The GODIVA model

Jason Bohland, PhD
Assistant Professor of Health Sciences and Speech, Language & Hearing Sciences
Faculty of Neuroscience
Boston University
QNL.BU.EDU

Sep 27 2012
jbohland@bu.edu
DIVA models speech sensorimotor control mechanisms

- Learning articulatory – auditory – somatosensory associations
- Learning sensorimotor chunks

DIVA

Guenther et al. (2006)
Linguistic performance errors
  obey structural constraints
  imply hierarchical organization
  Imply parallel planning

Chronometric data
  sequence length, familiarity effects
  imply parallel planning

Co-articulation
  carry-over and look-ahead in motor programming
  Imply parallel planning

Conceptual / theoretical non-biological models

from Levelt et al. (1999)
Competitive queuing architecture

Parallel “item and order” STM

Selection of next item for performance

Recordings from macaque F5 during serial shape drawing task

from Averbeck, Chafee, Crowe, and Georgopoulos (2002)
The GODIVA model

Extends DIVA to include explicit parallel representations of forthcoming utterances that interface with learned speech sensorimotor programs.
Main effects of sequence complexity

Syllable Complexity ($syl$)

<table>
<thead>
<tr>
<th>Sequence Complexity (seq)</th>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV-CV-CV ta-ta-ta</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC(C)CV-CC(C)V-CC(C)V stra-stra-stra</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV-CV-CV ka-ru-bi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC(C)CV-CC(C)V-CC(C)V kla-stri-splu</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Left inferior frontal sulcus / premotor
Right inferior Cb
Medial premotor regions
Anterior thalamus
Posterior parietal

aINS / FO junction
Main effects of syllable complexity
Sequence and syllable complexity strongly interact
e.g., the effect of sequence complexity is greater
when the individual syllables are complex
Phonetic invariance in IFS

Myers et al., 2009, Psychological Science
Phonological content queues

Left Inferior Frontal Sulcus

Columns categorically encode phonemes (or phoneme-like items) at a given position in the syllable.

Language specific phoneme set serves as a basis set for representing arbitrary speech sequences.

Input to specific slots can be learned (retrieval imposes a set of gradients), or can be the result of a parsing process (e.g. when stimuli are read or heard).

Use of positional representation effectively reduces the # of competing items in the queue (improves SNR).

The model uses 53 phonemes from the CELEX database (Baayen et al., 1995).

Shunting equations (ala Grossberg) provide automatic gain control

Linear feedback allows multiple inputs to remain active – self-normalization

Faster than linear feedback provides *winner-take-all* dynamics
Cells that code for entire syllable frames are planned as a CQ.

Competitive choice is gated by offset of activity in IFS choice region.

Choice of a syllable frame (g) initiates a serial chain of cells representing the constituent abstract positions that compose that frame.

8 syllable frames account for over 96% of syllable productions (Baayen et al., 1995).

Hypothesis: we can extract the patterns of regularity to enable abstract frame-like representations.
Planning loop through basal ganglia is used to enable competition within IFS zones
- Similar in spirit to action selection models (e.g., Mink and Thach, 1993)
- Sequence through syllable slots AND enable competition in the SSM
GODIVA posits new structure in DIVA’s Speech Sound Map (i.e., mental syllabary)

A set of phonemes chosen in IFS activates phonologically matching SSM representations

Gradient of activity represents degree of match

GO signal from SMA through BG/thalamus and (anticipatory) completion signal enables selection of next winning program
Complementary activity gradients are loaded into the IFS plan and pre-SMA plan layers.

The most active syllable frame (for the next syllable type) becomes active in the pre-SMA choice layer.

The corresponding cell in pre-SMA plan layer is suppressed.

The active pre-SMA choice cell initiates serial readout of a chain of cells corresponding to its abstract positions.

The active positional cell activates a BG loop channel, disinhibiting a thalamic cell, and enabling the appropriate positional zone in the IFS choice layer.

The most active phoneme in the IFS plan layer for this positional zone becomes active in the IFS choice field.

The corresponding cell in IFS plan layer is suppressed.

Simultaneous activation of one phoneme for each syllable position in the IFS choice layer activates potential sensorimotor program "matches" in the SSM plan layer.

Best matching program is activated in SSM choice layer.

The active program's constituent phonemes are suppressed in the IFS choice layer.

Motor plan cells are activated in the left motor cortex.

Overt production can be initiated for the currently active SSM choice cell (with SMA trigger).

Completion signal transiently suppresses SSM choice layer, allowing a new winner to be chosen.
GODIVA modules

Hypothesizes divisions of planning cells and output cells up through motor cortex

SMA / BG motor loop responsible for release / timing of sensorimotor programs
“go di va”: Simulation 1

(A) IFS plan cells, Zone 3

(B) IFS plan cells, Zone 4

(C) IFS choice cells, Zone 3

(D) IFS choice cells, Zone 4

(E) SSM plan cells

(F) SSM choice cells
“go di va”: Simulation 2
Errors occur naturally through addition of noise in CQ model planning layers.

But currently only the simplest of errors can be modeled.

Clusters are a major problem.

Q: Can syllable CV structures also slip?
Constraining the model

Neural models solve inverse problems

Many models can be built to yield the same results, matching the same data – but are they plausible?

The approach must be to add **constraints**

1. Better informed representations
   - Multi-voxel pattern analysis?
2. Anatomical and functional connections
   - Analysis of resting state and task driven fMRI datasets
3. Differences in the brains of healthy speakers and speakers with communication disorders
   - Better databasing efforts from the clinical world
Future directions

Connecting abstract *perceptual* representations (of items and order) with GODIVA’s *speech planning representations*

- Cf. Hickok 2012
Connecting abstract *perceptual* representations (of items and order) with GODIVA’s *speech planning representations*

- Cf. Hickok 2012

Further clarify (or modify?) the nature of phonological representations, specifically:

- Nature of consonant cluster representations
- Encoding phonotactic probabilities
- Abstract frames in perception and production

Clarify learning processes that can result in extracting phonological units from “pure motoric” units

Incorporate evidence from experimental data sets (MVPA, functional connectivity)
Acknowledgments

Frank Guenther
Dan Bullock
Joe Perkell
Oren Civier
Jason Tourville
Satra Ghosh