A hierarchical state feedback control model for speech simulates task-specific responses to auditory and articulatory perturbations

Benjamin Parrell[1]*, Vikram Ramanarayanan[2]*, Srikantan Nagarajan[3], John Houde[3] (*equal contribution)

[1]University of Delaware; [2]Educational Testing Service R&D; [3]University of California, San Francisco

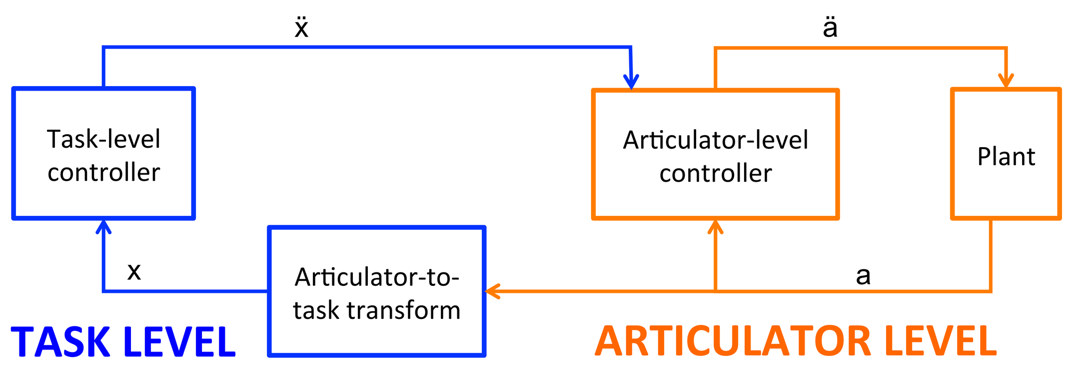
contact email: bparrell@wisc.edu

MATLAB code available on request

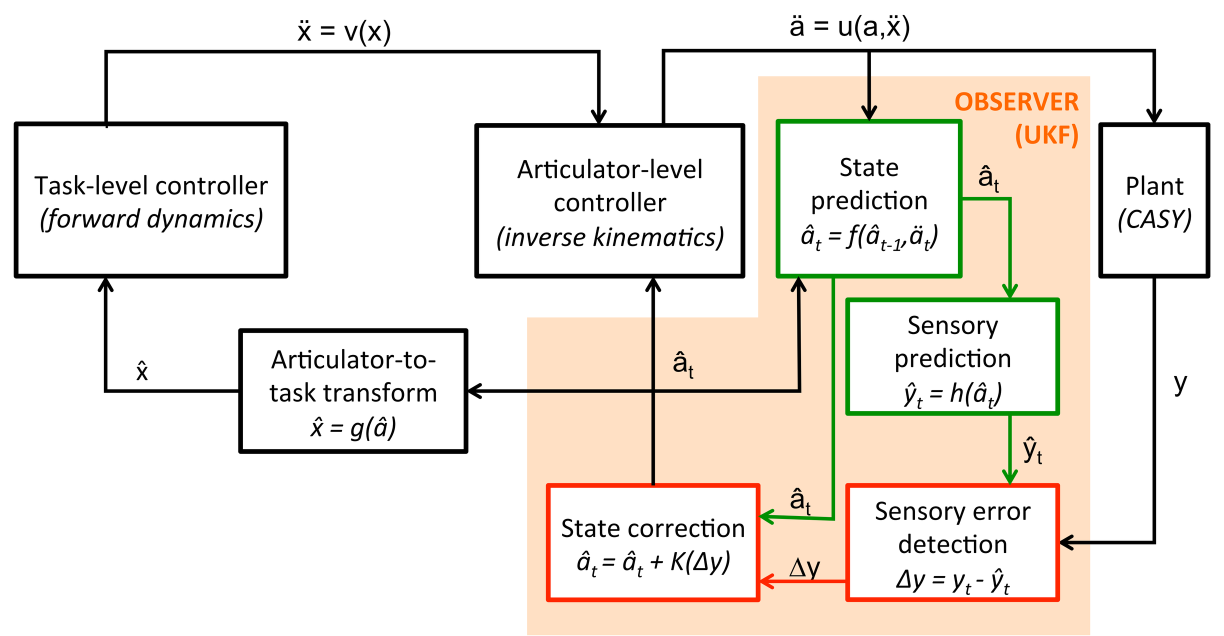
Model architecture

Goal: create a state feedback control model of speech motor control [1] that includes hierarchical control of both high-level tasks and low level articulation [2,3] and incorporates sensory feedback.

Basic hierarchical control



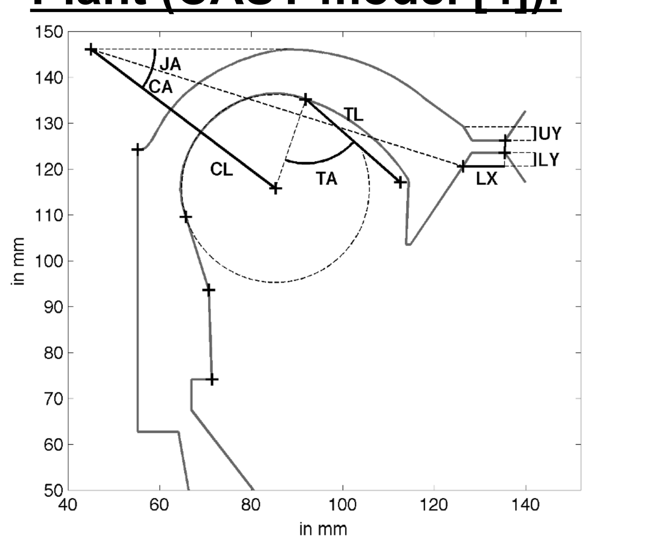
Current Model



Conventions:

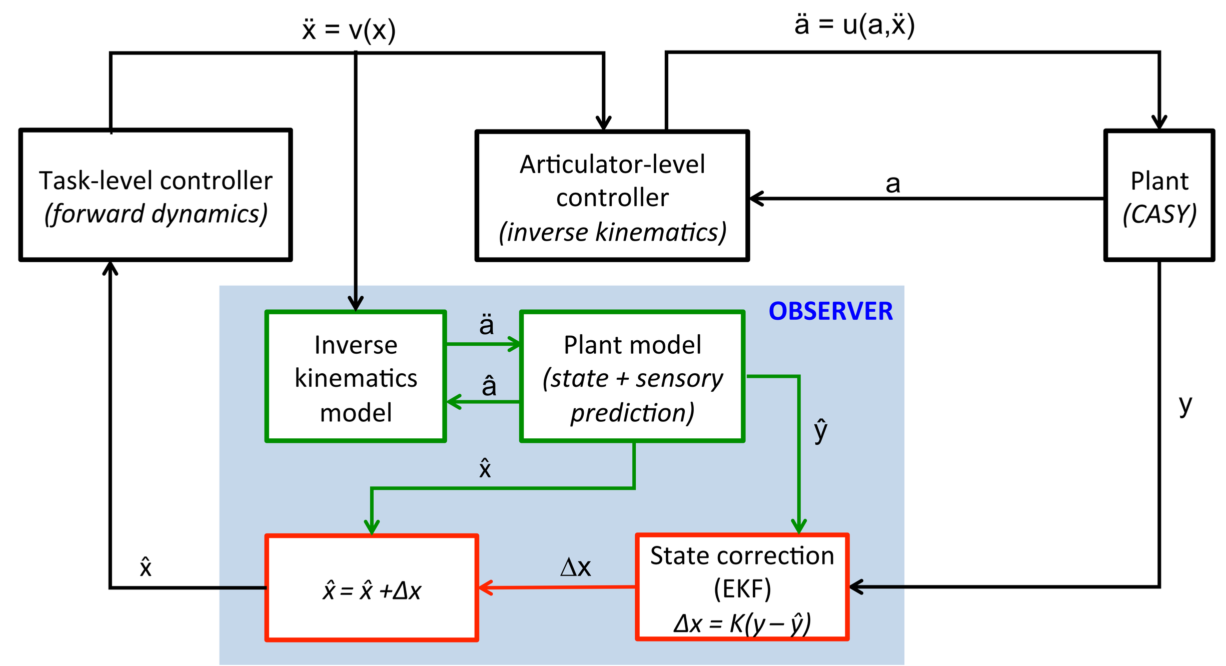
x: Task positions (x) and accelerations ($\dot{x}$)
$\ddot{x}$: Task accelerations
a: articulator positions (a) and accelerations ($\dot{a}$)
$\ddot{a}$: articulator accelerations (motor commands)
y: sensory information, includes both auditory + proprioceptive signals
^: prediction
→: predict
→: correct

Plant (CASY model [4]):



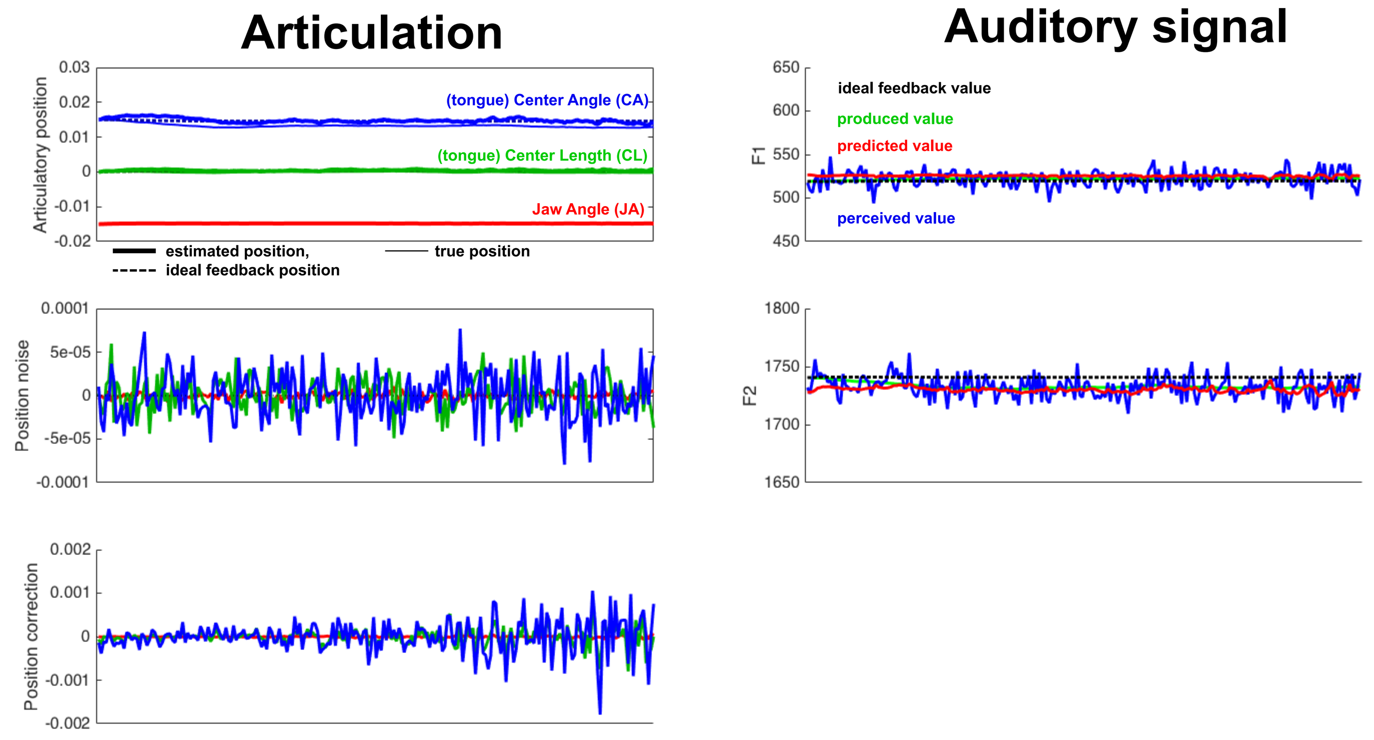
Note: we currently implement tasks as vocal tract constrictions [2,5], but similar architectures could be used for tasks in other spaces (e.g., auditory or somatosensory)

Previous Model [6]



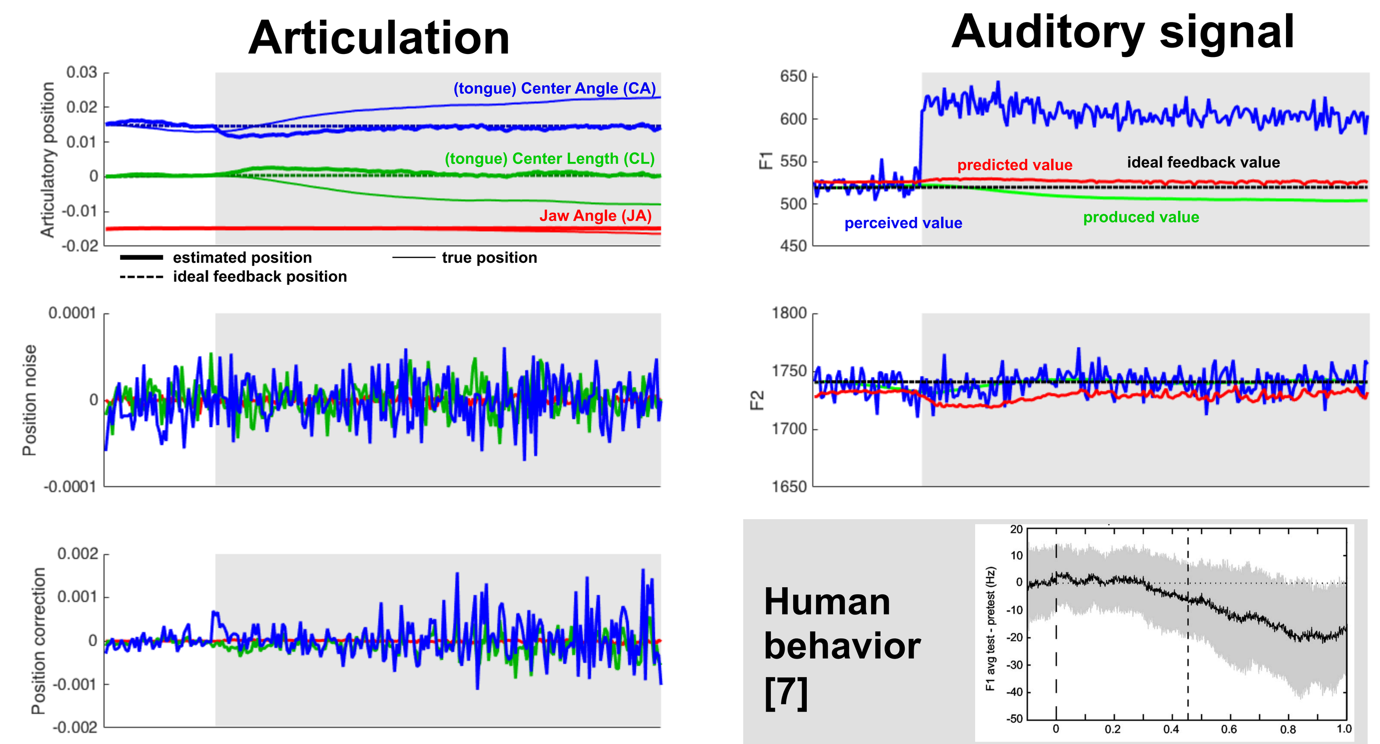
Simulation results

Model produces stable behavior in the presence of both motor and sensory noise



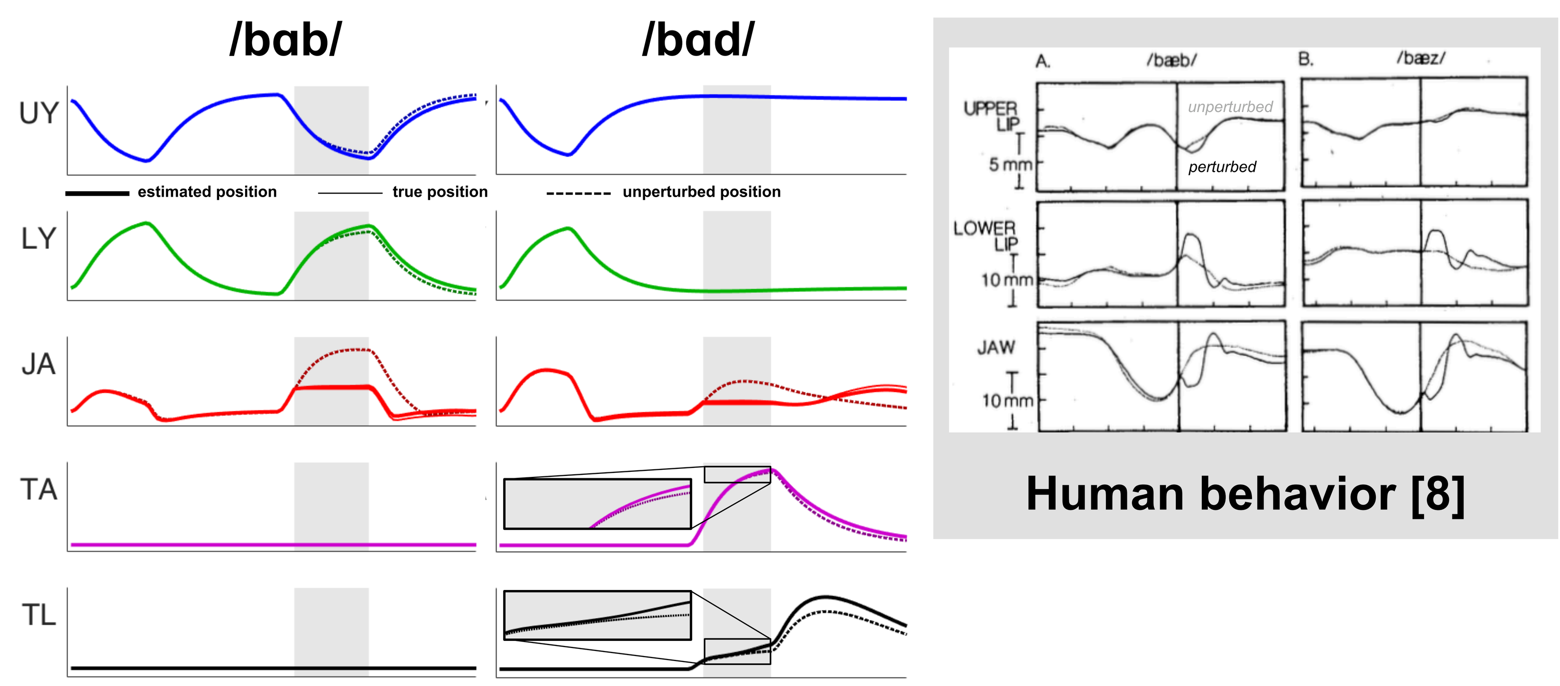
Compensation for auditory perturbations

Response to +100 Hz F1 perturbation



Compensation for articulatory perturbations

Response to downward jaw pull



Future directions

How to predict sensory and articulatory states?

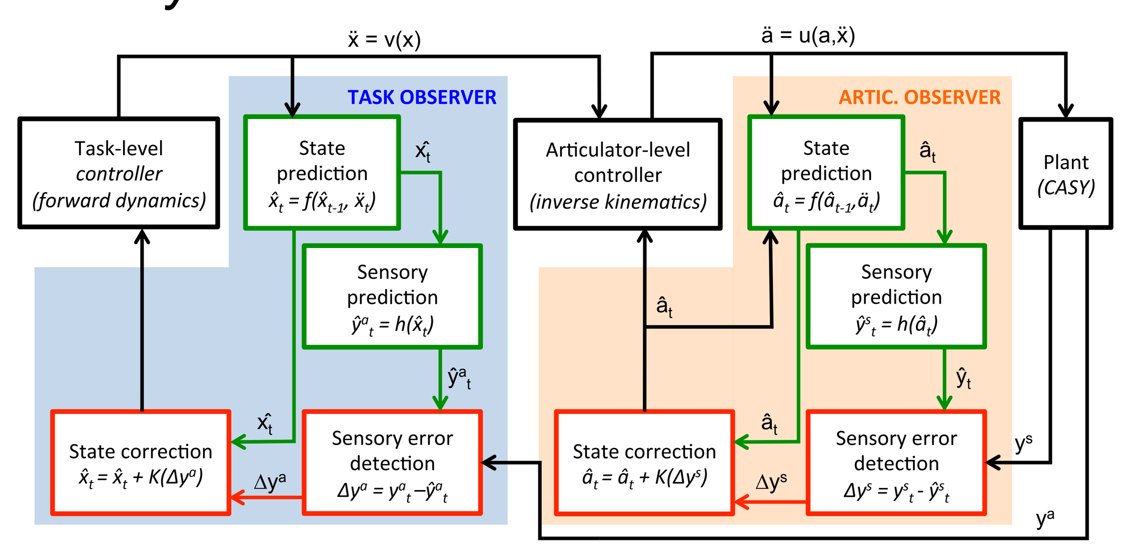
Currently, auditory predictions are learned via Locally Weighted Projection Regression [9] from a training set of noiseless vowel sweep data. Articulatory predictions are calculated from the differential equations that drive articulator motion in CASY.

How to estimate state from internal prediction and observation?

We have implemented versions with both an Extended Kalman Filter and an Unscented Kalman Filter. We are evaluating the behavior of both models. In either model, we are exploring if the Kalman gain (K) be fixed or free to vary.

How to integrate auditory and somatosensory feedback?

It is unclear whether auditory and somatosensory information are integrated into a single state estimate for low-level articulatory control as currently implemented. Alternatively, low-level control could rely on somatosensory feedback while auditory feedback is used for task-level control [10].



How to incorporate neural delays?

We are working to extend the state to include past time points, which has been used successfully to account for neural delays in non-speech models [11].

[1] Houde, J., & Nagarajan, S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, *5*, 82.
[2] Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecologoical Psychology*, *1*(4), 333-382.
[3] Todorov, E., Li, W., & Pan, X. (2005). From task parameters to motor synergies: A hierarchical framework for approximately-optimal control of redundant manipulators. *J Robot Syst*, *22*(11), 691-710.
[4] Iskarous, K, Goldstein, L., Whalen, D., Tiede, M., & Rubin, P. (2003). CASY: The Haskins configurable articulatory synthesizer. In *Proceedings of the 15th ICPhS* (pp. 185-188).
[5] Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. (2004). TADA: An enhanced, portable task dynamics model in MATLAB. *Journal of the Acoustical Society of America*, *115*, 2430
[6] Ramanarayanan, V., Parrell, B., Nagarajan, S., & Houde, J. (2016). A new model of speech motor control based on task dynamics and state feedback. In *Interspeech 2016.*
[7] Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(4), 2288-97
[8] Kelso, J., et al. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(6), 812-32.
[9] Mitrovic, D., Klanke, S., & Vijayakumar, S. (2010). Adaptive optimal feedback control with learned internal dynamics models. In *From motor learning to interaction learning in robots* (pp. 65-84). Springer.
[10] Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407-22.
[11] Izawa, J., & Shadmehr, R. (2008). On-line processing of uncertain information in visuomotor control. *The Journal of Neuroscience*, *28*(44), 11360-8.

Work supported by: NIH R01 DC013979, NIH R21 DC014525