Tongue motor control: deriving articulator trajectories and muscle activation patterns from an optimization principle

Pierre Baraduc, Tsiky Rakotomalala, Pascal Perrier

Tongue motor control: deriving articulator trajectories and muscle activation patterns from an optimization principle

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

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

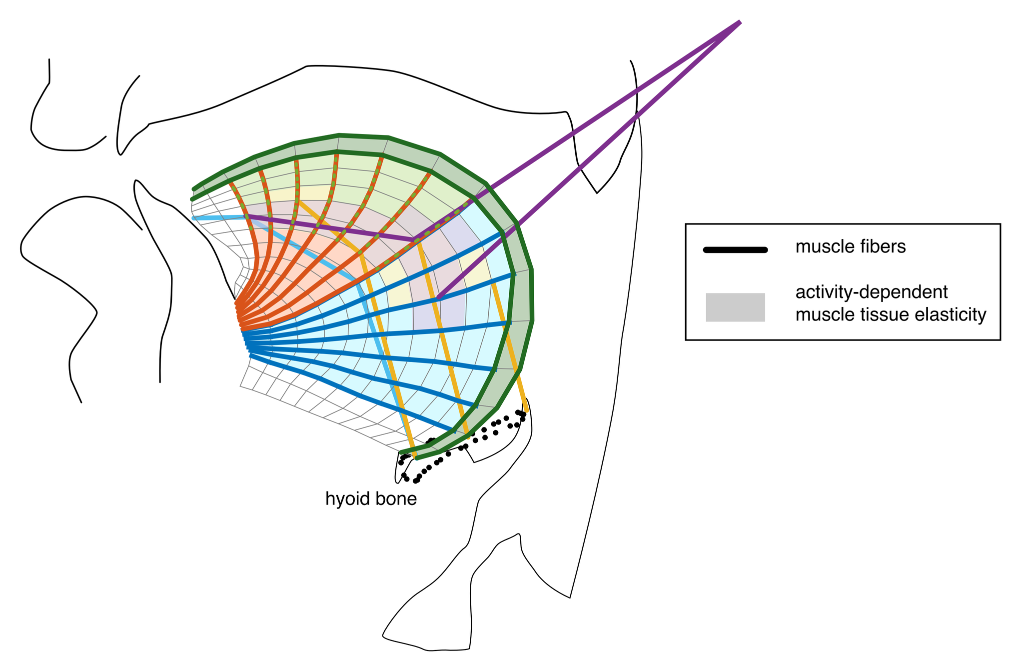
Speech production

- 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

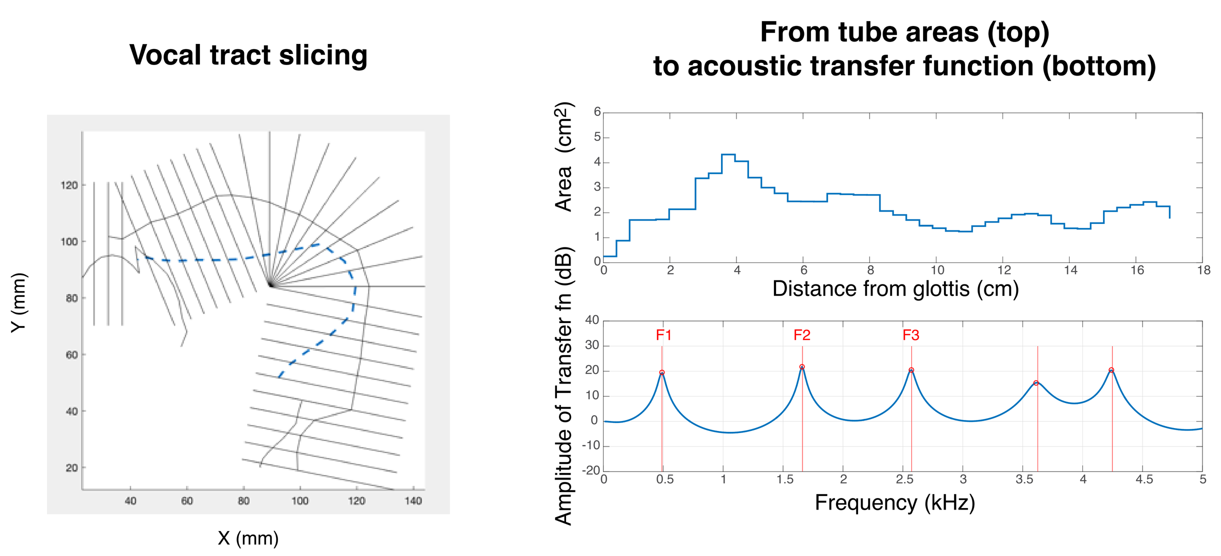
Tongue biomechanics:

- Finite element (FE) model of tongue deformation (sagittal 2D model)
- Seven muscles modelled: anterior genioglossus, posterior genioglossus, hyoglossus, styloglossus, verticalis, inferior longitudinalis, superior longitudinalis [1]
- Hill-type muscle model
- Activity-dependent tissue elasticity, small deformation approximation
- Fixed tongue floor
- Contacts with palate, velum and pharyngeal wall modeled as elastic interaction (wall: high stiffness)
- System continuous-time ODE solved with robust Runge-Kutta integration



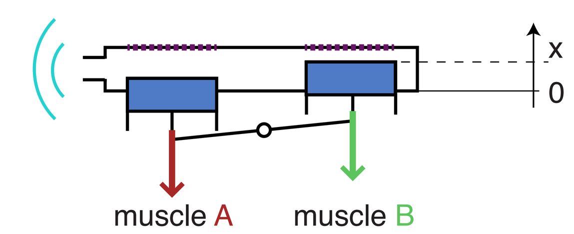
Vocal tract, from tongue shape to acoustics:

- 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 with an harmonic model following [3] after discretization of the tract in 44 tubes of identical length, and keep the first three formants.



Exploring the control of contacts:

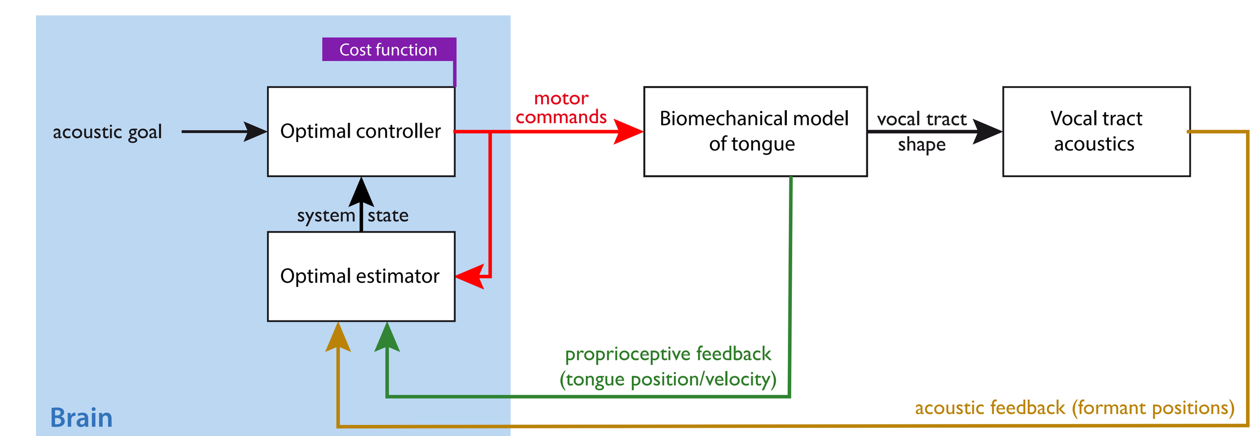
- One degree of freedom corresponding to principal component
- Tube model with auditory, proprioceptive and tactile feedback
- Muscular redundancy, inertia, elasticity towards neutral
- Intuition for more complex models, while convergence easier



Optimal control model:

Plant:

- Continuous time dynamics $\dot{s} = F(s(t), u(t)[+\varepsilon_u(t)])$ over fixed time T
- Proprioceptive feedback p modeled as 4D projection on principal axes of position and velocity, 3D acoustic feedback a (formant values)
- Acoustic F1-F2 goal
- Variability study:
  - linear reduced plant model identified over ~ 50,000 FEM simulations
  - only motor noise: additive (SD σ_A) and multiplicative (SD σ_M) Gaussian white noise on motor command
  - Internal state estimate via extended Kalman filtering



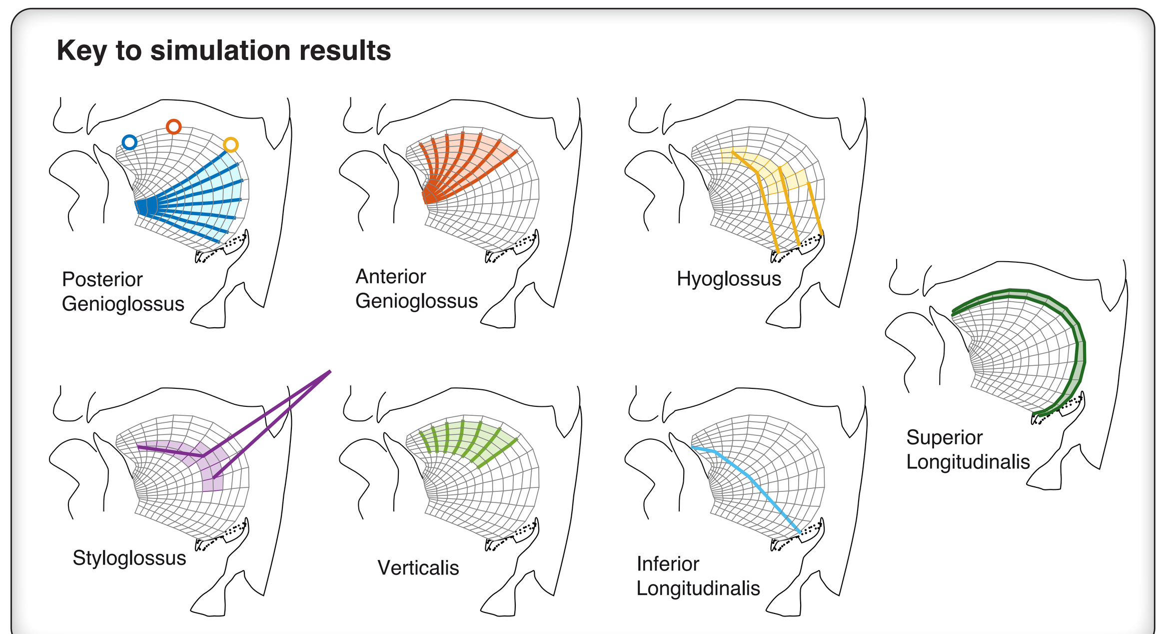
Optimization:

- Unconstrained optimization: cost function includes neuromotor effort and precision penalty:

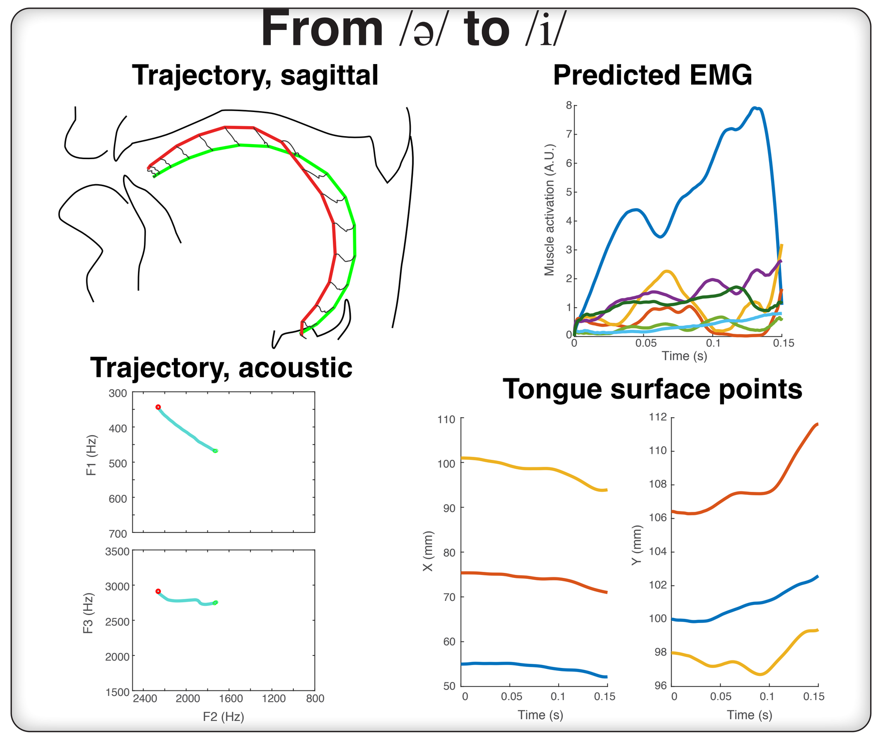
$$C = \int_0^T ||u(t)||^2(t)dt + \alpha||p(T) - p_{\text{goal}}||^2$$

- Indirect optimal control (Pontryagin based), gradient descent and/or Newton-Raphson method [4]
- Some checks of sensitivity to initial parameters

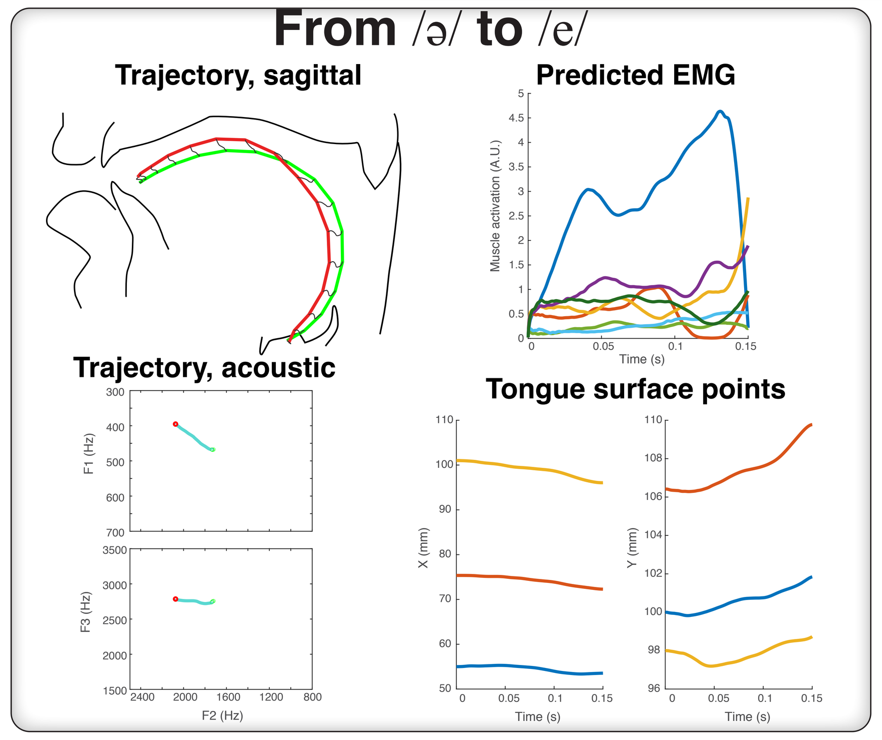
RESULTS



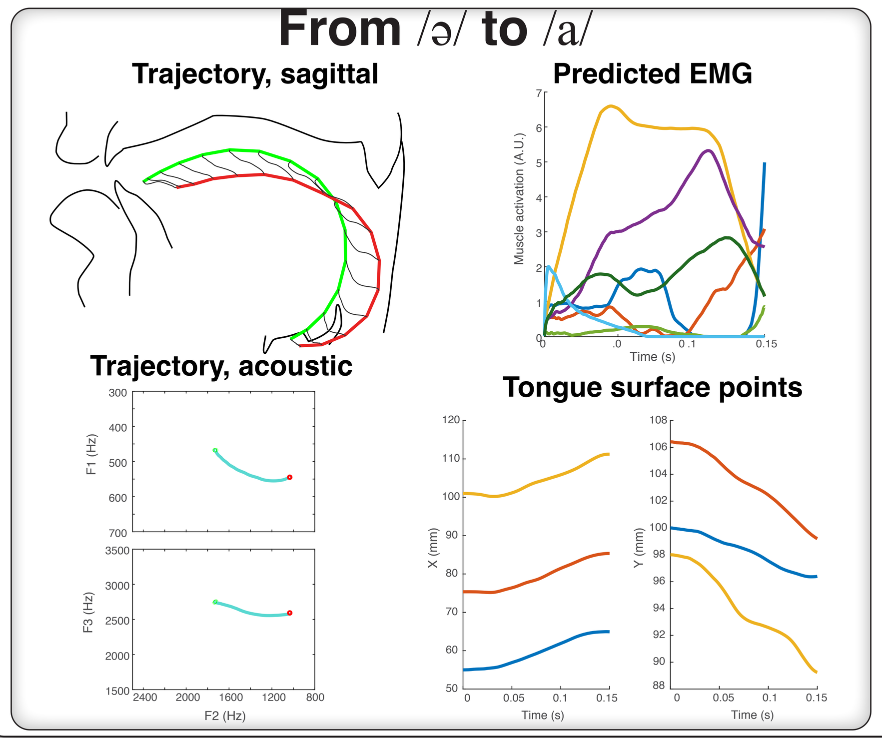
From /ə/ to /i/



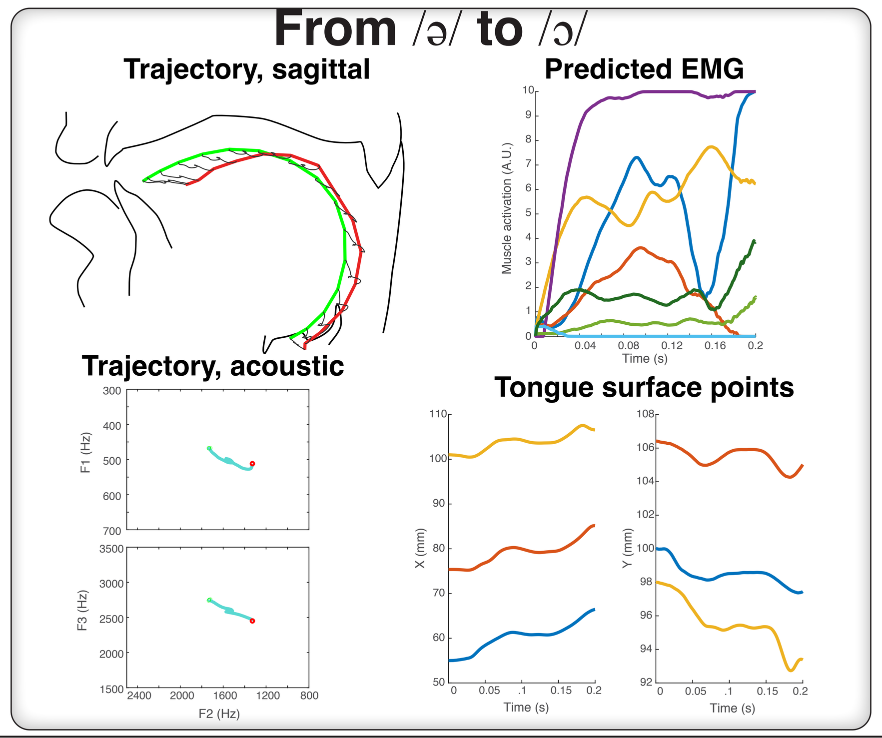
From /ə/ to /e/



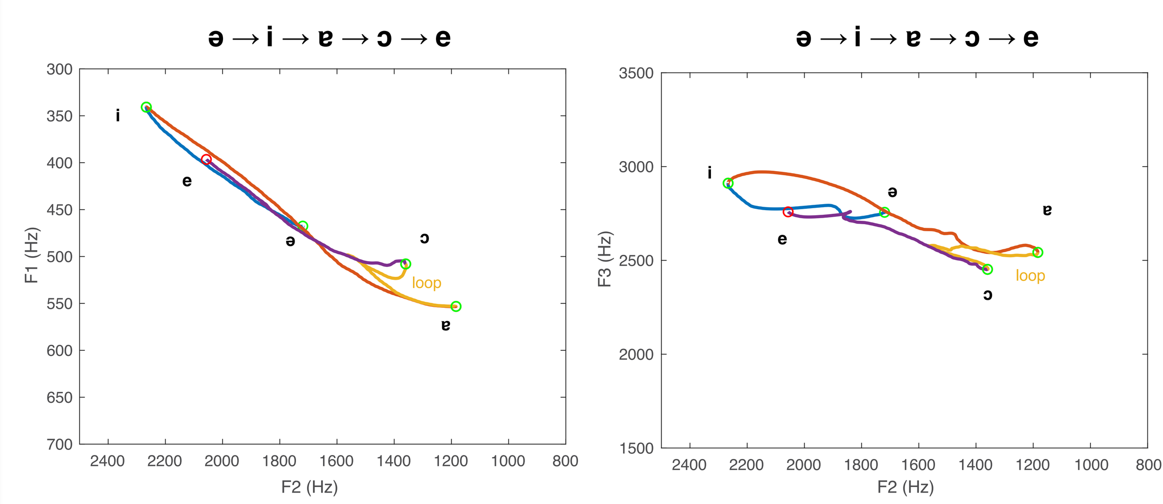
From /ə/ to /a/



From /ə/ to /ɔ/

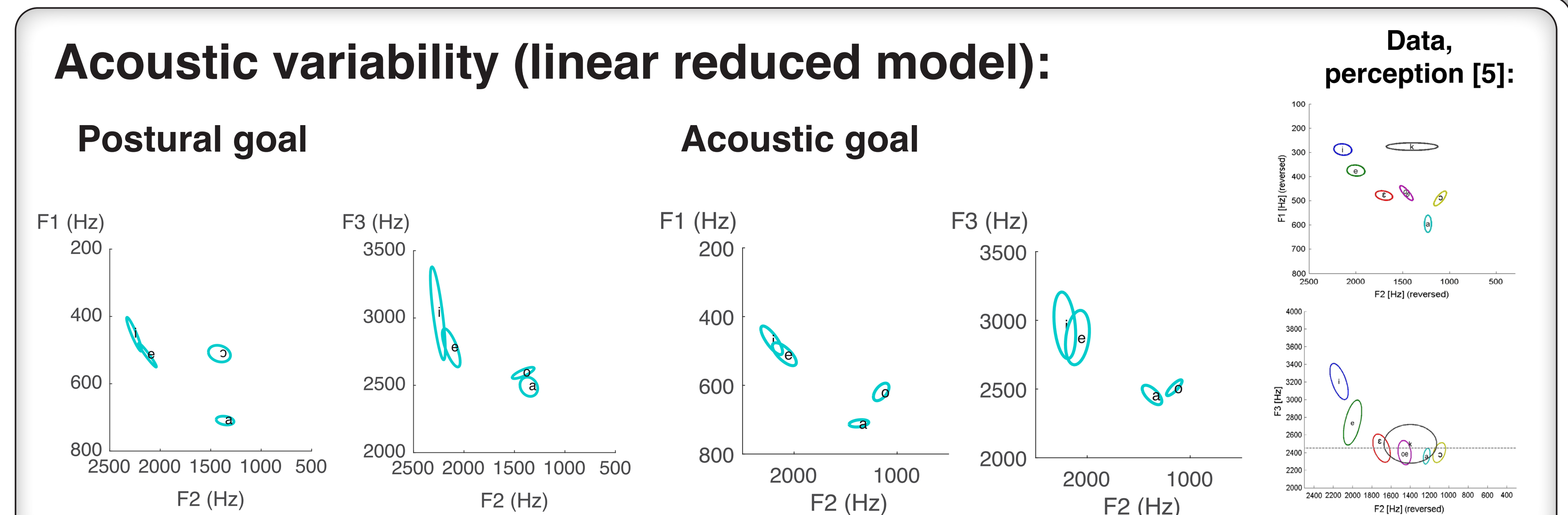


Vowel sequences (not center-out):

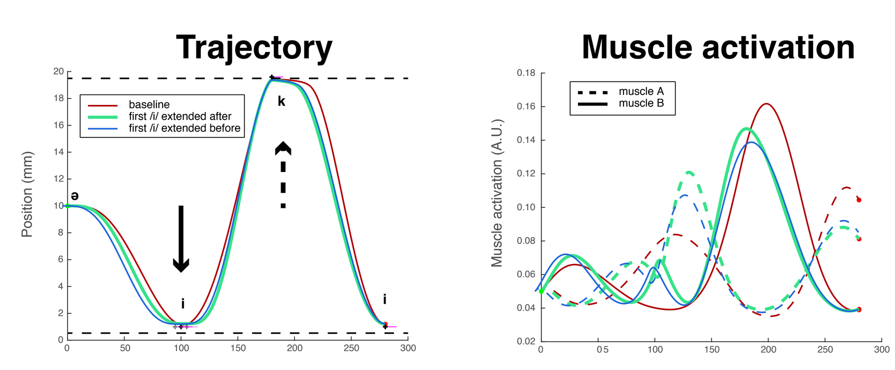


- Effort optimization leads to loopy trajectories in formant space
- Though goals are acoustic, intermediate postures (red) seem very similar to final postures from /ə/

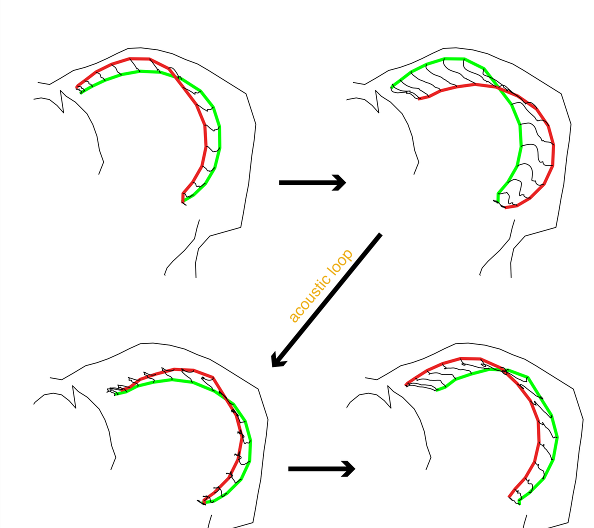
Acoustic variability (linear reduced model):



Control of contacts (VCV):



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



CONCLUSIONS

- **Minimization of effort produces plausible tongue trajectories (kinematics, EMG)**
- Part of phonemic variability linked to aspects of sensorimotor control?
- Toy model suggests coarticulation can be tackled by this method
- Model predictions should be validated with formant tracking, EMA recordings and intramuscular EMG

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.