Resonances and mode shapes of the human vocal tract during vowel production

Atle Kivelä, Juha Kuortti, Jarmo Malinen

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

NSCM 26
October 23-26, Oslo, Norway

Outline

Human voice production

Vowel production:

Mouth

↑

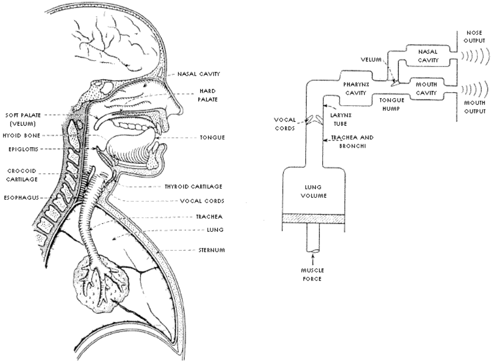
Vocal tract

(filter)

↑

Vocal folds

(source)



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

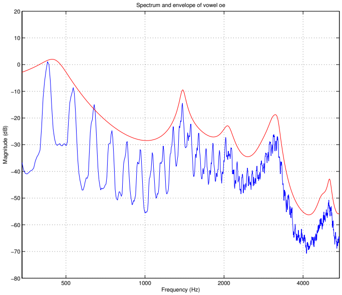
Other speech sounds:

- Vocal tract (VT) shape changes.
- Sound sources vary depending on the speech sound.

Formants

Definition

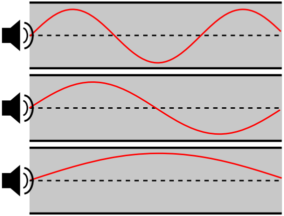
Formants are the envelope peaks of the sound spectrum.

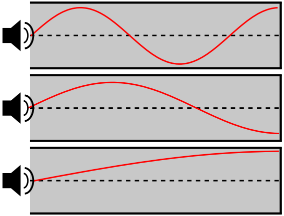


Resonances and formants

Definition

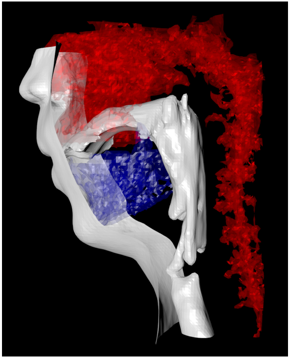
Resonances are computationally determined resonance frequencies for the VT.



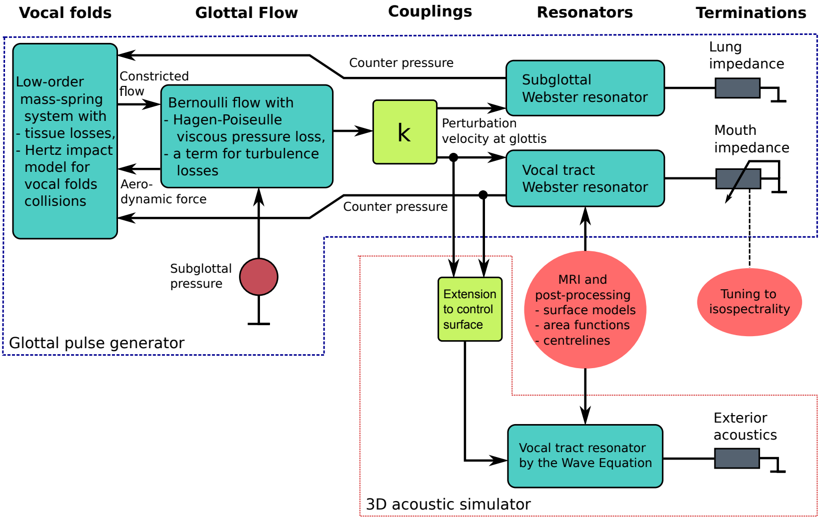


Data acquisition

- Simultaneous speech recording during 3D MR imaging.
- Formants are determined from the speech recordings after noise cancellation.
- Geometry for the computational model is constructed from MR images.



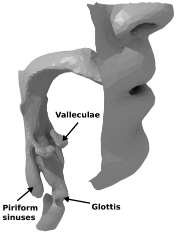
The models


Vocal folds
Glottal Flow
Couplings
Resonators
Terminations
Low-order mass-spring system with
- tissue losses,
- Hertz impact model for vocal folds collisions
Constricted flow
Bernoulli flow with
- Hagen-Poiseulle viscous pressure loss,
- a term for turbulence losses
Aero-dynamic force
Counter pressure
k
Subglottal Webster resonator
Lung impedance
Perturbation velocity at glottis
Vocal tract Webster resonator
Mouth impedance
Counter pressure
Subglottal pressure
Extension to control surface
MRI and post-processing
- surface models
- area functions
- centrelines
Tuning to isospectrality
Glottal pulse generator
Vocal tract resonator by the Wave Equation
Exterior acoustics
3D acoustic simulator

Helmholtz model

$$\begin{cases} \lambda^2 \Phi_\lambda = c^2 \Delta \Phi_\lambda & \text{in VT volume } \Omega \\ \Phi_\lambda = 0 & \text{on mouth opening } \Gamma_1 \\ \frac{\partial \Phi_\lambda}{\partial \nu} = 0 & \text{on VT walls } \Gamma_2 \\ \lambda \Phi_\lambda + c \frac{\partial \Phi_\lambda}{\partial \nu} = 0 & \text{on vocal folds } \Gamma_3, \end{cases} \quad (1)$$

- Φ_λ the velocity potential
- c the speed of sound
- ν exterior normal



Helmholtz model 2

The variational formulation of (1) is

$$\lambda^2 \int_\Omega p_\lambda \phi \, d\Omega + \lambda c \int_{\Gamma_3} p_\lambda \phi \, d\omega + c^2 \int_\Omega \nabla p_\lambda \cdot \nabla \phi \, d\Omega = 0. \qquad (2)$$

Using FEM with piecewise linear shape functions and a tetrahedral mesh with approximately 10^5 elements, we construct matrices **K**, **M** and **P**. Equation (2) can then be written as a quadratic eigenvalue problem

$$\lambda^2 \mathbf{K}\mathbf{x}(\lambda) + \lambda c \mathbf{P}\mathbf{x}(\lambda) + c^2 \mathbf{M}\mathbf{x}(\lambda) = 0.$$

This corresponds to the generalized eigenvalue problem

$$\begin{bmatrix} -c\mathbf{P} & -c^2\mathbf{M} \\ \mathbf{I} & 0 \end{bmatrix} \mathbf{y}(\lambda) = \lambda \begin{bmatrix} \mathbf{K} & 0 \\ 0 & \mathbf{I} \end{bmatrix} \mathbf{y}(\lambda), \quad \text{where} \quad \mathbf{y}(\lambda) = \begin{bmatrix} \lambda \mathbf{x}(\lambda) \\ \mathbf{x}(\lambda) \end{bmatrix}$$

Webster resonator

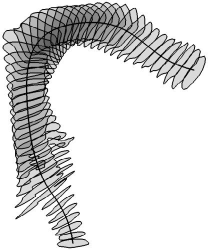
$$\lambda^2 \psi_\lambda = \frac{c^2 \Sigma(s)^2}{A(s)} \frac{\partial}{\partial s} \left(A(s) \frac{\partial \psi_\lambda}{\partial s} \right) \quad \text{for} \quad s \in [0, \ell], \qquad (3)$$

ψ_λ velocity potential

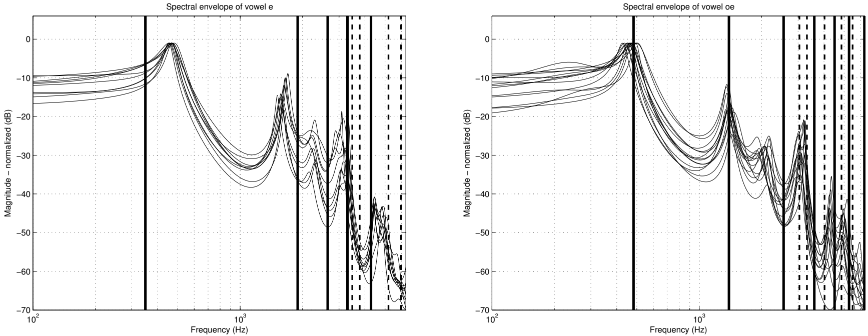
c speed of sound

$A(s)$ area at s

$\Sigma(s)$ curvature correction

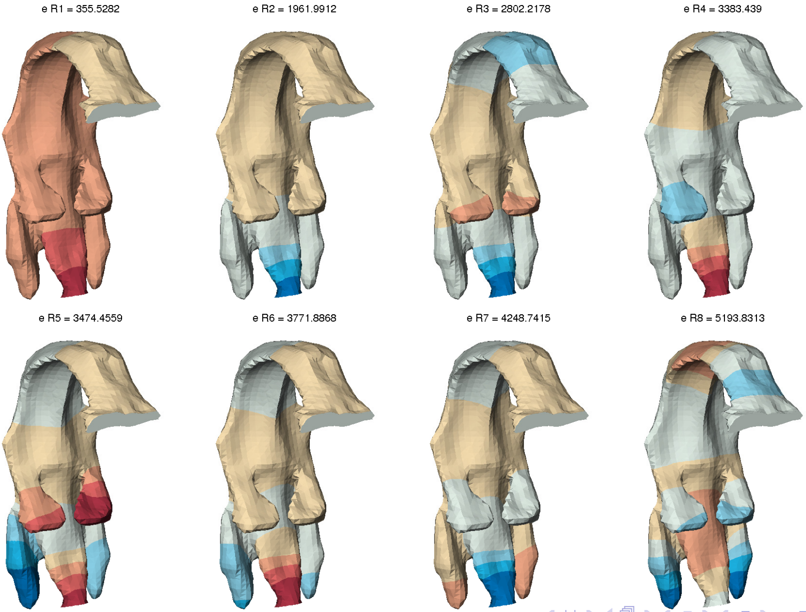


Resonances and formants for e and oe

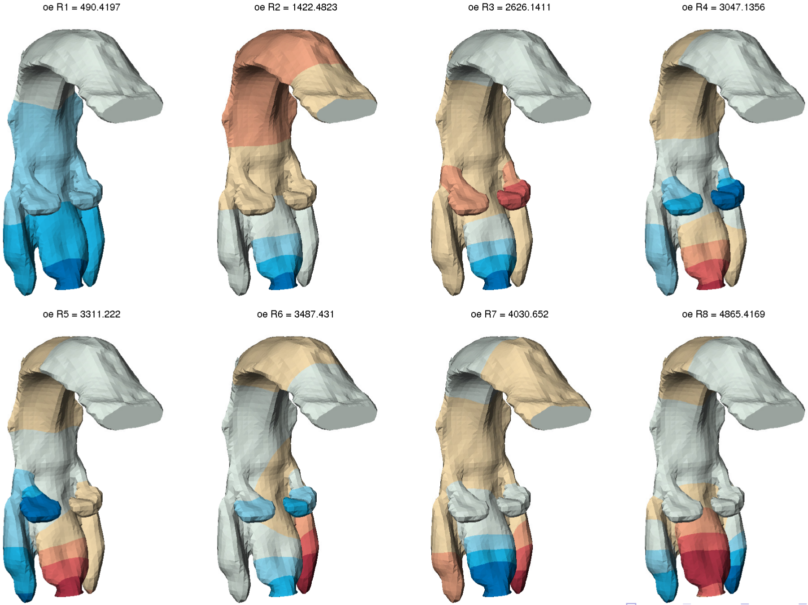


- As expected, different models somewhat corresponds.
- Helmholtz resonances form clusters.

Mode shapes for e



Mode shapes for oe



Classifying resonances

- Helmholtz resonances are complicated.
  - The resonance direction is not well defined.
  - There are Transversal components (the clusters).
- Webster model only capture longitudinal resonances.
  - The resonances "average" the clusters of the Helmholtz resonances.
  - There is no one-to-one correspondence.
- At least the longitudinal resonances show up as formants.
  - The LPC algorithm also does some "averaging".
  - There are other sources of error.
- Theres some discrepancy between formants and resonances .
  - The unrealistic mouth boundary condition is a likely cause.

Conclusions

- The VT can have significant transversal resonances.
  - Most articulatory models only consider longitudinal resonances.
  - Trained singers are known to be able to produce formants not visible in the spectra of untrained singers.
- A more realistic acoustic impedance at mouth should give better results.
  - Modelling of the exterior space is an ongoing project.
- Full classification of the resonance structure requires going through a larger set of data.
  - Dimensions of the vocal space.

The End

Questions?

http://speech.math.aalto.fi