Tongue motor control stability: integrating feedback, dynamical internal representation and optimal planning.

Pierre Baraduc, Pascal Perrier
GIPSA-lab, UMR 5216 CNRS / Univ. Grenoble-Alpes / Grenoble-INP
{pierre.baraduc, pascal.perrier}@gipsa-lab.grenoble-inp.fr

Funded by CNRS/PEPS grant

INTRODUCTION

Key features of sensorimotor systems:

- Multisensory integration
- Use of internal models to predict the sensory outcomes of actions
- Comparison of the sensory input with the internal prediction to optimally update the internal estimate of the system

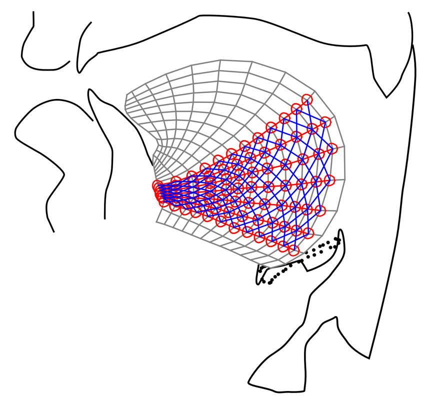
Speech production

- Coordination task: lips, jaw, tongue
- Resistance to external disturbances (inertial forces, objects in mouth, distorted audio feedback...)
- Can optimal feedback control theory illuminate the control of tongue movements during speech (tongue kinematics, coarticulation, use of feedback...)?

METHODS

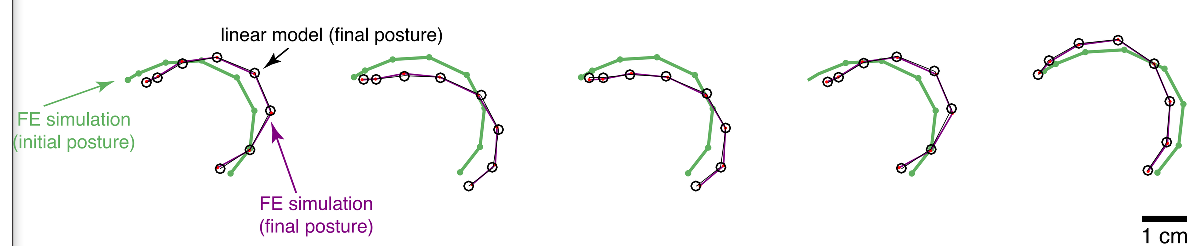
Tongue biomechanics:

- Six muscles modelled: anterior genioglossus, posterior genioglossus (illustrated), hyoglossus, styloglossus, verticalis, inferior longitudinalis [1]
- Finite element (FE) model of tongue deformation (sagittal 2D model)



Tongue reduced model:

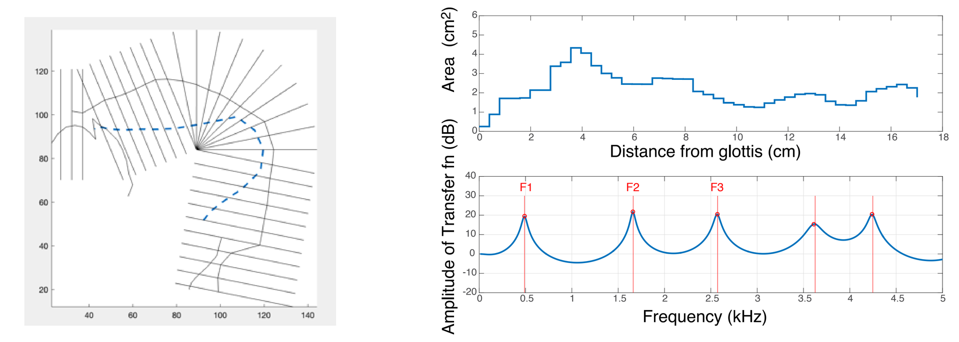
- Dimensionality reduction step: 17 contour points → 4D (SVD)
- Third order linear system, obtained by system identification of ~48,000 FE simulations at 100 Hz corresponding to control ramps of 33, 45, 65 and 120 ms duration, sampling command space



- Best linear approximation: $\Delta x_{k+1} = f(x_k, a_{k-1}, a_k)$ where x_k is tongue position, a_k the muscle activation at time t_k (RMS error = 0.42 mm)
- Nonlinear approx: 3-layer artificial neural network with 24 hidden units from ~55,000 FE simulations

Vocal tract, from tongue shape to vowel:

- For a given external tongue contour, a fixed jaw position, and a fixed lip aperture, we deduce the shape of the complete vocal tract using anatomical reference data (MRI, [2])
- We then compute the resonances of the vocal tract following the method of [3] after discretization of the tract in 44 tubes of same length, and keep the first three formants.

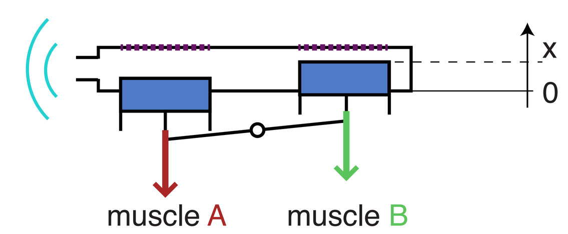


Contact management:

- Tongue contour reconstructed from reduced model
- Soft clamping of position and velocity as a function of margin to palate & margin velocity

Exploring the control of contacts:

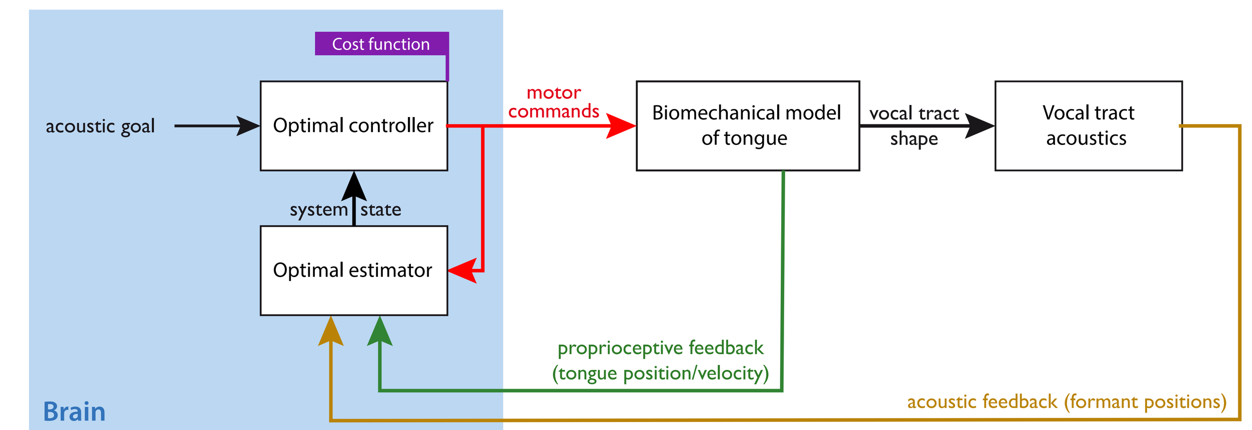
- One degree of freedom corresponding to principal component
- Tube model with auditory, proprioceptive and tactile feedback
- Muscular redundancy, inertia, elasticity towards neutral
- Intuition for more complex models, while convergence easier



Optimal control model:

Plant:

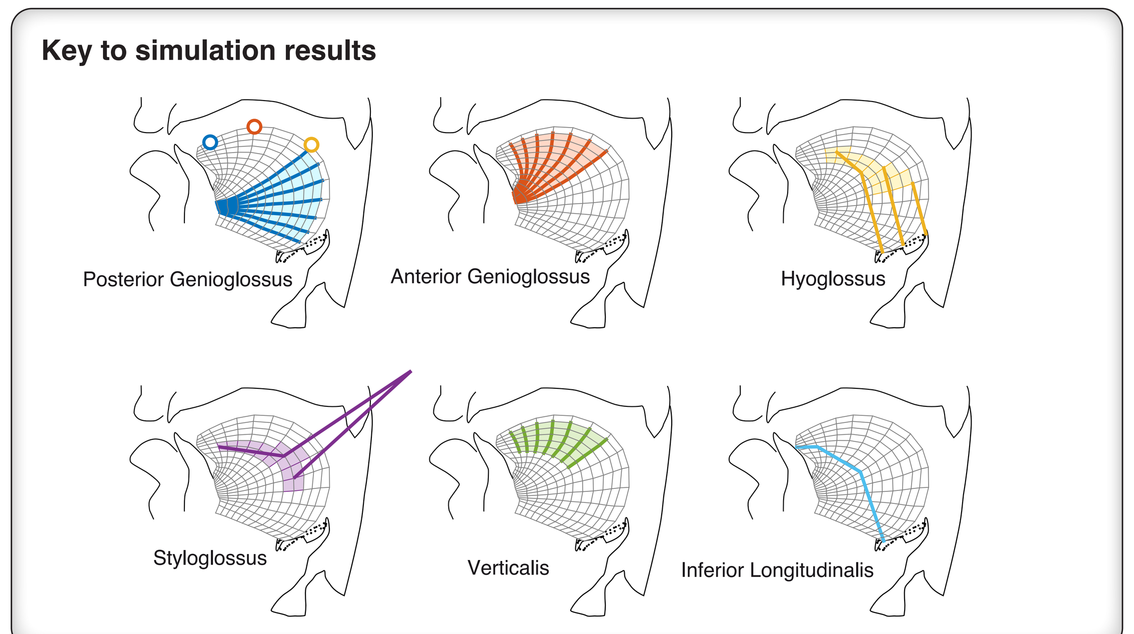
- Discrete time dynamics $s_{k+1} = F(s_k, u_k + \varepsilon(u_k))$ simulated at 500 Hz
- State vector $s_k = [x_k, \dot{x}_k, \dot{x}_{k-1}, \dot{x}_{k-2}, a_{k-1}, a_k, e_k]$
- Cascade of first order differential equation from command u_k to excitation e_k to activation a_k (τ = 10 ms)
- Only motor noise: additive (SD σ_A) and multiplicative (SD σ_M) Gaussian white noise on motor command
- Full (4D) proprioceptive feedback, 3D acoustic feedback



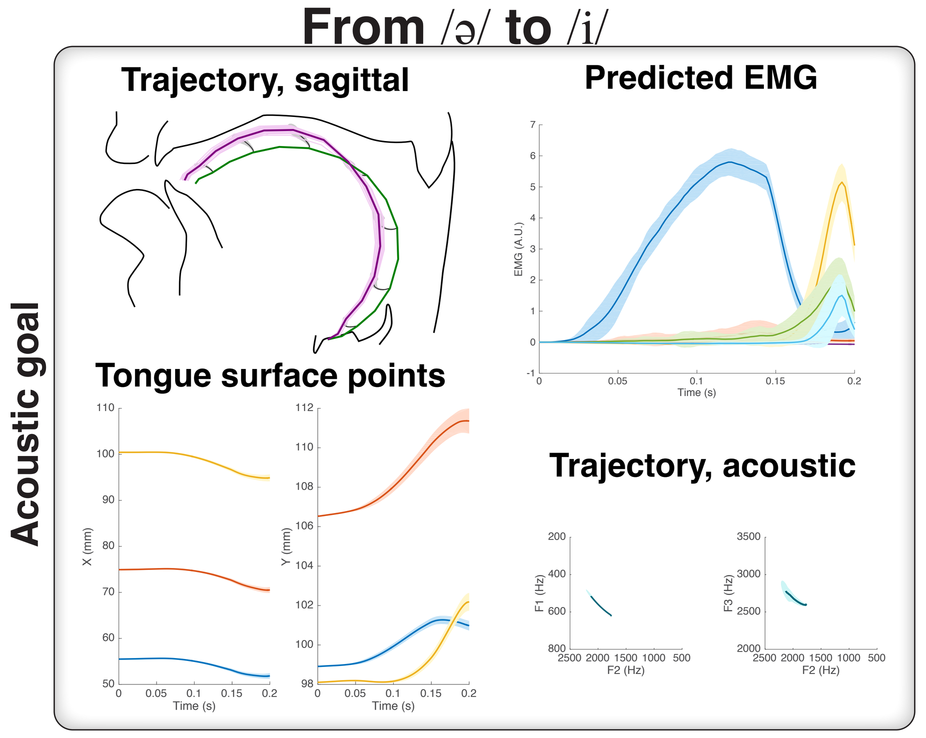
Optimization:

- Cost = integrated neuromotor effort $\sum_0^T u_k^2$
- Constrained:- final acoustic target or tongue posture
- Unconstrained: mix of effort and goal costs
- Indirect optimal control (Pontryagin based), gradient descent and/or Newton-Raphson method [4]
- Some checks of sensitivity to initial parameters

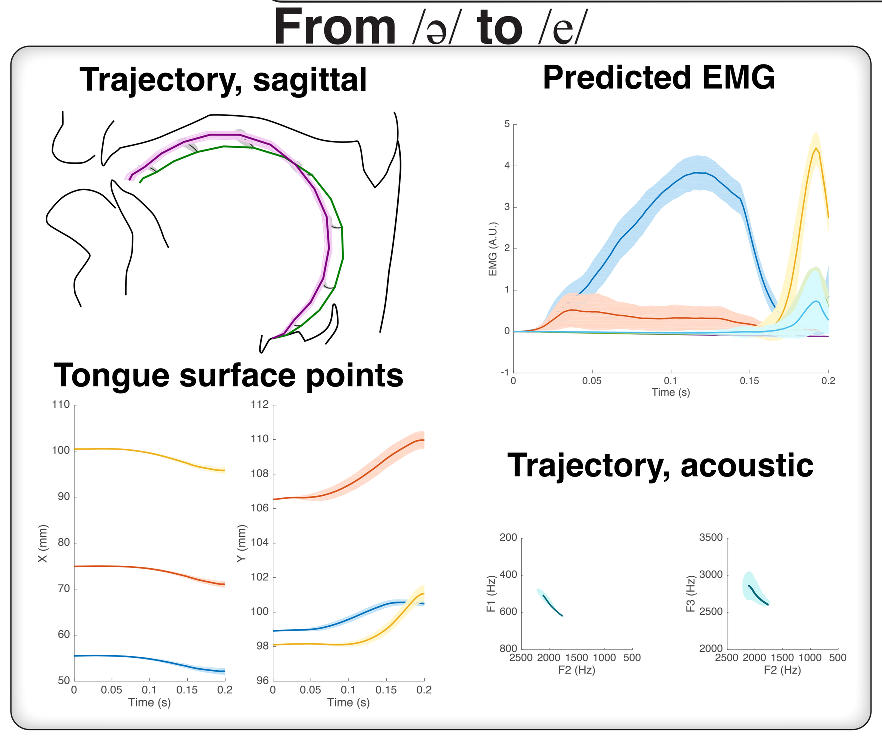
RESULTS



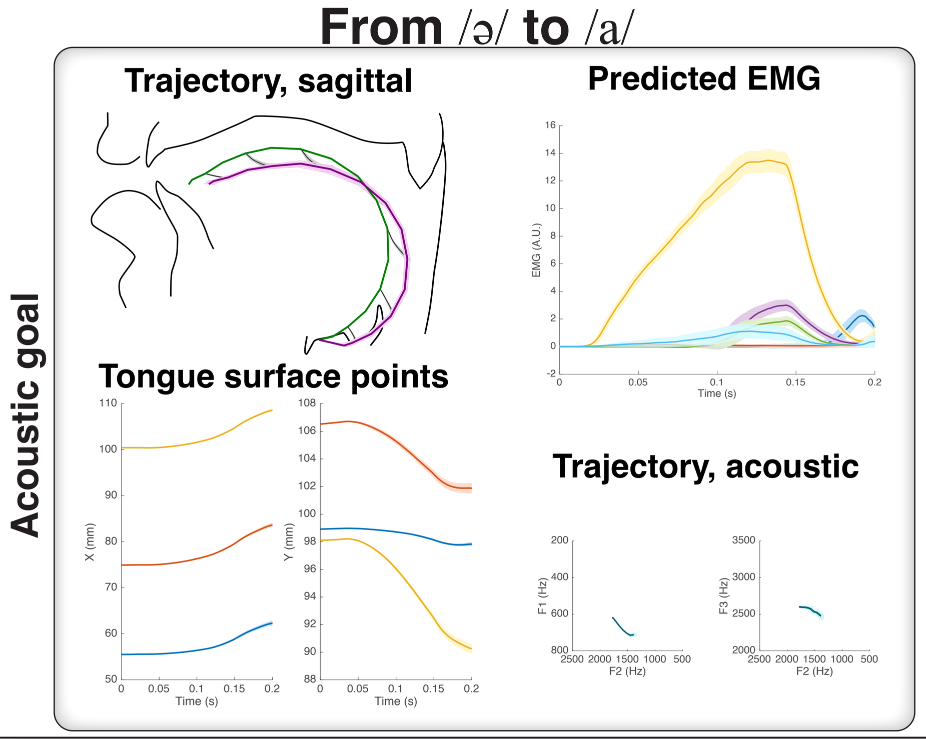
From /ə/ to /i/



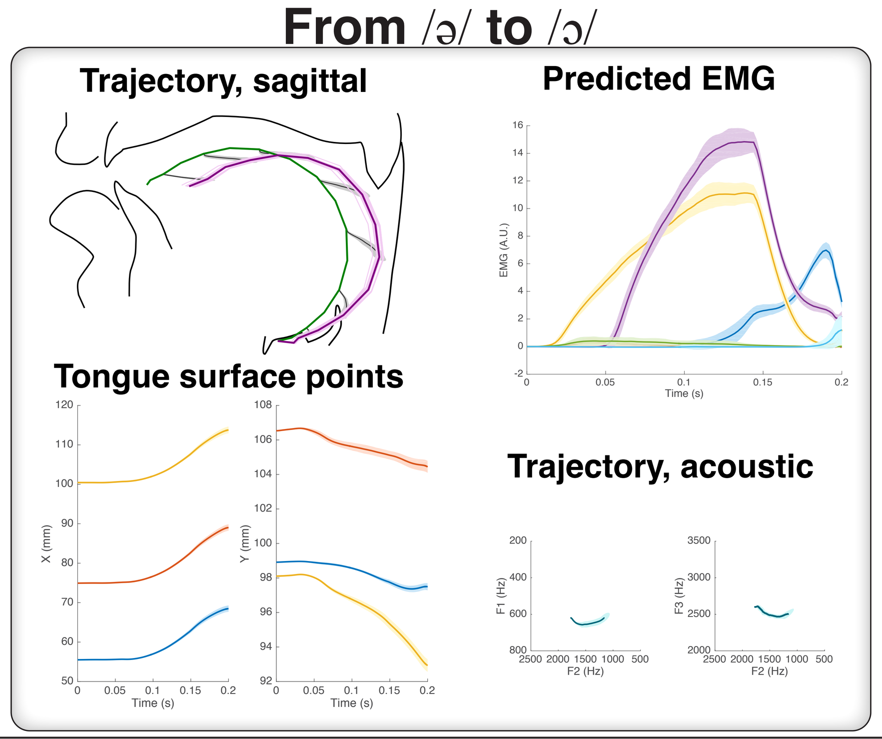
From /ə/ to /e/



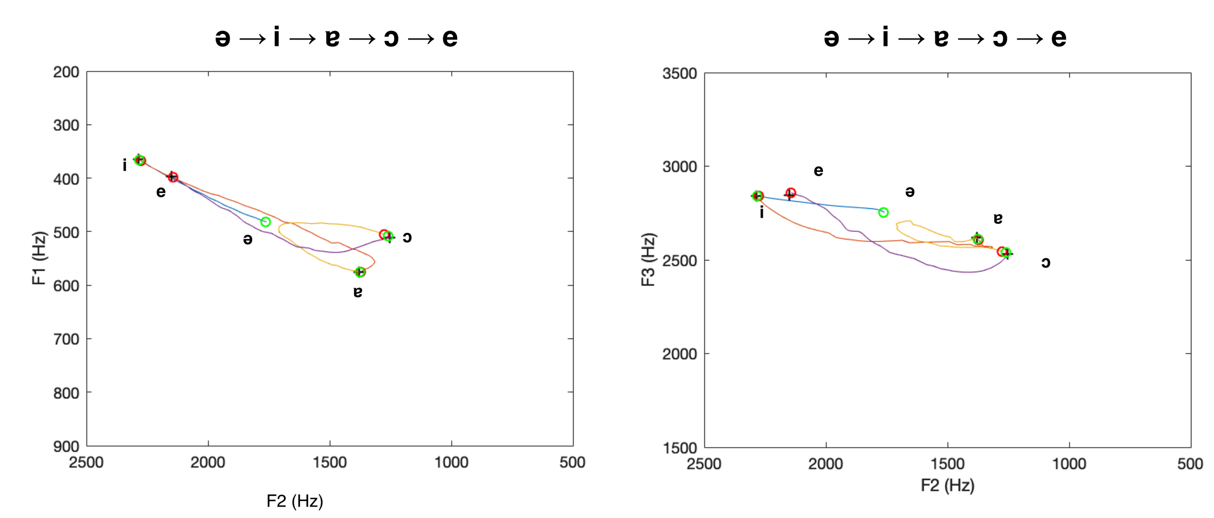
From /ə/ to /a/



From /ə/ to /ɔ/

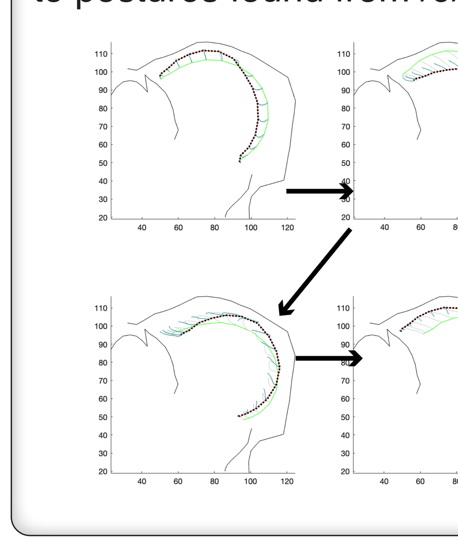


Vowel sequences (not center-out):

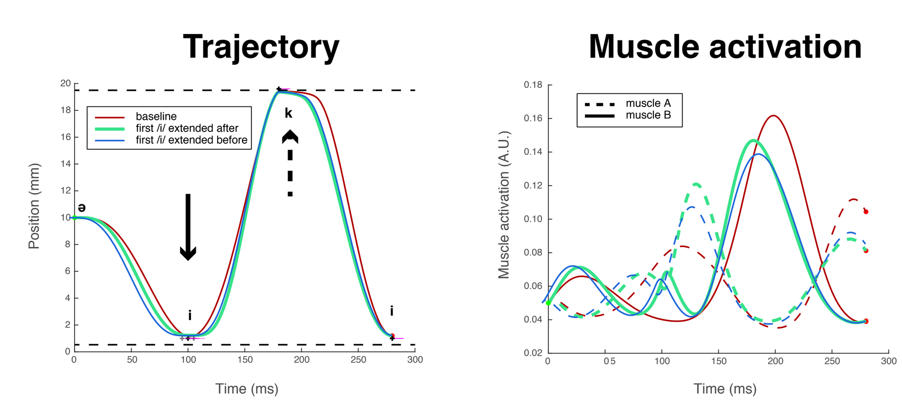


Caveat: difficulty in numerical optimization does not allow to assert found trajectories are a global optimum

- Effort-optimal trajectories to acoustic goals are far from straight in formant space
- Though goals are acoustic and starting point varied, final postures (red) seem very similar to postures found from /ə/

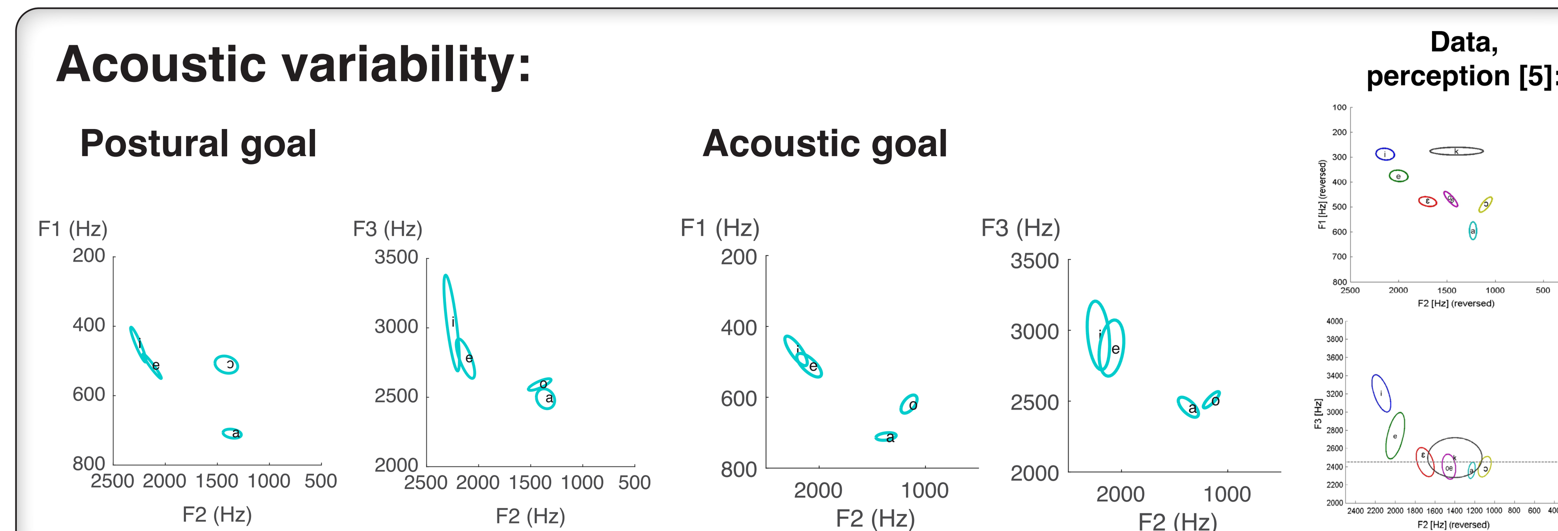


Control of contacts (VCV):



- Coarticulation emerges from effort optimization
- Delayed effects of earlier constraints
- (and large difference in optimization algorithms)

Acoustic variability:



CONCLUSIONS

- **Minimization of effort produces plausible tongue trajectories (kinematics, EMG)**
- Part of phonemic variability linked to aspects of sensorimotor control?
- Model predictions should be validated with formant tracking, EMA recordings and intramuscular EMG
- **Work in progress: optimization from FEM model**

REFERENCES

[1] Payan, Y., and Perrier, P. (1997). Synthesis of VV sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Comm* 22, 185–205.
[2] Badin, P., Elisei, F., Bailly, G. and Tarabalka, Y. (2008). An audiovisual talking head for augmented speech generation: models and animations based on a real speaker's articulatory data. In *Vth Conference on Articulated Motion and Deformable Objects*, pp. 132–143.
[3] Badin, P., and Fant, G. (1984). Notes on vocal tract computation. *STL-QPSR* 25, 53–108.
[4] Bryson A.E. (1999) *Dynamic optimization.* Addison-Wesley.
[5] Patri, J.-F., Diard, J., and Perrier, P. (2015). Optimal speech motor control and token-to-token variability: a Bayesian modeling approach. *Biol Cybern* 109, 611–626.