Glottal pulse modelling for vowel simulations

D.Sc. Jarmo Malinen

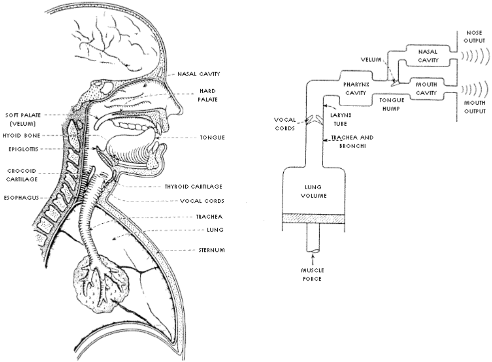
Aalto University, School of Science,
Department of Mathematics and Systems Analysis

University of Helsinki
March, 13th, 2015

Human voice production

Simplified vowel production:

$$\underbrace{\text{Mouth}}_{\text{(exterior load)}}$$
$$\uparrow$$
$$\underbrace{\text{Vocal tract}}_{\text{(filter)}}$$
$$\uparrow$$
$$\underbrace{\text{Vocal folds}}_{\text{(source)}}$$



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.

Purpose of the glottal pulse model

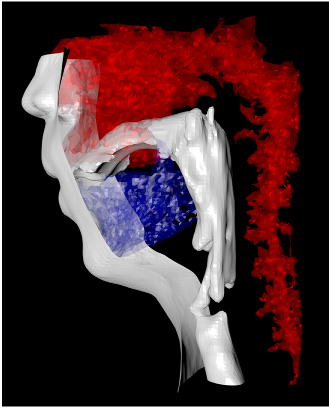
The glottal pulse model is designed for two purposes:

- As an adaptable and tunable **glottal source** for high-resolution acoustic model of vowel production “DICO”.
- As a **development platform** for speech processing algoritms. The glottal inverse filtering (GIF) is an example of this.

DICO is based on computational acoustics and MRI data.
It has been originally designed for predicting phonetic outcomes of maxillofacial surgery.

Modelling speech requires data

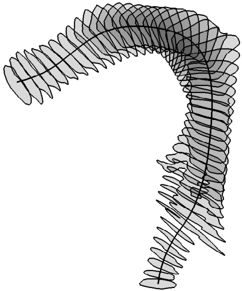
Simultaneous speech recording during 3D MR imaging.



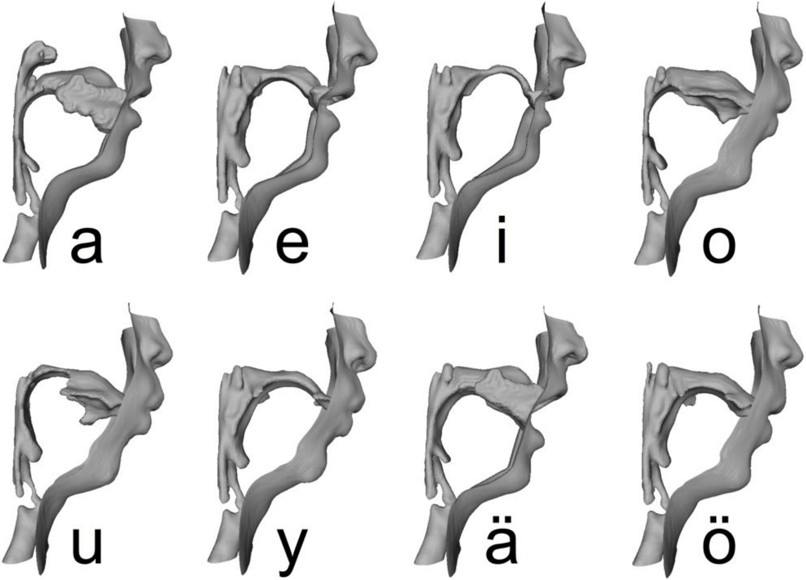
MRI is required for modelling.
Speech signal is required for model validation.

VT surfaces and area functions from MRI

The geometric data for acoustic modelling is obtained from Magnetic Resonance Images by custom segmentation software.

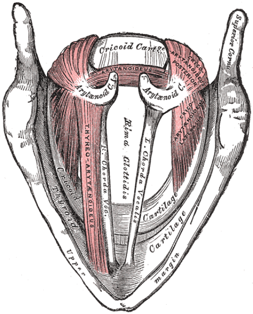


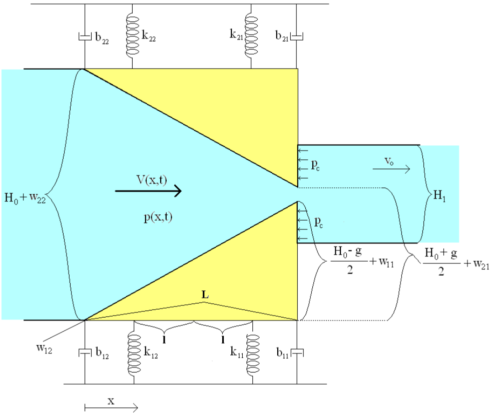
Geometries of Finnish vowels



Vocal folds

Vocal folds vibrate due to a self-sustained aerodynamic cut-off effect as a result of air flow from lungs.





Vocal tract as Webster's resonator

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

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

- c speed of sound
- ρ density of air
- α dissipation to tissues

- $A(0)$ Intersection area just above glottis
- $A(s)$ VT area function from MRI

Both $\Sigma(s)$ and $W(s)$ are corrections due to VT curvature.

Drawbacks of Webster's resonator

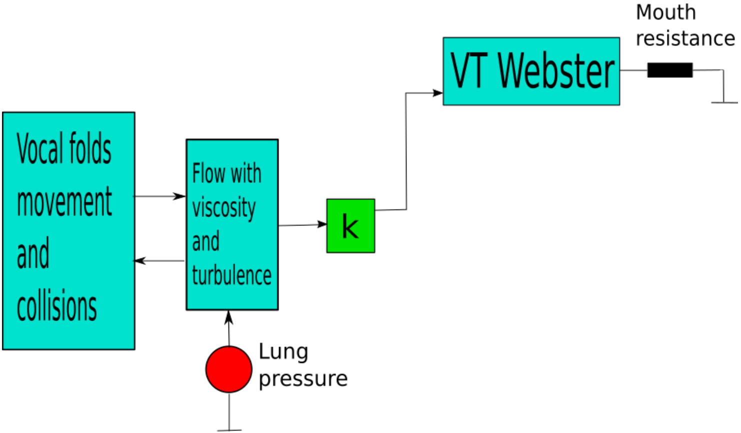
A high-resolution vocal tract acoustics model cannot be based on Webster's resonator:

- For three lowest formants under 4 kHz, Webster's resonator is **very accurate** and **numerically efficient**.
- Webster's resonator does not see cross-mode resonances in mouth cavity for, e.g., [ɑ].
- The resonance structure of valleculae and piriformis sinuses show up at 4+ kHz as a F_4-cluster. Not accounted by Webster's resonator at all.

However, Webster's resonator is sufficient for producing glottal pulse for a separate high-resolution model.

A simple source-filter model (1)

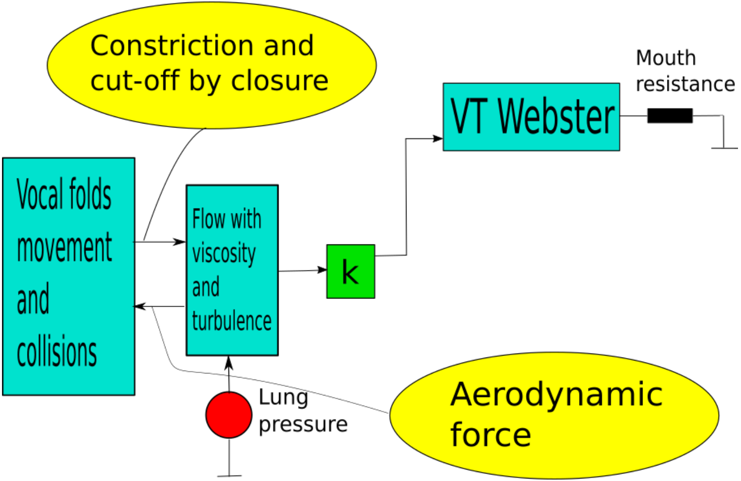
Coupling the glottis and flow model to Webster's resonator, we get a simple vowel synthesizer.



It does give (some kind of) vowels with correct formants, but does not really explain speech biophysics at all.

A simple source-filter model (2)

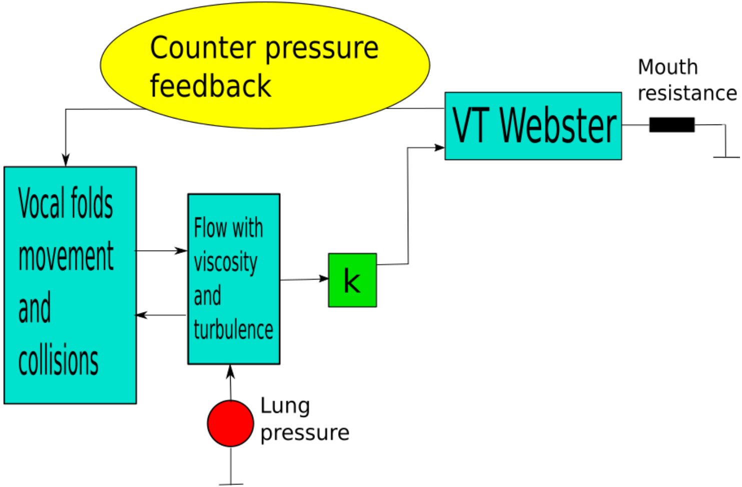
The vocal folds interact with the glottal flow in both directions.



This feedback produces sustained oscillations heard as sound.

Adding acoustic feedback

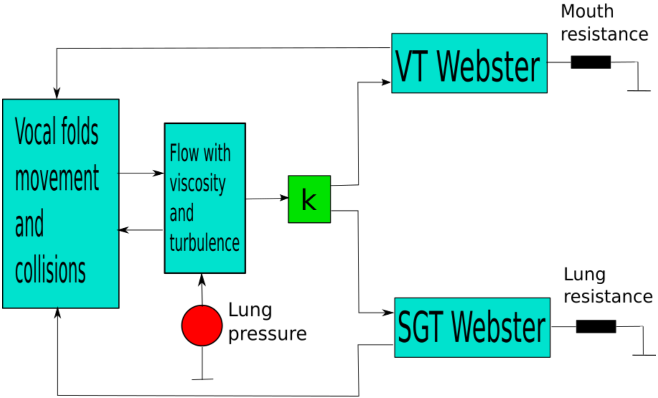
The next improvement is to let the vocal tract sound pressure affect the vocal fold vibrations.



This feedback is called source-filter coupling by I. Titze.

Adding acoustic feedback

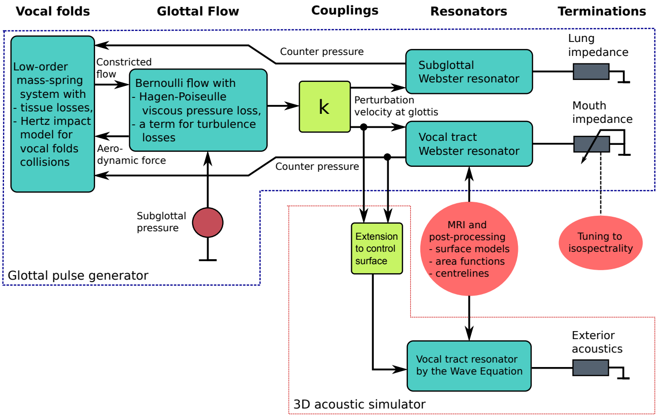
Even though lower airways under glottis are not actively used in speech, they do load the vocal folds in a similar manner as the vocal tract.



Now we have a fairly complete glottal pulse generator based on speech biophysics and geometric data from MRI.

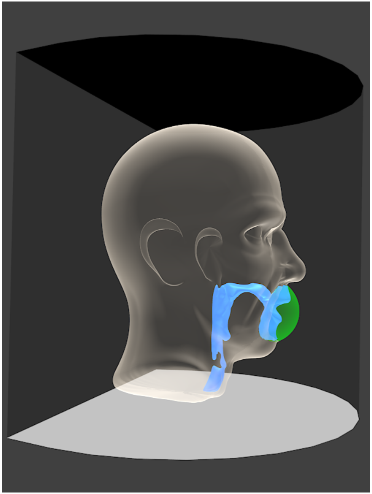
Multiphysics of vowel production: DICO

The full model feeds the glottal pulse to a high-resolution 3D acoustic model to obtain all the details.



Compromise between accuracy and numerical efficiency.

Environmental acoustics



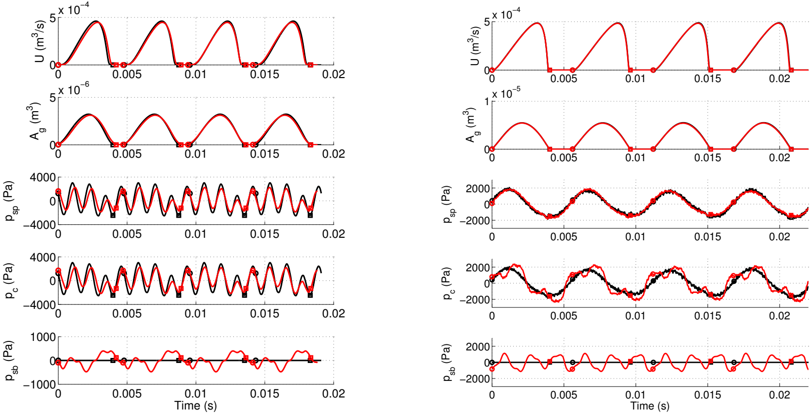
Even in otherwise empty space, the face of the speaker acts as a reflecting surface.

Speech in constrained space differs from speech in open space.

Speech data recorded inside the MRI coil is like **singing in a bucket**.

Dealing with the environment should be done in a numerically efficient way as it is, after all, a secondary feature.

Simulated glottal signals for [ɑ, i]



Signals in black are computed without subglottal resonator.

Technological side products

The novel technological solutions may have value that is independent of the modelling work:

- **Segmentation and registration of MRI material from VT** can be used for other purposes: e.g., numerical studies of flow mechanics in upper airways, related to Obstructive Sleep Apnea or asthma.
- High-quality **speech sampling during dynamic MRI** for studies of articulation, singing, etc.
- MRI experiments during prolonged vowel utterances.

Almost all of the potential applications are uncharted territory right now... and we are interested in cooperation.

Thanks for your patience.
Any questions?

http://speech.math.aalto.fi

Joint work of groups in Aalto University, University of Helsinki, University of Turku, Turku University Central Hospital, and Medical Imaging Centre of South-West Finland.

“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, T. Murtola, and M. Vainio.
Modal locking between vocal fold and vocal tract oscillations: Simulations in time domain. *arXiv:1211.4788*, 2012–2015 (and still going on and on).

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*, 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. *Applied Acoustics* 83(1):64–75, 2014.

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

T. Lukkari and J. Malinen.
A posteriori error estimates for Webster’s equation in wave propagation. *Journal of Mathematical Analysis and Applications*, 2015.