simplified vowel production:

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



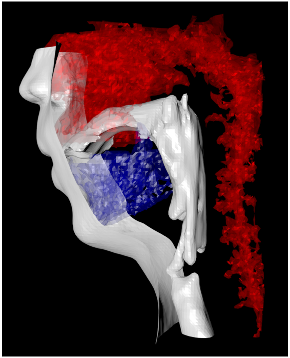
Flanagan, J. L. (1972). Speech Analysis Synthesis and Perception, Springer-Verlag.

Notable facts:

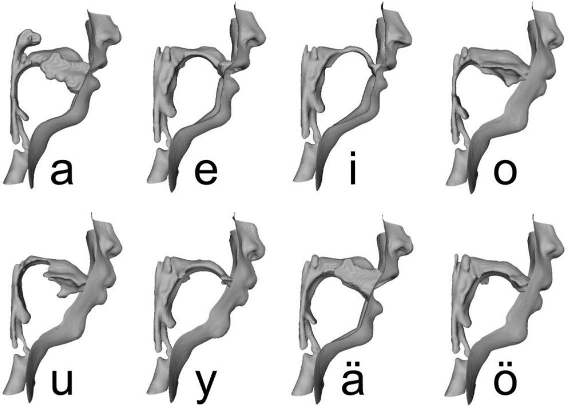
- Vocal tract (VT) shape changes, and there are feedbacks.
- Not all speech sounds originate in vocal folds.

Modelling speech requires data

- Simultaneous speech recording during 3D MR imaging.
- Geometry for the computational model is constructed from MR images by custom software.

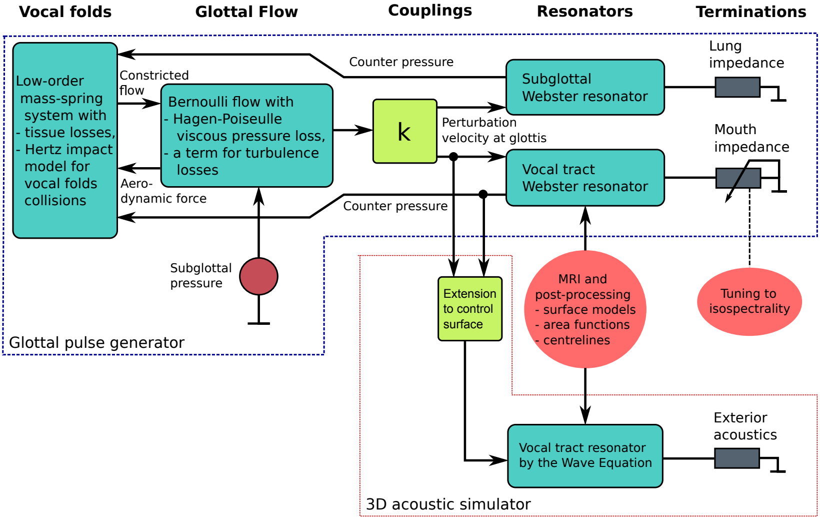


Geometries of Finnish vowels



PDE's of acoustics should be solved in these domains.

Multiphysics of vowel production, Dico



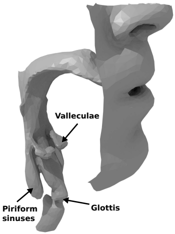
Wave equation model ("Dirichlet mouth")

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?

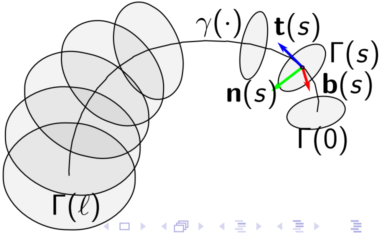
Let $\Omega \subset \mathbb{R}^3$ be a variable diameter, curved tube. Now, is there an approximate equation for the averages

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

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

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

- ℓ length of Ω
- $\gamma(\cdot)$ centreline of 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).

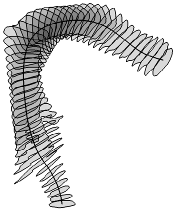
c, ρ, α as above

ℓ length of the VT

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

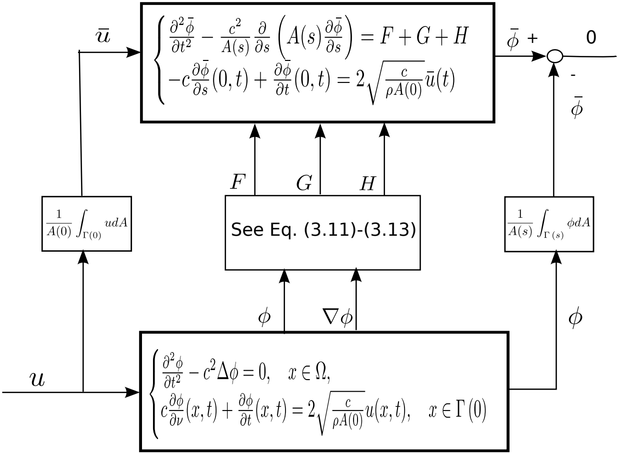
$\Sigma(s)$ curvature correction

$W(s)$ stretching correction



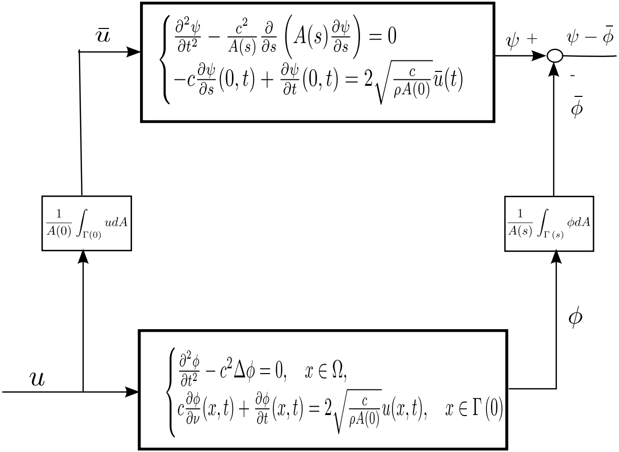
From now on, we restrict ourselves to the conservative case $\alpha = 0$.

Approximation by Webster's model? (1)



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

Approximation by Webster's model? (2)

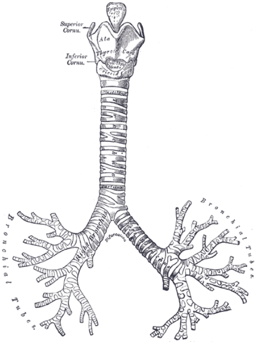


To make a long story short: $F + G + H \to 0$ as $\phi - \bar{\phi} \to 0$, giving an *a posteriori* estimate for the approximation error $\psi - \bar{\phi}$.

Transmission graphs

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

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



We just use Webster's model for exponential horn in "Dico".

Resonance equations

Ceteris paribus, the measured resonance structure from vowel sounds should match the computed resonances from the model.

Wave Equation $\rightarrow$ Helmholtz equation:

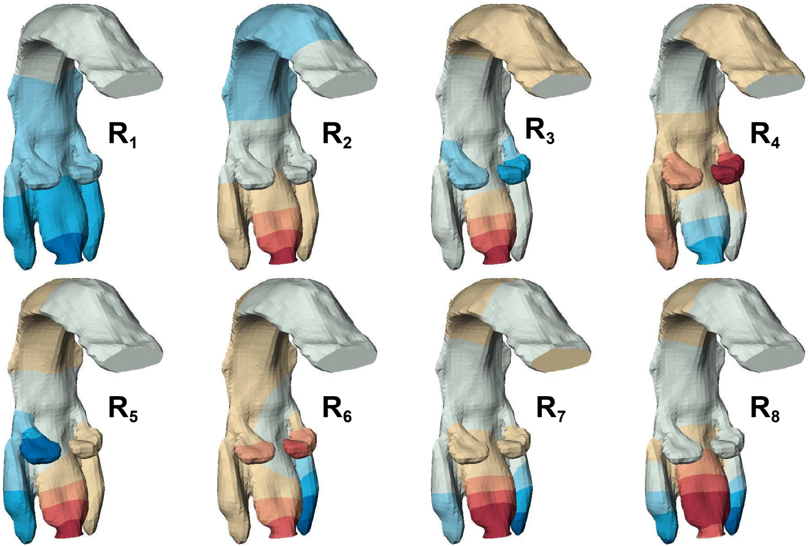
$$\lambda^2 \Phi_\lambda = c^2 \Delta \Phi_\lambda \quad \text{in VT volume } \Omega.$$

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

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

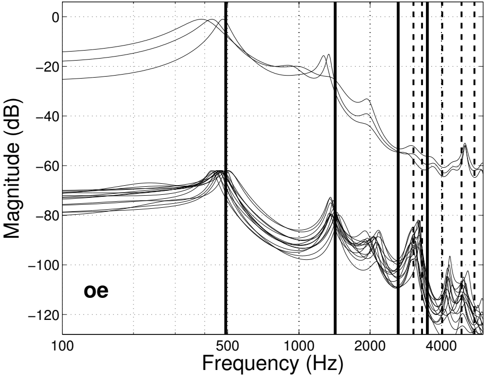
- The boundary conditions for the time-variant PDE give the corresponding boundary conditions of the resonance PDE.
- Discrete resonance frequencies: $R = \frac{1}{2\pi} Im(\lambda)$.

Helmholtz mode shapes Φ_λ for [oe]



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

Matching measurements and computations

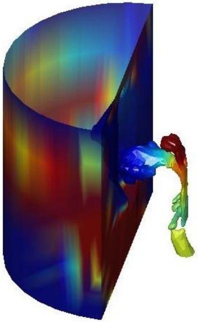


- Vertical lines: Helmholtz resonances with “Dirichlet mouth”.
- Curve families: Spectral envelopes from recorded speech. The upper during MRI, the lower in anechoic chamber.

Exterior acoustics (1)

Until now, the exterior space acoustic have been omitted, and the Dirichlet boundary condition at mouth has been used instead.

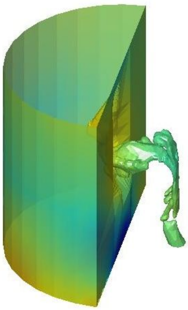
The mixed resonance of a nasal [ɑ] at 1625Hz. Both the vocal tract and the idealised, semi-cylindrical exterior domain ($d = 30$ cm) are excited.



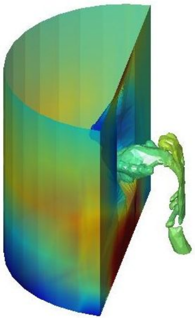
Neglecting the exterior acoustics leads to a frequency-dependent discrepancy of ≈ 2.5 semi-tones between VT resonance measurements from speech and Helmholtz computations.

Exterior acoustics (2)

For speech, we need "High Fidelity" in the vocal tract volume but in the exterior acoustic space, "Low Adultery" will suffice.



The mixed resonance of [ɑ] is found at 1625Hz when using 8900 D.o.F. for the exterior domain.



The mixed resonance of [ɑ] is found at 1637Hz when using 26 D.o.F. for the exterior domain.

Exterior acoustics (3)

Typical numbers of exterior space reduction for the Helmholtz problem:

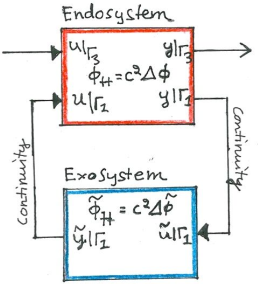
	# of tetr. F.E.	D.o.F	Reduced D.o.F.
Vocal tract	115000	26600	26600
Exterior space	38500	8900	26

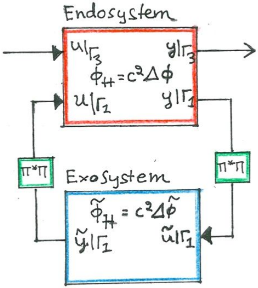
The dimension reduction $8900 \rightarrow 26$ in degrees-of-freedom of the exterior acoustics produces an error of ≈ 0.8 semi-tones in the three lowest pure resonances R_1, R_2, and R_3 of the vocal tract.

"Pure" vocal tract resonance means that the exterior acoustic space is not significantly excited.

Partial dimension reduction (1)

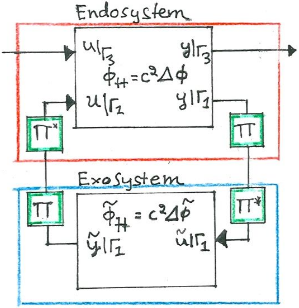
Let us start with a dissipative BCS that is first splitted spatially into two subdomains: interior and exterior.

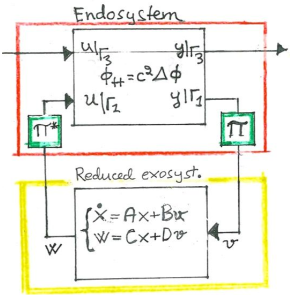




- $\Pi : L^2(\Gamma_1) \to \mathbb{C}^n$ is a finite rank co-isometry. For example, it may map to averages on disjoint parts of the interface Γ_1.
- The orthogonal projection $\Pi^*\Pi$ removes energy from the feedback loop, thus preserving passivity.

Partial dimension reduction (2)





- The new endosystem has finite-dimensional internal input and output spaces.
- Finally, the exosystem is replaced by a finite-dimensional approximate system.

Conclusions

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

“Opera magna”

A. Hannukainen, T. Lukkari, J. Malinen, and P. Palo.
Vowel formants from the wave equation. *Journal of the Acoustical Society of America*, 122(1):EL1–EL7, 2007.

A. Aalto, D. Aalto, J. Malinen, and M. Vainio.
Modal locking between vocal fold and vocal tract oscillations. *arXiv:1211.4788 (submitted), 2013.*

A. Aalto and J. Malinen.
Composition of passive boundary control systems. *Mathematical Control and Related Fields*, 3(1):1–19, 2013.

T. Lukkari and J. Malinen.
Webster’s equation with curvature and dissipation. *arXiv:1204.4075 (submitted), 2013.*

A. Aalto, T. Lukkari, and J. Malinen.
Acoustic wave guides as infinite-dimensional dynamical systems. *ESAIM: Control, Optimisation and Calculus of Variations (to appear)*, 2014.

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. *Applied Acoustics (to appear)*, 2014.

Thanks for your patience.
Any questions?

http://speech.math.aalto.fi