Mathematical modeling of speech acoustics

D. Sc. Daniel Aalto

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Ultimate goal

Predict the **speech** outcome of oral and maxillofacial surgery patients

How to reach the goal?

- By synthesizing the speech of a **virtually** operated vocal tract

This talk

- Orthognathic patients and model validation
- Objective measurement of speech acoustics
- Presentation of the computational vowel model

Orthognathic patient data I

- Anatomic changes are predictable (advancement of mandible and/or maxilla)
- Comparison of measured acoustics and computed speech characteristics from the MR images
- Comparison of the prediction and the real speech outcome

Orthognathic patient data II

- 20 (10 women, 10 men) patients undergoing an orthognathic treatment in Turku are enrolled
- so far: 6 pre-treatment recordings
- 70 sustained sound productions (10 s) for each patient
- 12 short sentences
- Measurable changes in formants (Niemi et al. 2006)

Orthognathic patient data III

- Phonetically rich sentences
- Vowels occur in phonetically controlled contexts
- Coarticulation and motility in sagittal videos
- Changes in vowel space are predicted

What is speech for?

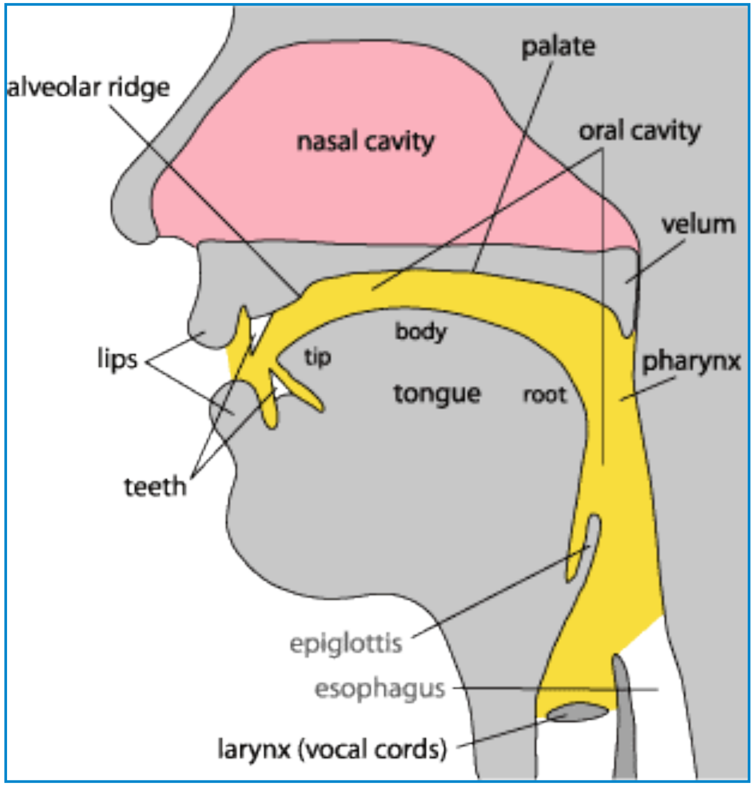
Linguistic code:

Communicating the “thoughts” of a person through language e.g. “Mom, do pharaohs exist?”

Biological code:

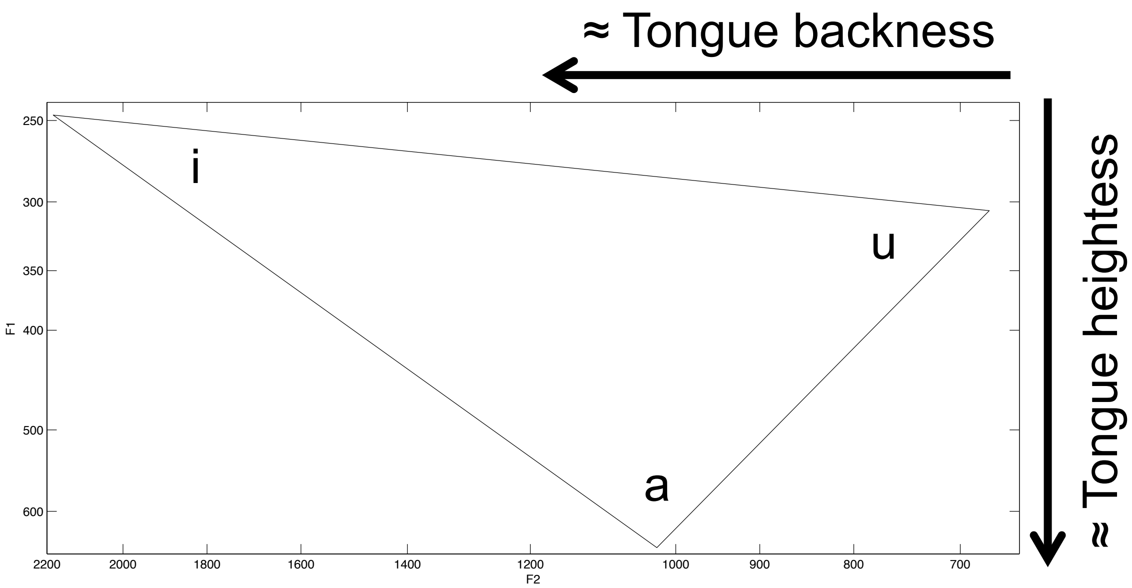
The speaker reveals information about him/herself e.g. the speaker sounds assertive, healthy, happy, tired, 5-6yo boy, speaker of Finnish (Turku dialect)

Vocal tract resonances



Thanks to:
A. Suni, T. Raitio, P. Alku

Vowel space: Formants



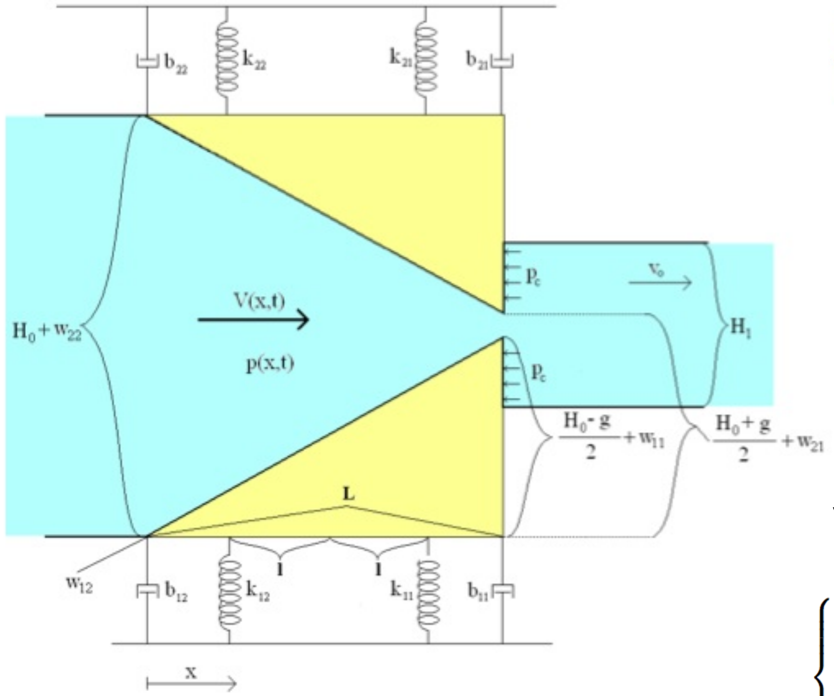
- Size of the vowel space correlates with intelligibility
- In addition, vowel formants provide information of the speaker

Dico: modeling speech acoustics

An integrated vowel acoustics model:

- 3D-image is transformed to a sound!
- A mechanical model for vocal folds
- An acoustic model for the air vibrations inside the vocal tract
- Interaction between vocal folds and vocal tract
- Model = simplified reality!

Dico: Submodel for vocal folds



Vocal fold movement:

$$\begin{cases} \mathbf{M}_1\ddot{\mathbf{W}}_1(t) + \mathbf{B}_1\dot{\mathbf{W}}_1(t) + P\mathbf{K}_1\mathbf{W}_1(t) = -\mathbf{F}(t), \\ \mathbf{M}_2\ddot{\mathbf{W}}_2(t) + \mathbf{B}_2\dot{\mathbf{W}}_2(t) + P\mathbf{K}_2\mathbf{W}_2(t) = \mathbf{F}(t) \end{cases}$$

Air flow through glottis:

$$p_{sub} = \underbrace{\widehat{C}_{iner}\dot{v}_o(t)}_{\text{inertive pressure}} + \underbrace{\frac{C_g}{\Delta W_1(t)^3} v_o(t)}_{\text{viscous pressure loss}}$$

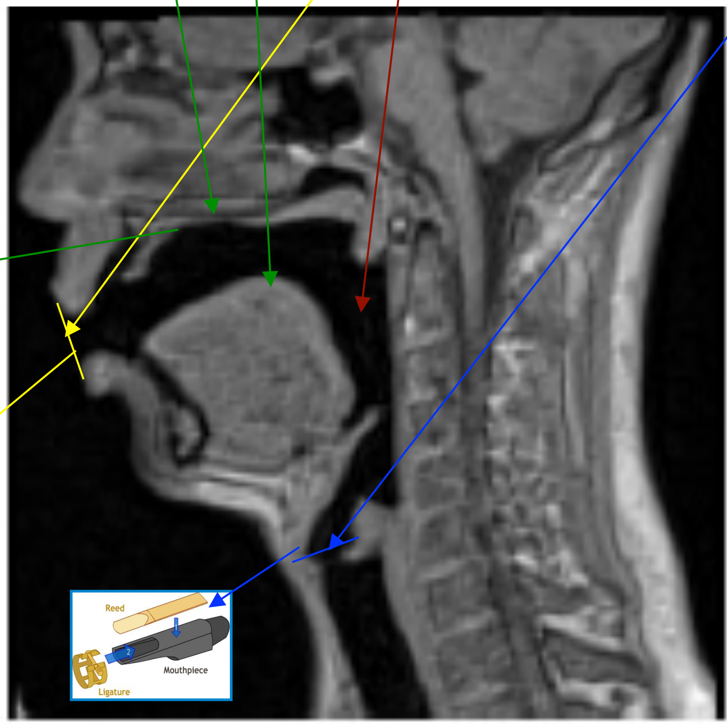
Vocal tract impact:

$$\begin{cases} \frac{\partial^2 \Psi}{\partial t^2}(x,t) = \frac{c^2}{A(x)}\frac{\partial}{\partial x}\left(A(x)\frac{\partial \Psi}{\partial x}(x,t)\right), \qquad x \in [0, L_{VT}], \\ \Psi_x(0,t) = -v_o(t), \\ \Psi_t(L_{VT},t) + \theta c \Psi_x(L_{VT},t) = 0, \\ p_c(t) = \rho \Psi_t(0,t). \end{cases}$$

Dico: Acoustic equations

$$\begin{cases} \Phi_{tt} = c^2 \Delta \Phi \text{ on } \Omega, \ \Phi = 0 \text{ on } \Gamma_1, \\ \frac{\partial \Phi}{\partial \nu} = 0 \text{ on } \Gamma_2, \ \Phi_t + c\frac{\partial \Phi}{\partial \nu} = 2\sqrt{\frac{c}{\rho_0}}\, u \text{ on } \Gamma_3, \end{cases}$$

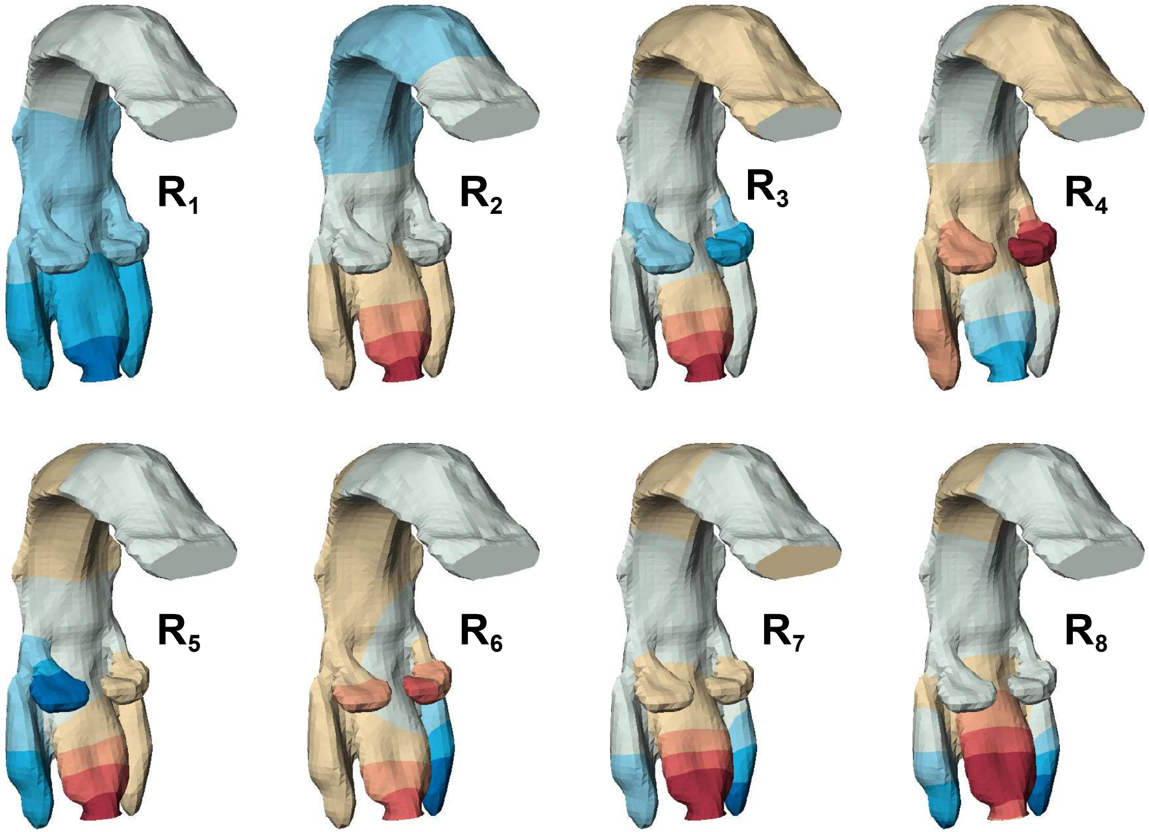
Φ = virtual sound



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

λ = Formant/resonance frequency

Dico: Solving the equations

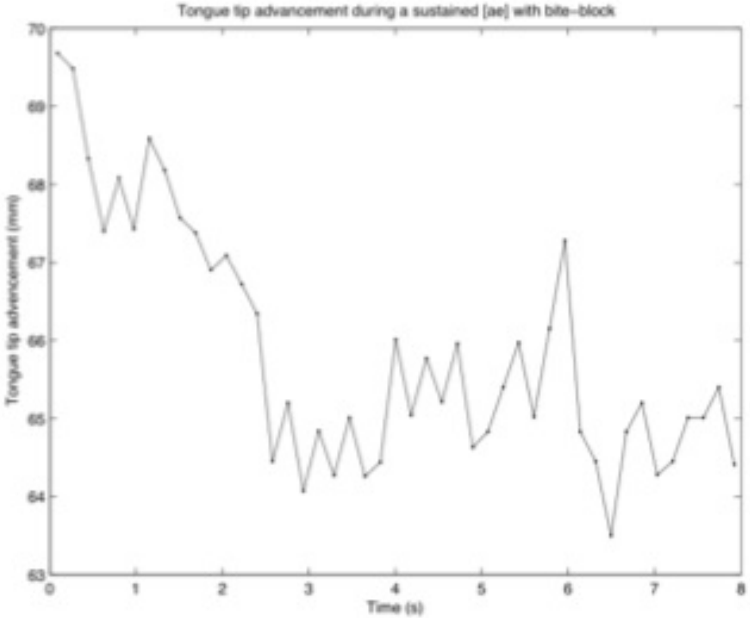


Computed standing waves

Validation details

- Are we solving the right equations?
- Possible error sources:
  - Instabilities in sustained vowel production
  - Spatial inaccuracies in MRI and image processing
  - Room acoustics, measurement equipment, noise cancellation

Vowel production stability



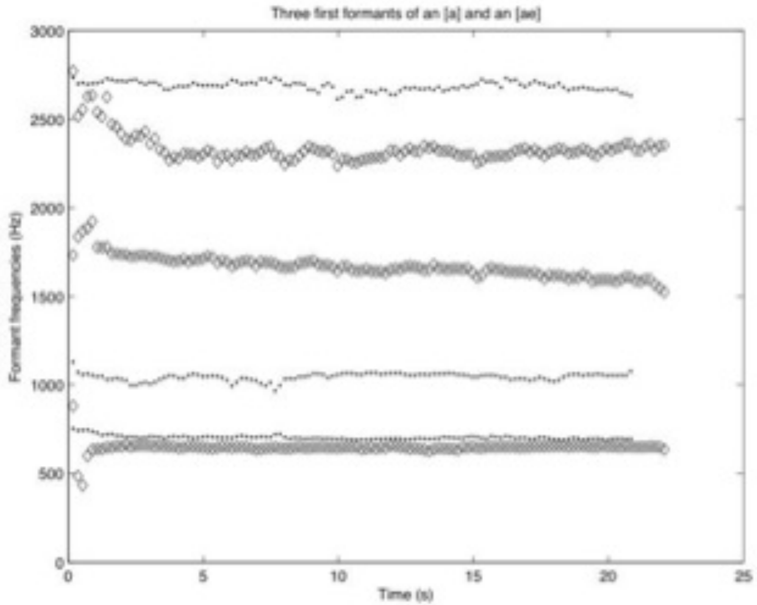
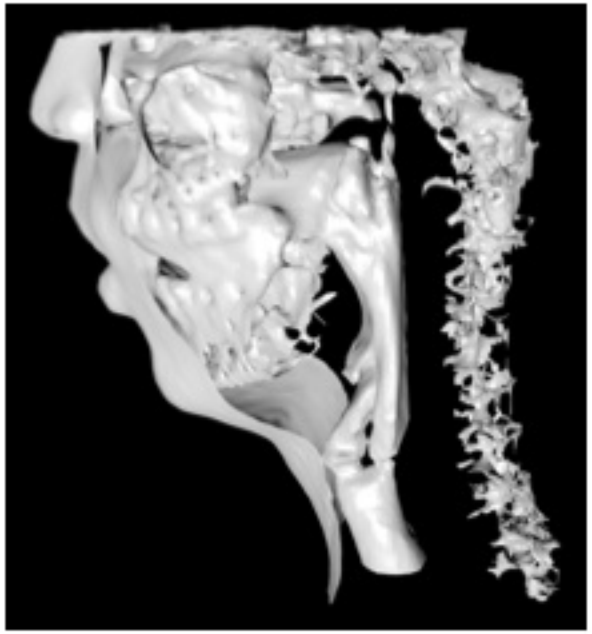
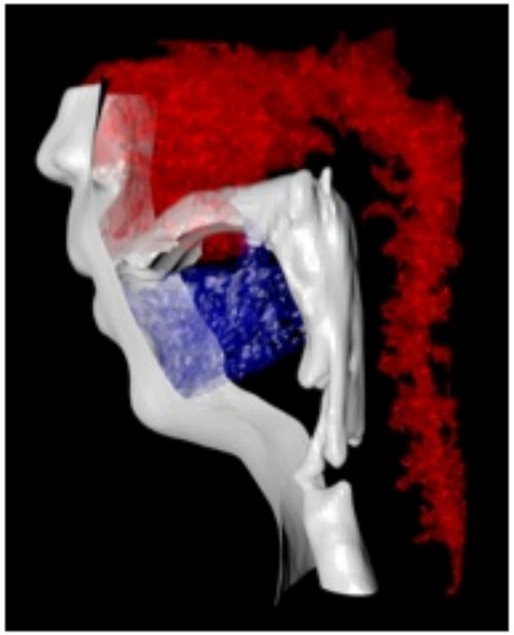


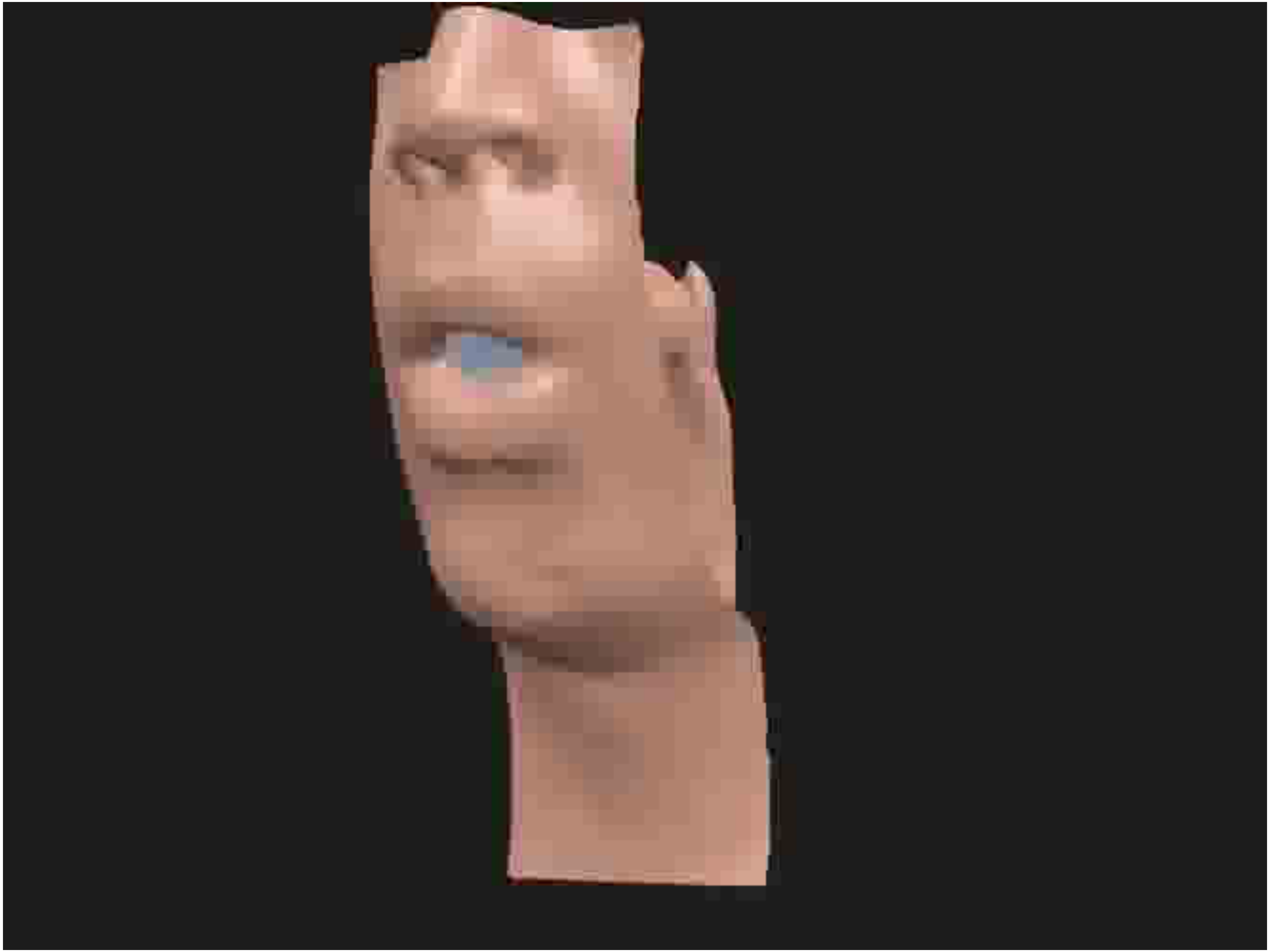
Image processing



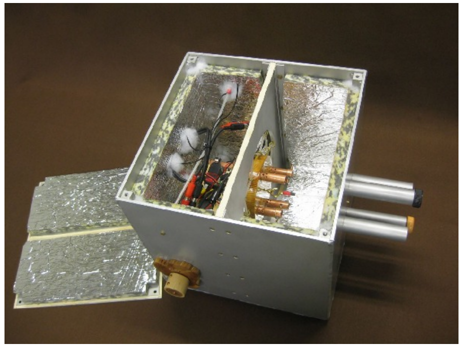


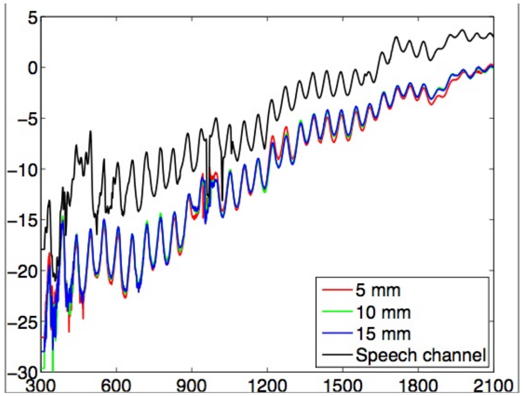
- Semi automatic volume detection
- Point Cloud Library
- Teeth superposition
- Surface model

Aalto, et al. 2013



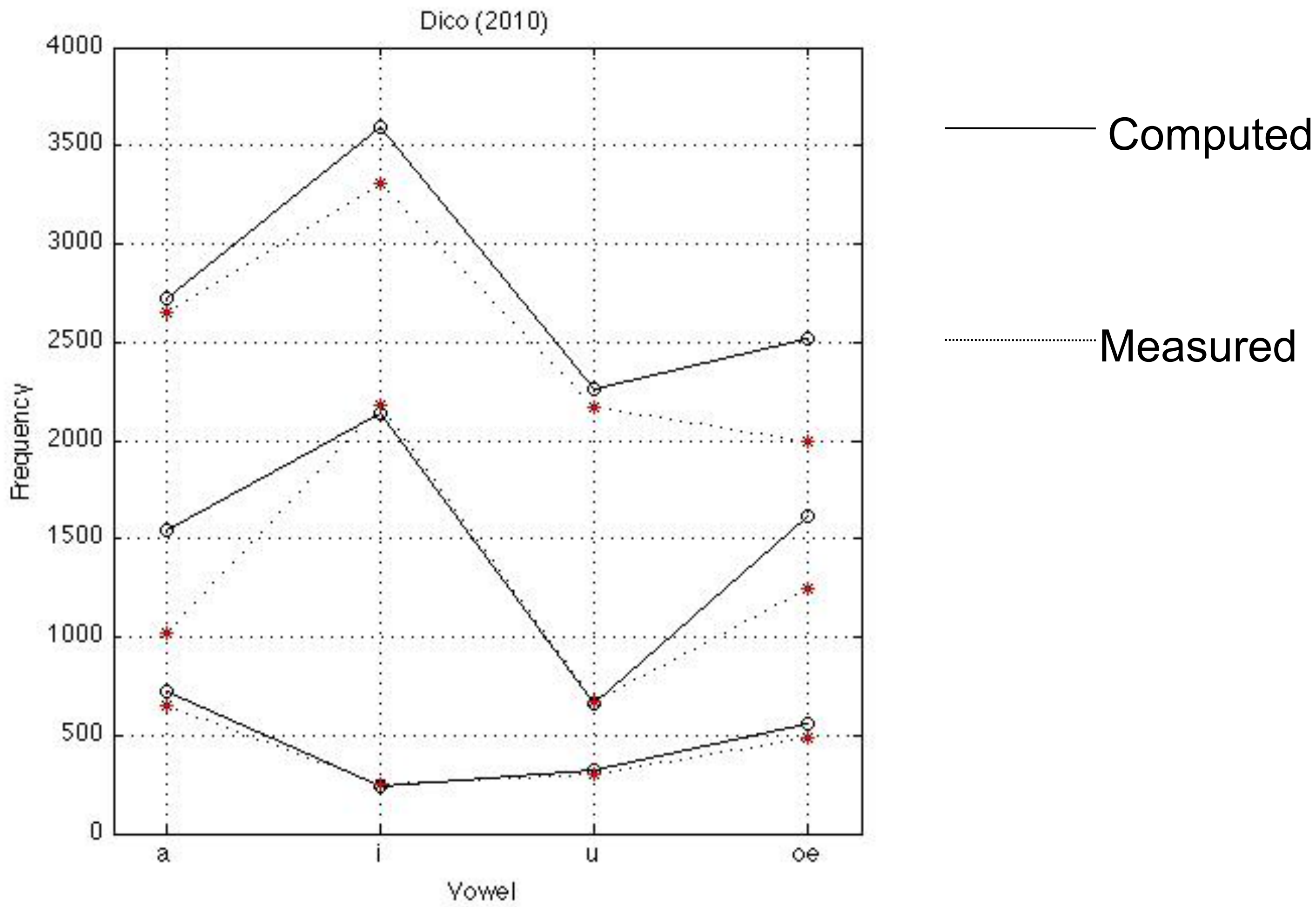
Sound recording





- Compensation for sound deformations due to the wave guides
- Noise cancellation
- Room acoustics still challenging

First results (pilot data)



Next steps

- Model improvement: lip boundary
- Complete analysis of the data
- Predictions for the post-op conditions
- Extending the methods to other patient groups
- Mobility analysis based on mid-sagittal videos
- Sensitivity analysis based on 3D geometries
- Visualization by multicolored 3D-prints

Generalizability of the approach

- The model is based on anatomy and physics: potential for reliable simulation of the vocal tracts independent of age, sex, ethnic group
- Model focuses on the passage from anatomy to acoustics: feasible for most patient groups
- Here the data is Finnish but every language in the world contains several vowels
- Some consonants are also captured by the model

Future questions

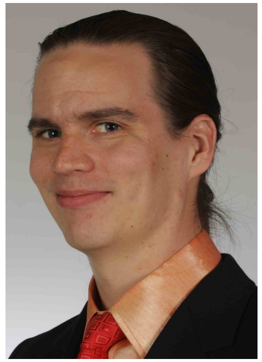
- Combining precise anatomical images (MRI) with moving vocal tract videos, could we pre-synthesize the speech following any surgery?
- Could a speech synthesizer be **personalized** to sound like the patients' own voice (e.g. cerebral palsy patients) based on the anatomy of the vocal tract?
- Could we predict the post-op **intelligibility** and/or **ease of articulation** of a patient undergoing oral and maxillofacial surgery?

University of Turku &
Turku University Hospital






University of Helsinki

Aalto University

Thanks!