# Computational modelling of speech

Docent Jarmo Malinen, DSc (tech.)

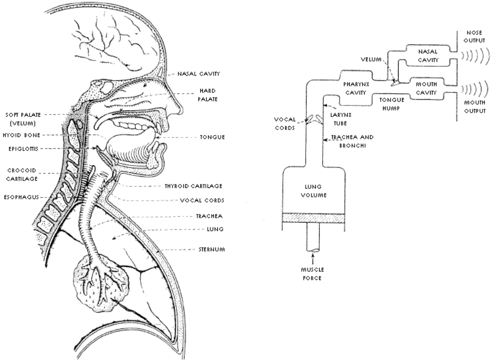
Aalto University,
Dept. Mathematics and Systems Analysis

October 22nd, 2016

# 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.

# Mathematical modelling of speech acoustics

# DICO – a model for vowel production based on MRI

DICO -vowel production model in nutshell:

- The mathematical modelling of speech based on vocal tract 3D anatomy, obtained by Magnetic Resonance Imaging.
- Numerical simulation of speech sounds based on MRI.
- Fine-tuning and validation is based on comparison between simulated speech and speech recorded during the MRI examination.
- Model experiments using 3D prints of vocal tracts.
- The total error in modelling and data acquisition must be significantly smaller than the effect of surgery on speech.

# DICO (1)

In the core of model DICO, there are two partial differential equations describing acoustics. For exact 3D vocal tract MRI geometries

$$\frac{\partial^2 \phi}{\partial^2 t} = c^2 \Delta \phi$$

and for 1D waveguides derived from 3D data

$$\frac{\partial^2 \psi}{\partial^2 t} = \frac{c(s)^2}{A(s)} \frac{\partial}{\partial s} \left( A(s) \frac{\partial \psi}{\partial s} \right) - \frac{2\pi \alpha W(s) c(s)^2}{A(s)} \frac{\partial \psi}{\partial t}.$$

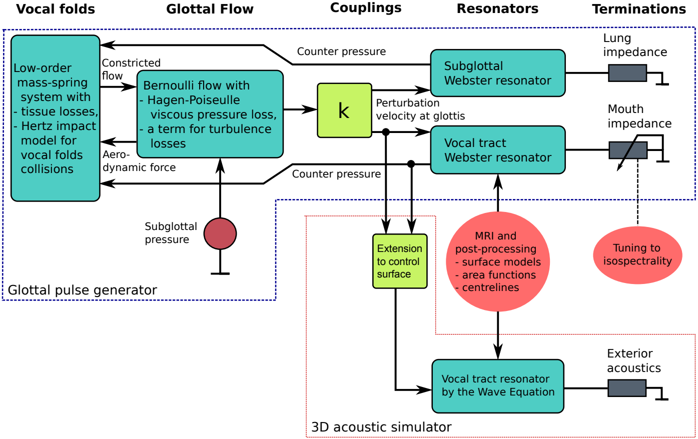
The are the wave equation and Webster's equation.

These equations are solved numerically by Finite Element Method in vocal tract configurations, obtained by MRI during speech.

The velocity potentials $\phi$ and $\psi$ produce a simulated sound signal that resemble the spoken vowel, if boundary conditions are used at the larynx end to apply an input signal that resembles the vocal folds vibrations.

# DICO (2)

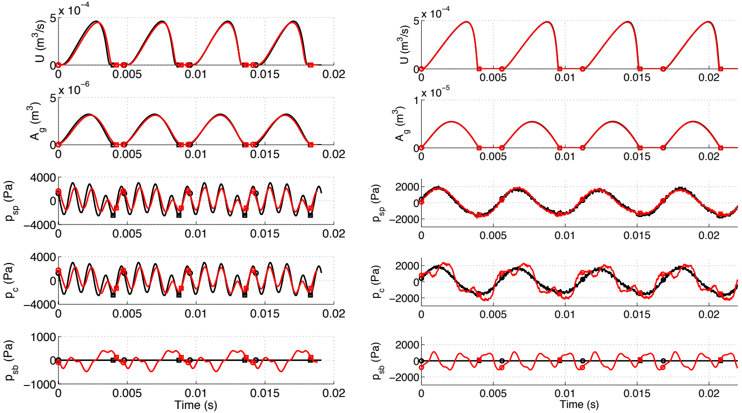
In addition to the vocal tract, we need to model vocal folds, lower airways, and even the exterior space so as to simulate vowels.




DICO is an interconnected network of mechanical, flow-mechanical and acoustic partial models.

# DICO (3)

Simulated glottal opening areas and sound pressures at larynx during production of the Finnish vowels [ɑ] and [i].




[ɑ] [i]

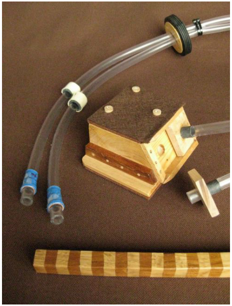
# Acquisition of patient data

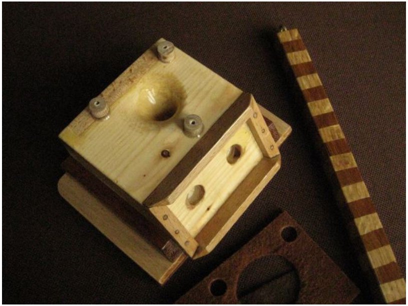
# Recording speech and MRI (1)


- Modelling speech and validation of the model requires **simultaneous** recording of speech and 3D MRI from test subjects.
- Metal or electronics cannot be taken inside the MRI scanner.

Some kind of stealth technology is needed!

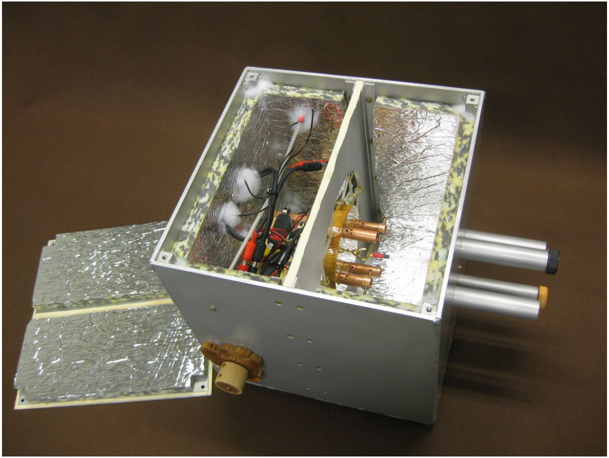
# Recording speech and MRI (2)





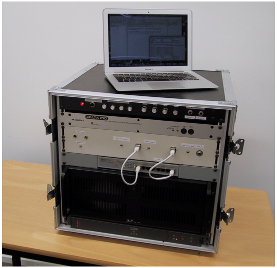
- Speech and the noise sample from MRI scanner are transmitted in "garden hose" the first 3 meters.
- Two-channel sound collector and the waveguides.
- The sound collector fits on the head and neck coils of Siemens Avanto 1.5T MRI scanner. No moving parts, immune to vibrations.

# Recording speech and MRI (3)



The waveguides lead to a microphone assembly inside a sound-proof Faraday cage, beside the MRI scanner.

# Recording speech and MRI (4)



- The signals are transmitted using shielded cables from microphones to a custom-made pre-amplifier and analogue sound processing unit.
- Signals are digitised by M-Audio Delta AD-converter.
- All electronics and the computers are in a movable rack so as to speed up its installation in MRI laboratory.
- The whole experimental arrangement is (almost fully) automatised for improved efficiency.

# Recording speech and MRI (5)

Unfortunately, Fast Prototyping Methods can be nowadays used to produce unromantic acoustic instruments lacking all soul and style.

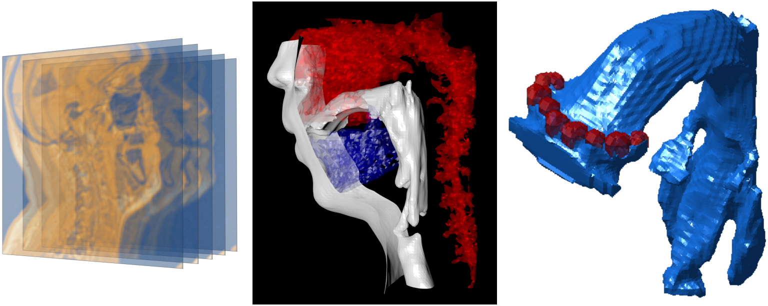
A MRI noise collector for an newer version of Siemens Avanto 1.5T that has been produced by Ultimaker Original in PLA.

# Post-processing of MRI and sound signals

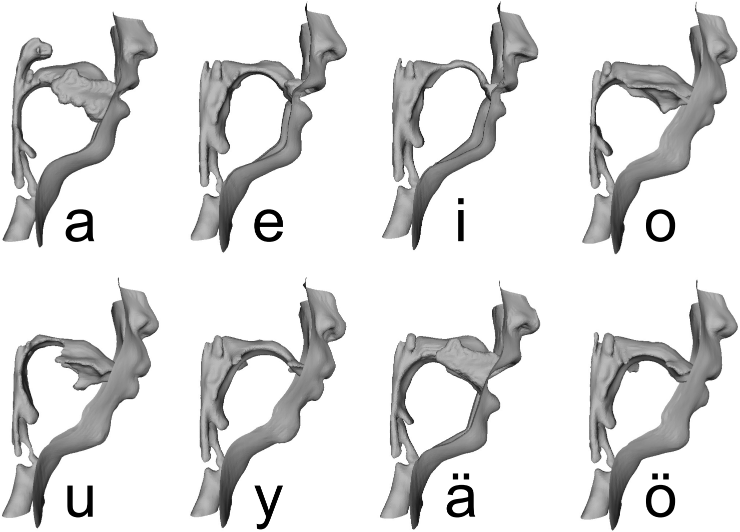
# From pixels to surface models...

Custom 3D image processing software produces (almost) automatically surface models from the (somewhat) blurry pixel data of the MRI scanner.

The air-tissue interface is separated from other anatomic structures in surface models.



To solve acoustic equations, the surface models need be further processed to Finite Element Meshes.

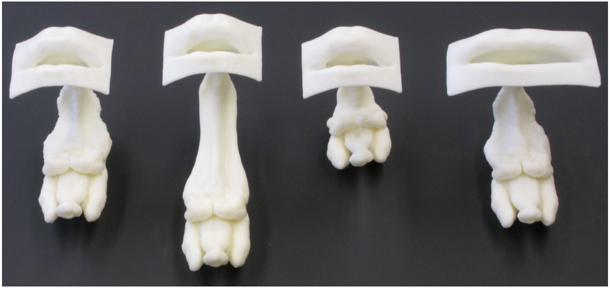

a
e
i
o
u
y
ä
ö

# ...and even to animations.

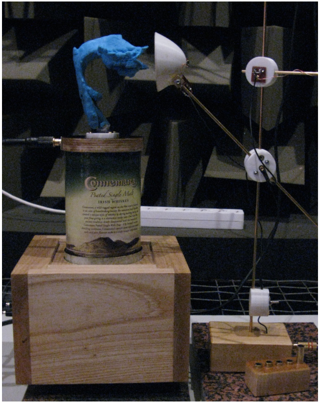
[Let us show them on full screen]

# Acoustic measurements from vocal tract models produced by fast prototyping

# Physical models by 3D printing (1)



# Physical models by 3D printing (2)



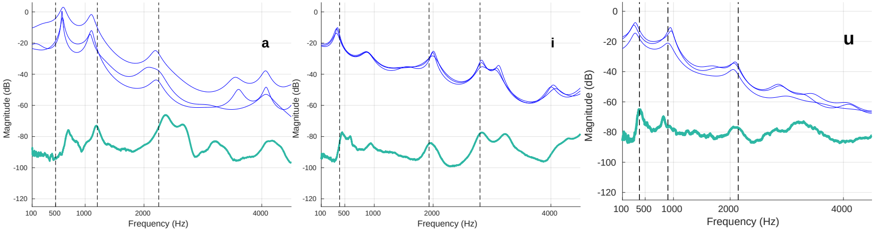

By carrying out laboratory measurements using 3D prints, it is possible to obtain independent comparison data for, e.g., model validation.

# Physical models by 3D printing (3)

# Accuracy of modelling and measurements of speech production at current stage

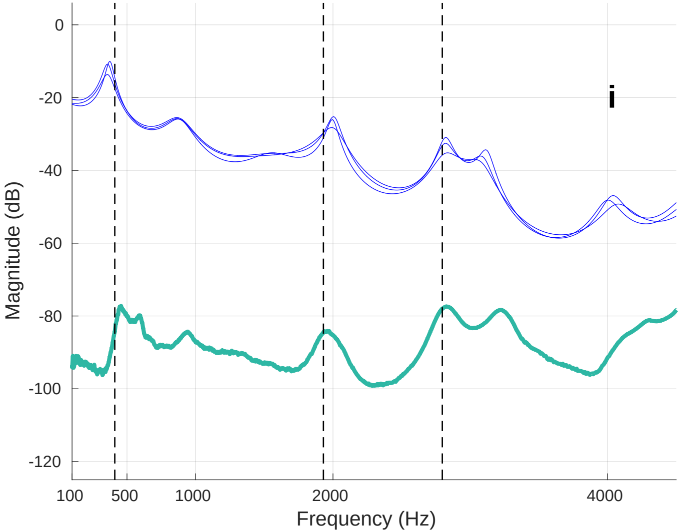
# Accuracy (1)

Vowel spectrograms ([ɑ], [i], and [u]) of one test subject measured from actual speech during MRI and from 3D prints by "sweeping".




Vertical dashed lines are resonances computed from MRI numerically.

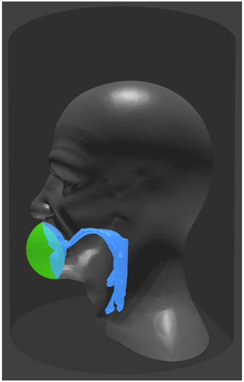
This is so nice that we better magnify it!

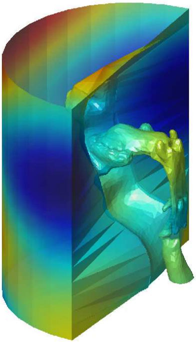


**Environment acoustics**
(an epilogue)

# The effect of the surrounding space (1)

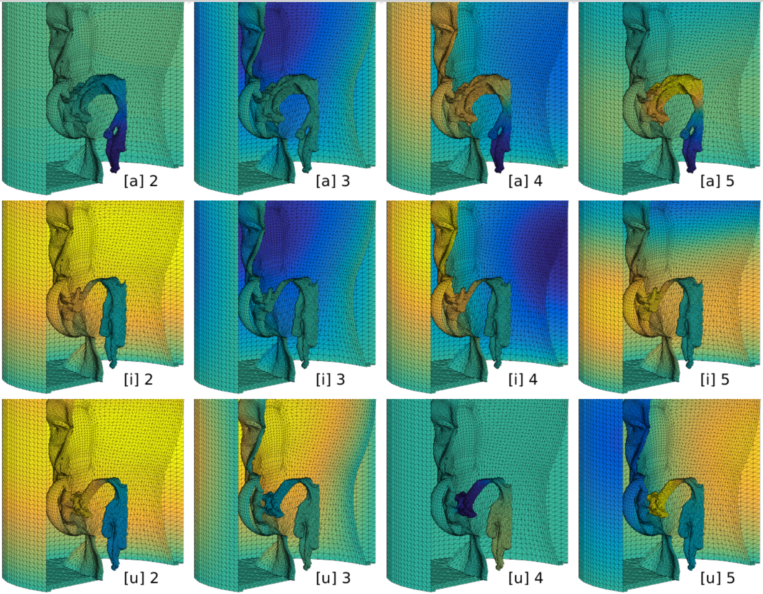
The single most significant remaining acoustic source of error is the effect of the surrounding space. This is work in progress.





The environment may be, e.g., open space or the MRI scanner coil.

# The effect of the surrounding space (2)




Resonant sound pressure fields of [a], [i], [u] in a constrained environment.

# Members of the COMSPEECH Research Consortium

Institute of Dentistry, University of Turku and Turku University Hospital

Institute of Behavioural Sciences, University of Helsinki

University of Alberta; Edmontor

Department of Mathematics and Systems Analysis, Aalto University, Helsinki