Computational modeling of human speech

... FEM in the human vocal tract

Atle Kivelä
and
A. Aalto, D. Aalto, J. Malinen, T. Murtola, A. Hannukainen,
A. Huhtala
(from Aalto University)

FemGrill 2012-09-20

Computational speech modelling

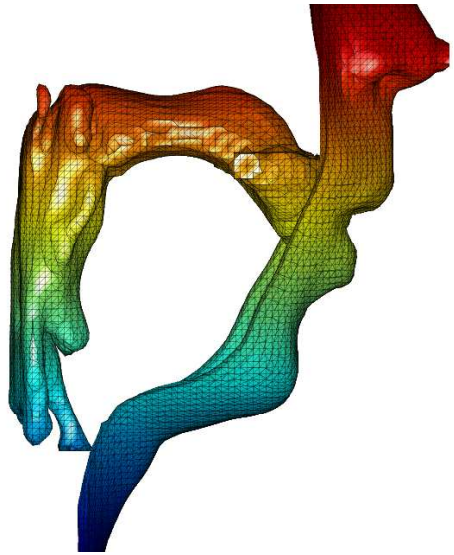
- Multidisciplinary research combining technology, modelling, phonetics and medicine.
- Telephony
- Speech synthesis
- Planning of maxillofacial surgeries

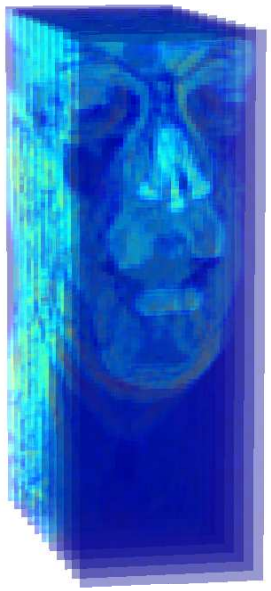
$$\Phi_{tt} = c^2 \Delta \Phi \quad \text{in} \quad \Omega \times \mathbb{R}$$

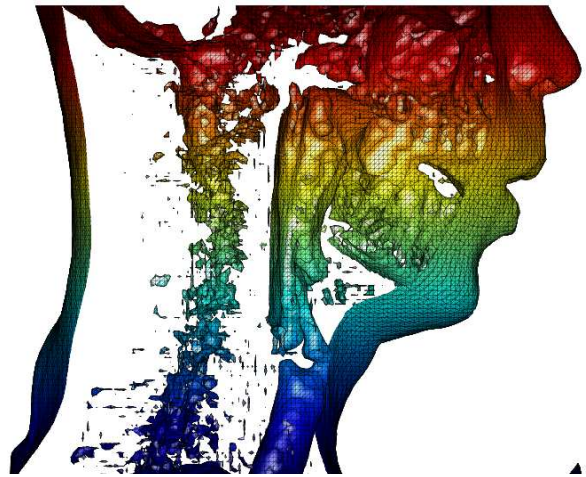
$$\Psi_{tt} = \frac{c^2}{A(s)} \frac{\partial}{\partial s} \left(A(s) \frac{\partial \Psi}{\partial s} \right) \quad \text{in} \quad [0, l] \times \mathbb{R}$$

Geometry extraction

- Magnetic Resonance Imaging
- Things to worry about:
  - Teeth
  - Artefacts
  - Registration
  - Automation

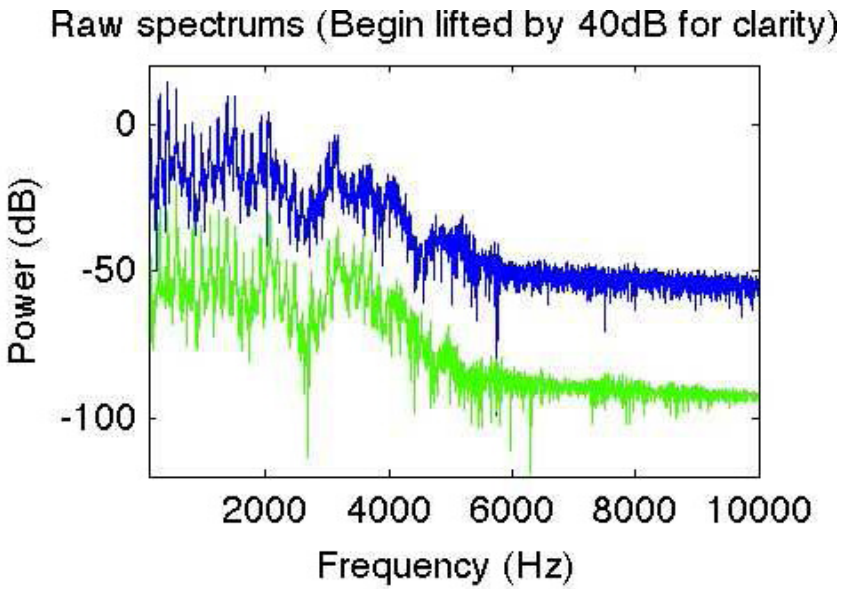


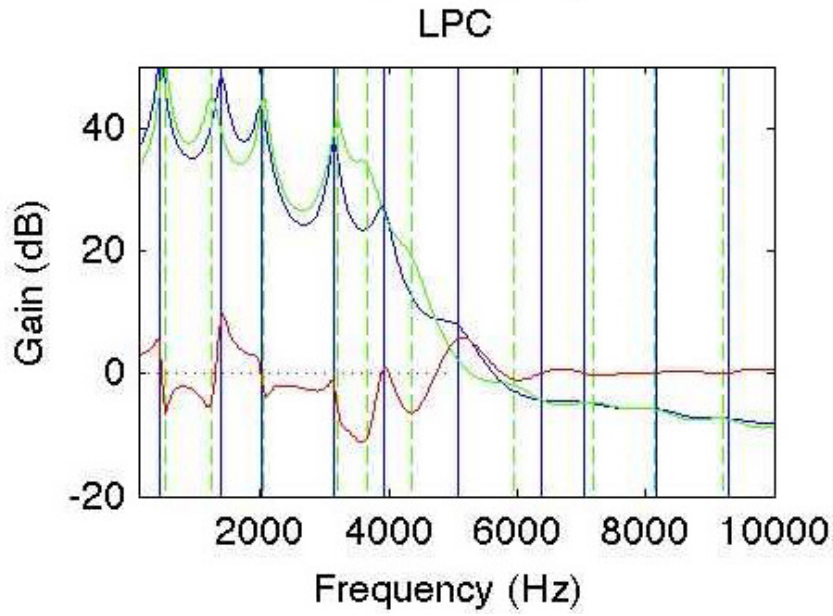




Formant

Formant A resonance frequency determined from audio recording F_i





Resonance

Resonance A numerically determined resonance frequency λ_i

$$\begin{cases} \lambda^2 \Phi_\lambda = c^2 \Delta \Phi_\lambda, & \text{in VT} \\ \Phi_\lambda = 0, & \text{at mouth} \\ \frac{\partial \Phi_\lambda}{\partial \nu} = 0, & \text{on VT walls} \\ \lambda \Phi_\lambda + c \frac{\partial \Phi_\lambda}{\partial \nu} = 0, & \text{at glottis} \end{cases}$$

$\longrightarrow$ formulation as an eigenvalue problem

$\longrightarrow$ estimates for λ_i

Some results

Vowel	R_1	R_2	R_3	R_4
[ɑ]	720	1547	2721	4138
[i]	246	2135	3592	4667
[u]	324	659	2262	3091
[œ]	562	1612	2519	3602

Table: First four resonances (in Hz) for the Helmholtz problem by FEM.

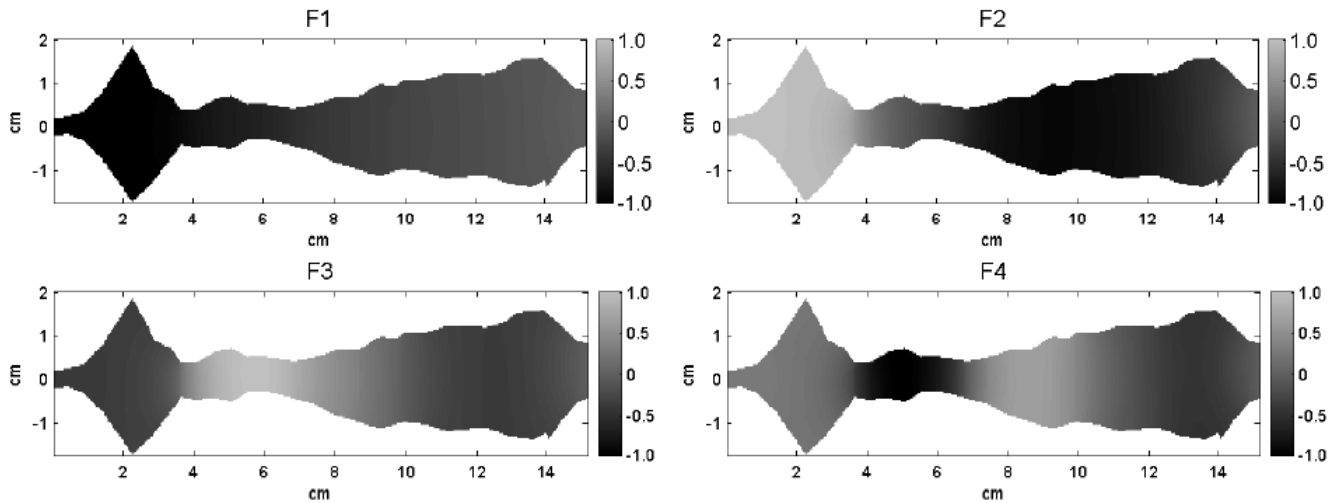
Vowel	F_1	F_2	F_3	F_4
[ɑ]	651 ± 7	1024 ± 35	2647 ± 117	3679 ∓ 36
[i]	247 ± 9	2183	3304 ∓ 46	4407 ∓ 251
[u]	306 ∓ 37	675 ∓ 39	2173 ± 13	3242 ± 139
[œ]	483 ∓ 35	1249 ± 74	1994 ∓ 50	3188 ∓ 17

Table: Formants (in Hz) computed as means of those extracted from the beginnings and ends of the sound samples.

How good is it?

Vowel	D_1	D_2	D_3	D_4	mean discr.
[ɑ]	1.7	7.1	0.5	2.0	2.8
[i]	-0.1	-0.4	1.4	1.0	0.7
[u]	1.0	-0.4	0.7	-0.8	0.7
[œ]	2.6	4.4	4.0	2.1	3.3

Table: Discrepancy (in semitones) between computed resonances and mean formant frequencies.



Future

- Geometry extraction is (to this point, more or less) robust
  - so new vowels can be computed
  - the code will most likely also be released in some form
- Getting data from patients undergoing orthognathic surgery.
- Teeth models
- Mechanical models