Aalto University
School of Science
and Technology

Mouth impedance optimisation for vocal tract of vowels

Juha Kuortti, Joonas Kivi, Jarmo Malinen, and Antti Ojalammi

Department of Mathematics and Systems Analysis
Aalto University
School of Science

October 29, 2015

Human vowel production

Vowel production: Human speech production has traditionally been modelled with *source-filter model*

Mouth

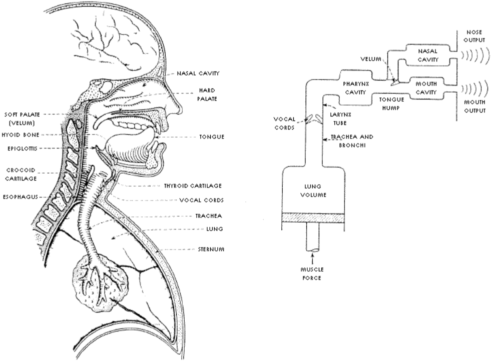
↑

Vocal tract
(filter)

↑

Vocal folds
(source)

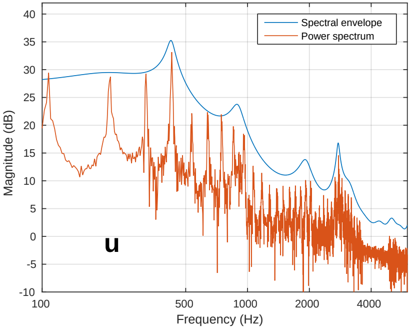
Assumption: Source and filter are **independent.**



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

Defining characteristic: formants

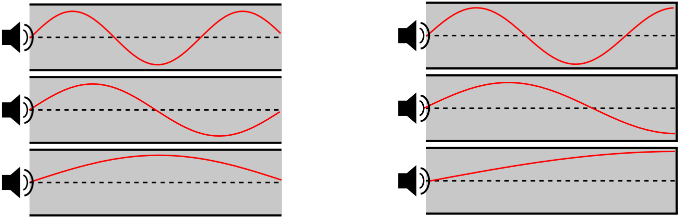
Assume that speech signal $x[n]$ is autoregressive:
$x[n] = a_1 x[n-1] + a_2 x[n-2] + \ldots + a_p x[n-p]$, i.e. the filter of previous slide is *all-pole*. By estimating the coefficients $\mathbf{a}$ we can define $H(z) = A(z)^{-1}$, and plot $H(i\omega)$.



The peaks of the *spectral envelope* are called *formants*.

Defining characteristic: resonances

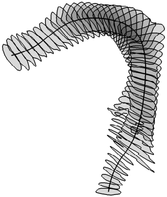
A resonator (such as glottis) inside of a cavity (such as vocal tract), will cause high amplitude oscillations at certain frequencies determined by the geometry.



These computationally obtained frequencies are called resonances.

Determining the characteristics: data acquisition

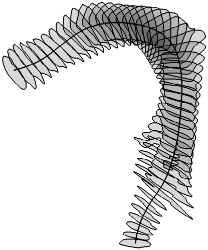
MR Imaging→Image data →Surface model →Area functions



Determining the characteristics: Webster's equation

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

- ψ_λ velocity potential
- c speed of sound
- $A(s)$ area at s
- $\Sigma(s)$ curvature correction



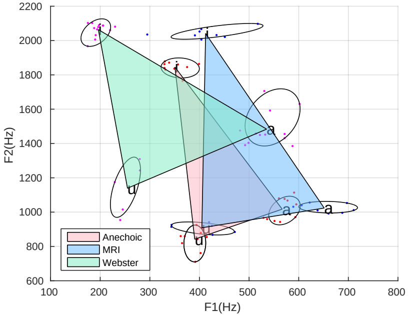
Determining the characteristics: Boundary conditions

If we assume that vocal tract behaves like a semi-closed cylinder:

- Dirichlet boundary condition at open end (mouth) $\frac{\partial \phi}{\partial t}(\ell, t) = 0$.
- Neumann boundary condition at glottis: $c\frac{\partial \phi}{\partial \nu}(0, t) + \frac{\partial \phi}{\partial t}(0, t) = 0$ to prevent energy leaking back to vocal tract.

Boundary conditions and exterior acoustics

Unfortunately, the simple assumptions won't work. The exterior space influences the acoustics heavily.



Determining the characteristics: Webster's equation 2

We improve the boundary conditions of the equations at the mouth to include acoustic *impedance*:

$$\begin{cases} \left(\frac{{\lambda_\theta}^2}{c^2\Sigma(s)^2} + \frac{2\pi\alpha W(s)\lambda_\theta}{A(s)}\right)\psi_{\lambda_\theta} = \frac{1}{A(s)}\frac{\partial}{\partial s}\left(A(s)\frac{\partial\psi_{\lambda_\theta}}{\partial s}\right) \quad \text{on} \quad [0,\ell] \\ \left(\frac{\lambda_\theta}{c}\right)\psi_{\lambda_\theta}(0) - \frac{\partial\psi_{\lambda_\theta}}{\partial s}(0) = 0, \\ \left(\frac{\lambda_\theta}{c}\right)\psi_{\lambda_\theta}(\ell) + \theta\frac{\partial\psi_{\lambda_\theta}}{\partial s}(\ell) = 0, \qquad \theta = \nu + i\eta. \end{cases}$$

Unfortunately, determining the value of θ through measurements is impossible, so it will have to be determined computationally.

Motivation for boundary condition

Arnold Sommerfeld defined a boundary condition for Helmholtz equation that forces the energy from source to scatter into the infinity, and thus prevents the ”energy from infinity” to be radiated back into to the field.

$$\left(\frac{\partial}{\partial |x|} - ik \right) u(x) = 0.$$

The boundary condition employed by us is actually a lossy version of the Sommerfeld boundary condition

$$\left(\frac{\partial}{\partial s} + \frac{\lambda_\theta}{\nu + i\eta} \right) \psi_\theta = 0.$$

Optimisation: setup

Goal: We want to match the three first resonances produced by Webster's equation to the first three formants measured from the speech signal.

We have one parameter to optimise, θ in the boundary term
$\left(\frac{\lambda_\theta}{c}\right)\psi_{\lambda_\theta}(\ell) + \theta\frac{\partial\psi_{\lambda_\theta}}{\partial s}(\ell)$

There are some issues pertaining to optimisation

- Three objectives far apart in frequency range.
- We are essentially interpolating boundary condition from Dirichlet to Neumann, which will cause a change in wavenumber at certain values of θ

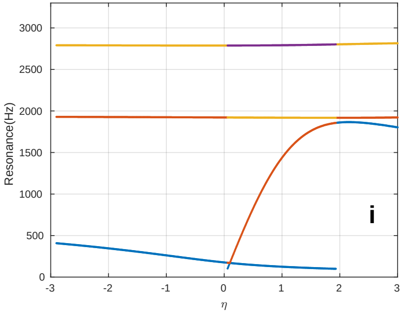
Optimisation: implementation

Since discrepancies in formant frequencies are measured in semitones, the first issue is solved by using a objective function f to reflect this:

$$f(\theta) = \sum_{j=1}^{3} \left| \log_2 \left(\frac{R_j(\theta)}{F_j} \right) \right| .$$

Optimisation: implementation

Fundamental change in the type of boundary condition can cause the wavenumber of the waveform to change.



Solved by determining the feasible areas manually, and imposing constraints for optimisation

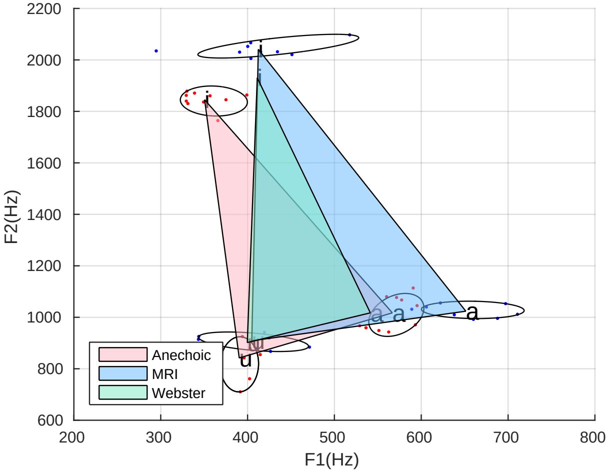
Results:Discrepancy

After optimisation we measure the discrepancies in semitones for the first three resonances and formants:

$$D_j = 12\log_2\left(R_j(\theta_{opt})/F_j\right)$$

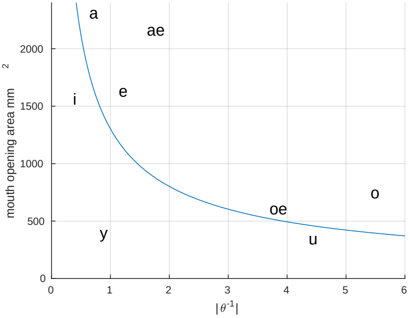
	[ɑ]	[e]	[i]	[o]	[u]	[y]	[æ]	[œ]	abs. avg
D_1	-0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
D_2	0.0	0.0	-0.9	0.0	0.0	-0.7	1.8	2.2	0.7
D_3	-3.0	1.1	0.0	3.5	1.4	1.8	2.8	2.3	2.0

Results:Discrepancies



Relation between θ and geometry

Impedance is a measure of how much the motion induced by pressure applied to a surface is impeded. The assumption is that the optimal impedance is somehow tied to the inverse of the mouth opening.



The End

Questions?

http://speech.math.aalto.fi