Simulating adaptation in the FACTS model of speech motor control: current progress, problems, and potential paths forward

Benjamin Parrell1*, Vikram Ramanarayanan2,3*, Srikanth Nagarajan2, John Houde3 (*equal contribution)

1University of Wisconsin-Madison; 2Educational Testing Service R&D; 3University of California, San Francisco

Background

FACTS model
(Feedback Aware Control of Tasks in Speech)

- FACTS is a hierarchical control model which links the control of high-level speech tasks with lower-level control of articulation.
- FACTS builds on previous task- and feedback-based controllers (Task Dynamics, SFC).
- FACTS is able to replicate online behavioral responses to auditory and somatosensory perturbations of speech.
- Currently, FACTS does not include adaptive control to account for changes in behavior over time.

Modelling sensorimotor adaptation in a state feedback control model

- Adaptation is driven by sensory errors (as caused by, e.g., an external perturbation of vowel formants).
- Errors can update either the forward model or the control policy, or both.
- If errors update the forward model, this model must be used in planning future movements.
- If errors update the forward model, do they update the state prediction model, the sensory prediction model, or both?

Learning the state and sensory prediction models

- Configurable Articulatory Synthesizer (CASY) used as the vocal tract model for FACTS simulations.
- Auditory prediction requires both process and observation models.
- **Process model:** predict the next articulatory state from the current state and current motor command (x[t], u[t] → x[t+1]).
- **Observation model:** predict the current sensory state from the current articulatory state (x[t+1] → y[t+1]).
- Somatosensory prediction uses an identity function.
- **Training Data for learning models:** ~2900 sweeps of the CASY synthesizer covering different regions of the vocal tract.

Locally Weighted Projection Regression

- Point-to-point prediction.
- Learns a local receptive field mapping for different regions of the input-output space.
- Used for process and observation models.
- More interpretable relative to DNN-based models.
- As implemented, inaccurate for process model, which leads to model instability.

Recurrent Neural Networks (LSTMs)

- **Sequence to sequence prediction**
- Learns a nonlinear mapping of the input-output space.
- Black box model. Interpretability not straightforward.
- Investigated for process model because of instability in LWPR models.

Modeling adaptation in FACTS

Sensorimotor adaptation as changes to the sensory prediction model

- Updates to LWPR sensory prediction model on a trial-by-trial basis.
- After baseline trial (black), model exposed to 20 trials +100 Hz perturbation of F1(red→blue).
- A final washout trial (green) with no perturbation tests for adaptation.

Sensorimotor adaptation as changes to the state prediction model

- Updates to LWPR model do cause changes to behavior.
- Learning does not oppose perturbation.
- Model learns to predict auditory perturbation, leading to loss of compensatory response (as seen on trial 1).
- Potential for adaptation only if model could be used to optimize motor command.

Current problems and potential solutions

- **Sequence prediction:**
  - Model currently trained on whole-trial sequences.
  - Need to predict a single time point.
  - Lose desirable smoothing with single-point prediction.

- **Model accuracy:**
  - How to assess accuracy.
  - What is accurate enough?