Coupled Oscillator Model of Speech Production Planning

Louis Goldstein

Haskins Laboratories

and

University of Southern California

Acknowledgments

- Experimental Work
  - Dani Byrd
  - Ioana Chitoran
  - Tine Mooshammer
  - Shri Narayanan
  - Marianne Pouplier
  - Mark Tiede
- Model
  - Hosung Nam
  - Elliot Saltzman

Catherine P. Browman

1945-2008

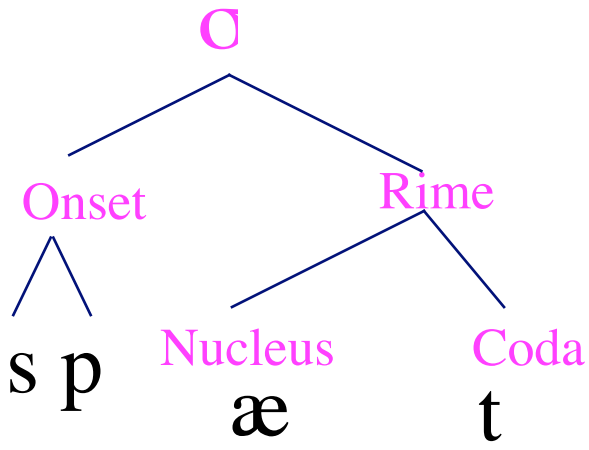
Timing in speech production

- How is timing among speech production events regulated to provide
  - stability
  - flexibility

- Coupled Oscillator model
  - provides one solution to this problem
  - leads to a grounded theory of syllable structure

Syllable Structure

- Basis for generalizations
  - Universality:
    CV syllables only universal type
  - Combinatoriality:
    - Onsets & rime combine relatively free
    - Combinations within onset and rime can be more restricted
  - Acquisition: CV structures acquired earlier than VC
  - Planning Time: CV structure planned faster than VC
- Structural differentiation
  - between similar sequences in different languages that are syllabified differently.



Grounding of syllable structure?

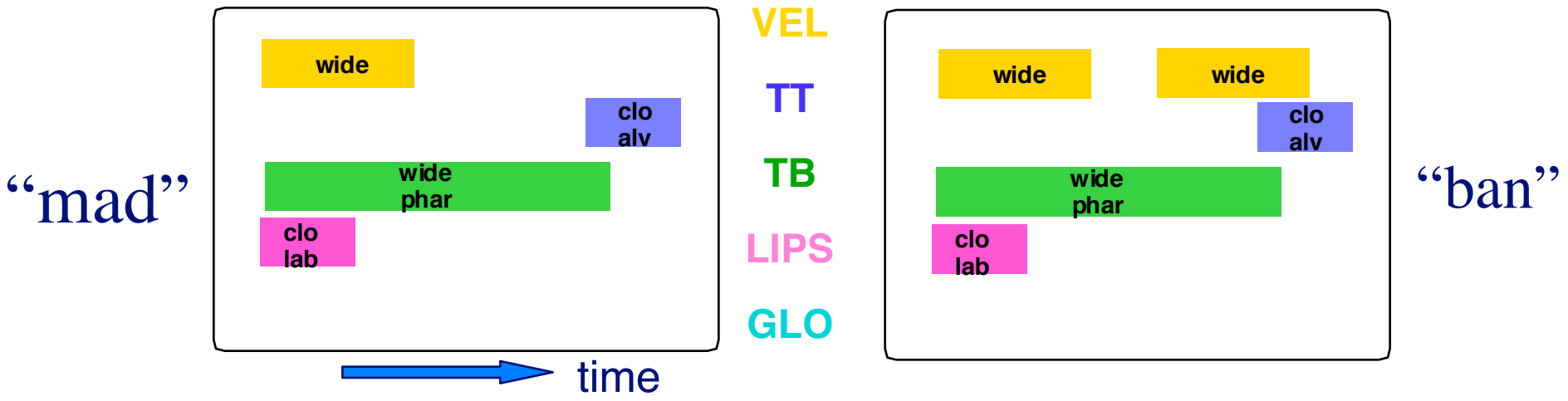
- Frame-Content Theory (MacNeilage, 2000)
  - Syllable frame develops (phylogentically, ontogenetically) from cycles of mandibular oscillation.
  - Syllable content (Cs, Vs) develops later
  - However..
    - Not clear how this theory can provide an account of these syllable-based generalizations.

Articulatory Phonology

- Phonology as combinatorial system
  - discrete, particulate units
  - glue that holds the units together in combinations.
- Articulatory Phonology
  - Units = Gestures
    speech production actions performed by one of the constricting "devices" of the vocal tract.
  - Glue =
    Coupling of *planning oscillators* associated with gestures that holds them together in stable temporal patterns

Controlling the timing of speech events

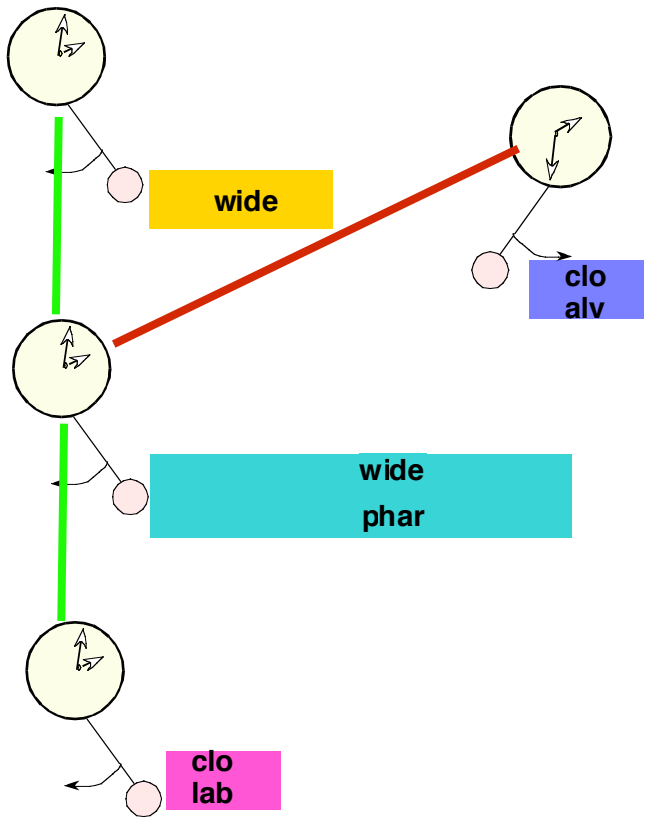
- Word forms as molecules composed of multiple gestures.
- Relative timing of gesture activation is significant information as can be shown in a gestural score.



How is appropriate relative timing maintained?
What is the glue?

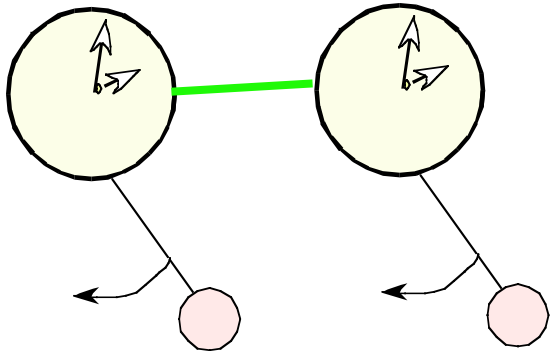
Model of inter-gesture timing during speech production

- Each gesture is associated with a planning oscillator, or clock, responsible for triggering that gesture's activation.
- Relative phase of oscillators (and therefore time of triggering) is controlled by coupling the clocks to one another.



Why coupling?

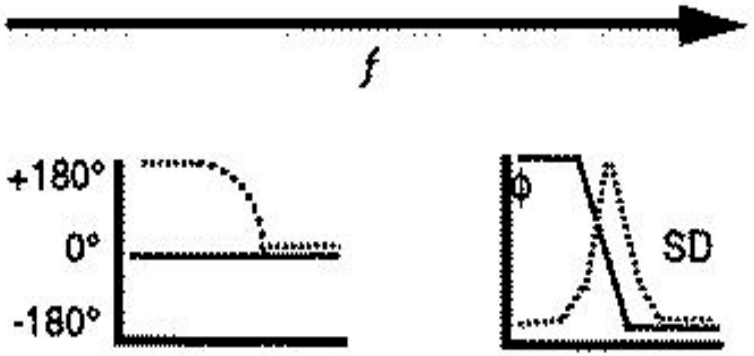
- Coupled oscillators exhibit entrainment. They synchronize with one another.



- Systems of coupled oscillators exhibit distinct modes of synchronization.
- These modes have been shown to underlie the timing of movements of multiple limbs in human action. (e.g.,Turvey, 1990).
- The same modes can be used to coordinate speech actions and form the basis of syllable structure.

Synchronization modes for limb coordination: phase-locking

- Two relative phase modes are *spontaneously* available (require no learning)
  Haken, Kelso & Bunz, 1985
  - 0° (in phase) most stable
  - 180° (anti-phase)
- Other phase locks can be learned (with difficulty).
- Abrupt transitions to most stable mode (0°) as frequency increases
  (Haken, Kelso & Bunz, 1985)

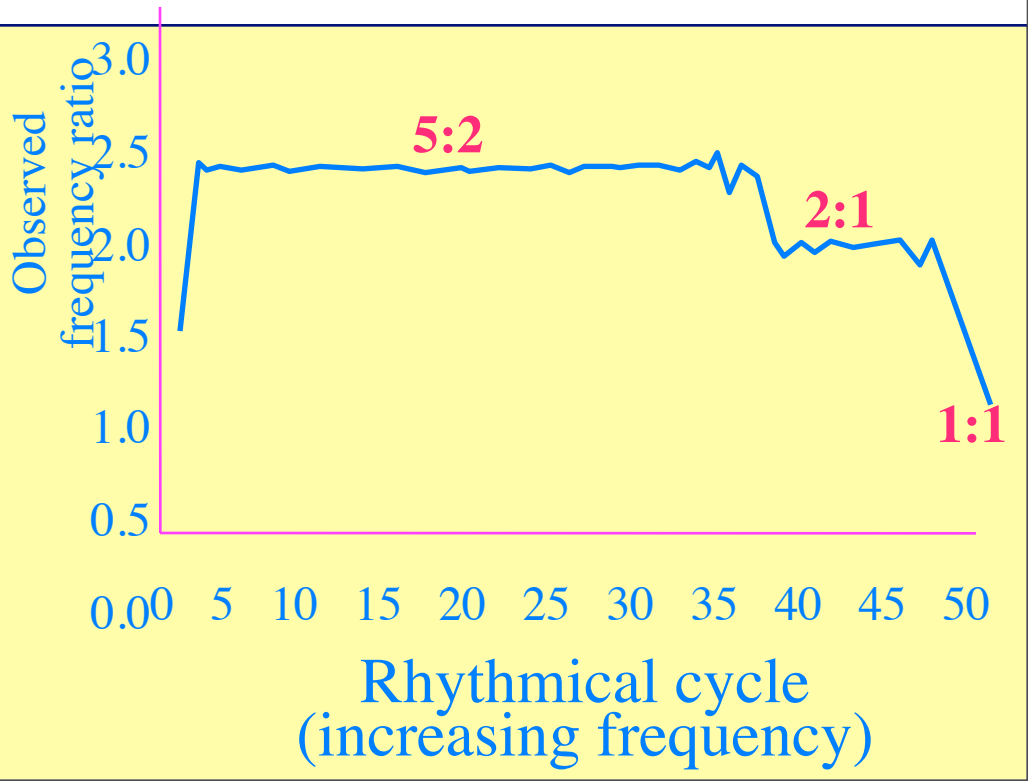


Turvey, 1990

Synchronization modes: Frequency-locking

- When performing oscillatory motions of multiple body parts
  - Simple rhythms (e.g., 1:1, 2:1) can be performed without learning or practice.
  - Complex multi-frequency rhythms (e.g., 4:3, 5:2) can be learned.
- Spontaneous transitions are observed from complex to simpler modes.

Skilled drummers begin by bimanual tapping with a 5:2 frequency-locking and gradually increase movement frequency.



Haken, H., Peper, C. L., Beek, P. J., & Dafferthofer, A. (1996).

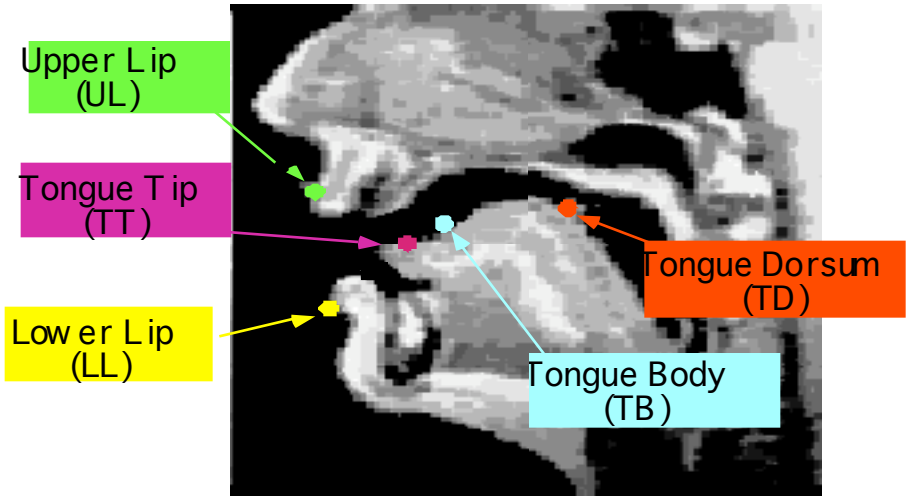
Evidence for modes in speech: Speech Errors (w/Pouplier)

- Speech errors have been characterized as errors in sequencing of segmental units.
  - The most frequent sublexical units involved in errors are single segments (Shattuck-Hufnagel, 1983).
  - Errors are thought to arise during planning when segments are inserted into the wrong positions within prosodic frames.
  - Utterances thought to be well-formed.
- Observing kinematics of speech errors leads to a very different picture:
  - Errors can be analyzed as shifts to more stable modes of coupled oscillators.

Experimental Technique

(Goldstein, Pouplier et al. 2007)

- Use of EMMA to provide movement data

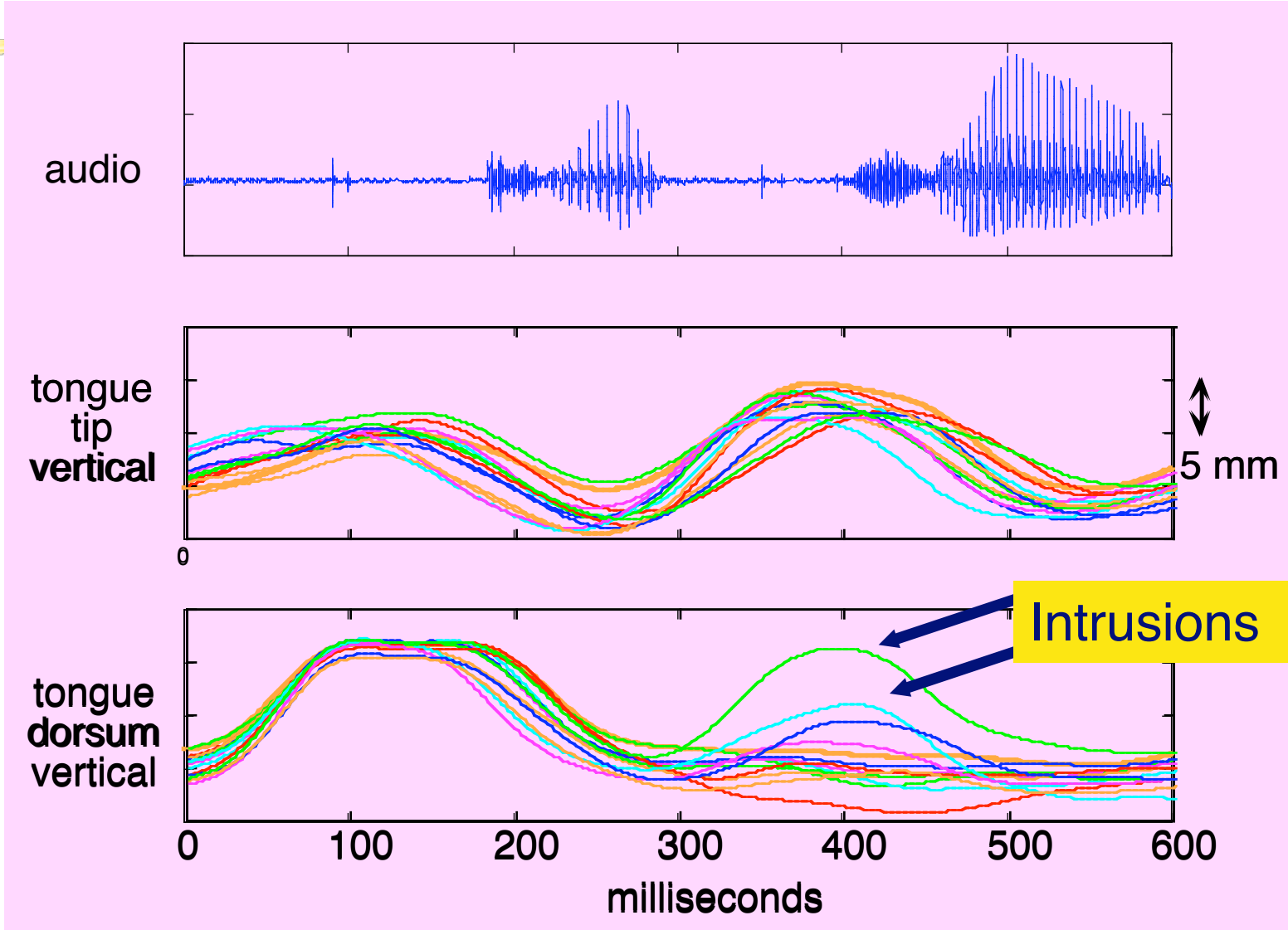


- Error Elicitation: Repetition of phrases with "alternating" consonants
  - cop top, kip tip
    bad bang,
  - 10-15 seconds, synchronized to metronome
  - Variations in rate, stress, order
  - Non-alternating controls
    - cop cop, top top

Main finding:

- An “extra” copy of a gesture appears to be activated at an inappropriate time.
  - e.g., in “cop top,” tongue dorsum ([k]-like) raising gesture during the [t].
  - Such errors can be statistically identified based on properties of control utterances.
- Errors vary continuously in magnitude and have different perceptual consequences (Pouplier & Goldstein, 2005)
  - Small movements are perceived as “normal” (e.g. /t/).
  - Large movements perceived as substitution errors (e.g. => /k/)

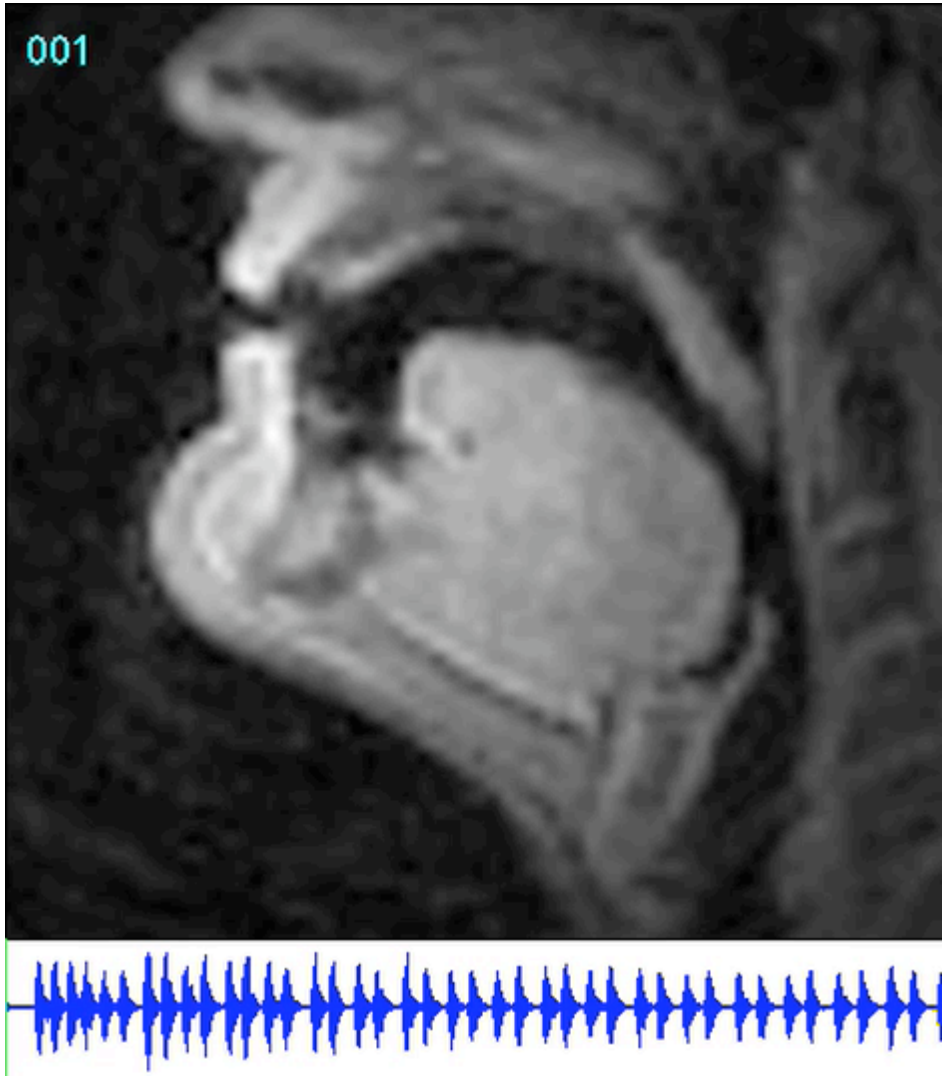
Intrusion Errors: “cop top”

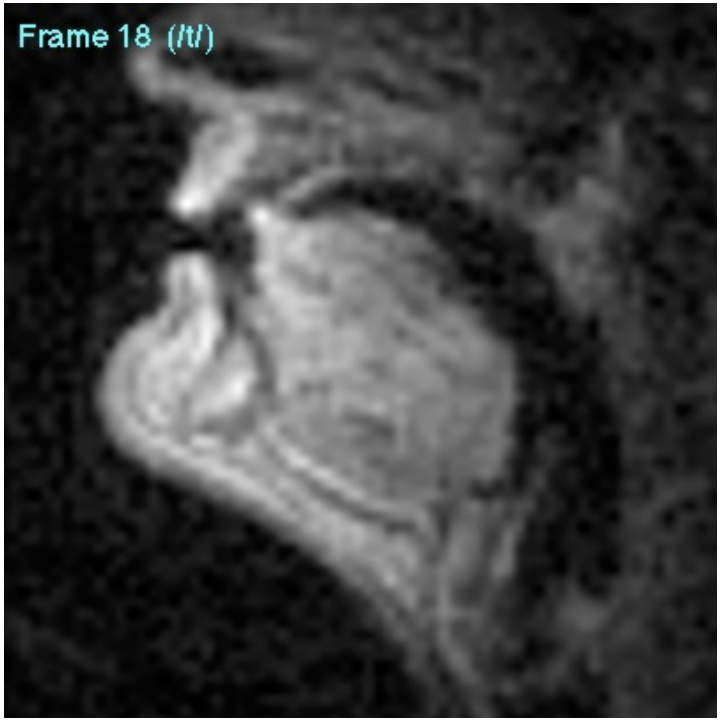


Intrusion vs. Reduction Errors

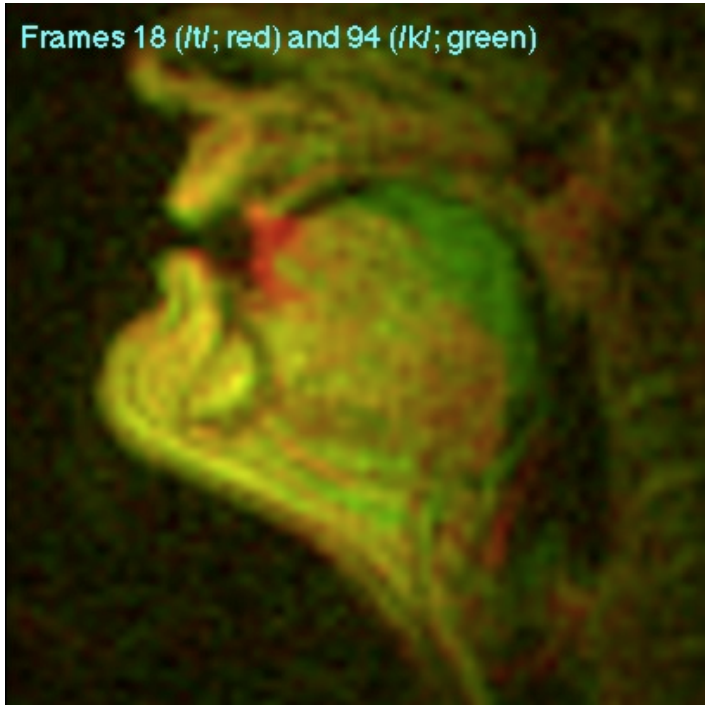
- Intrusion is usually not accompanied by reduction of the target gesture
  (e.g. tongue tip gesture during [t] of “top”).
- Errors of reduction do occur, *much* less frequently.
- System often appears to be producing two gestures concurrently.

Real-time MRI of apraxic (SPAN USC: Byrd, Narayanan)



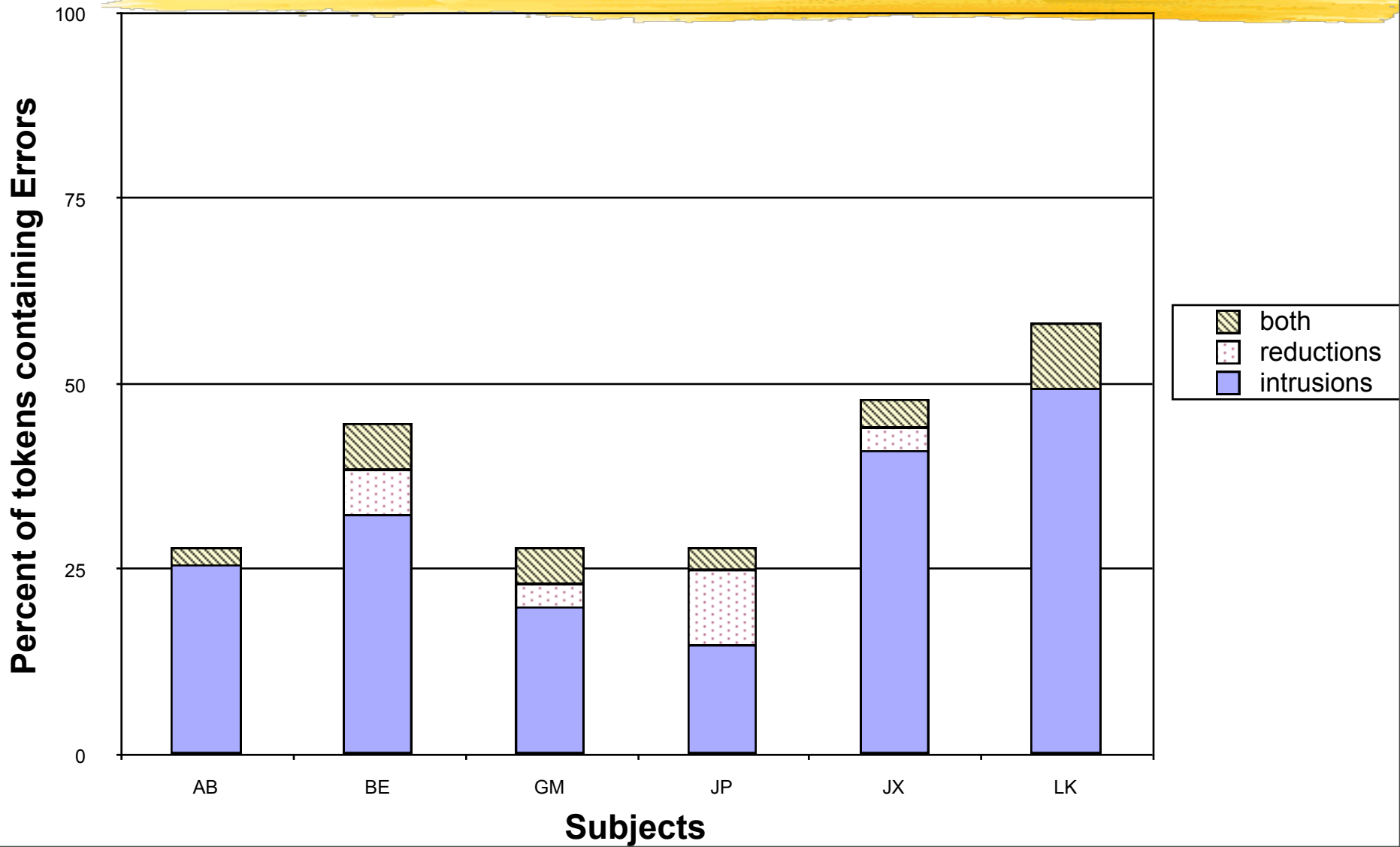

Frame 18 (/t/)


Frame 94 (/k/)


Frames 18 (/t/; red) and 94 (/k/; green)


Frame 556 (/tk/)

Dominance of Intrusion



How do intrusion errors arise?

- Intrusion errors are systematic, but cannot arise from mis-sequencing well-formed segments within a well-formed plan.
  - More than one unit is being produced concurrently.
  - Errors can be partial in magnitude.
- Intrusions can be explained if speech production planning engages coupling of oscillators.
  - Some modes of coupling are more stable than others.

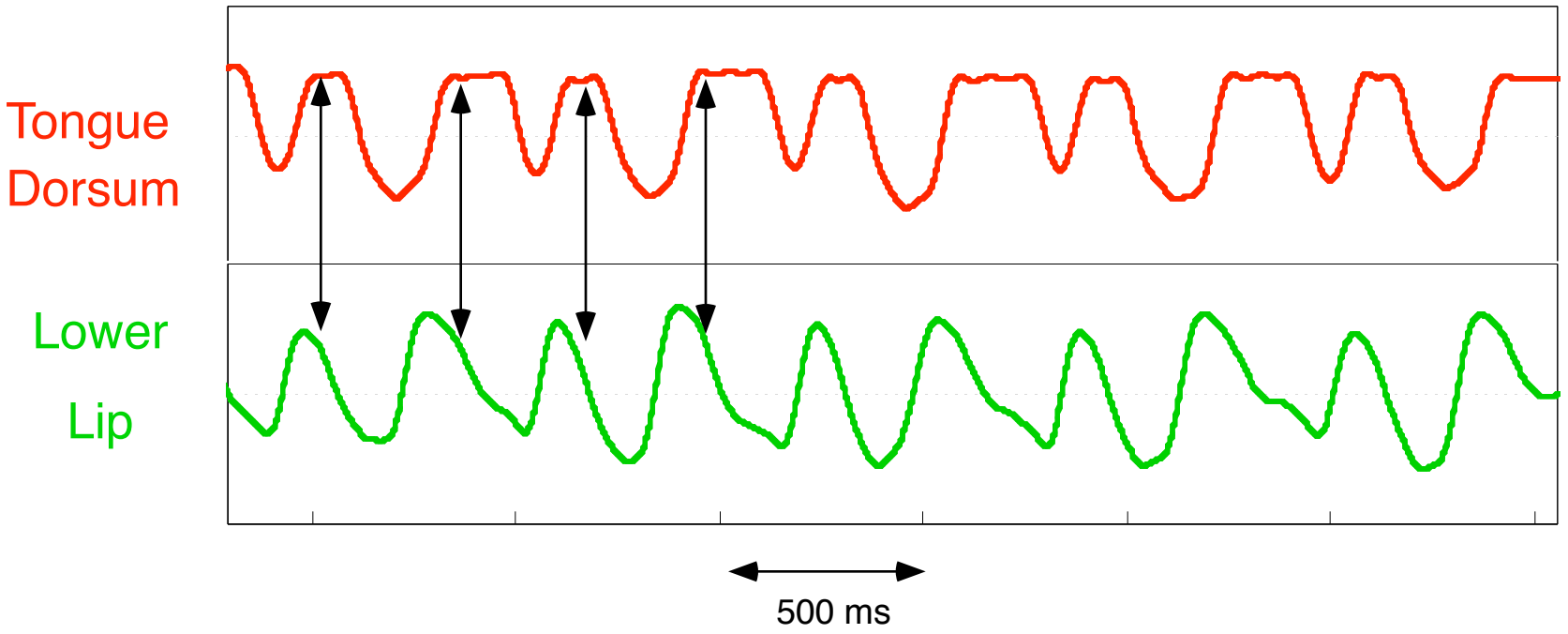
Intrusion errors and frequency- locking modes

Transition to a simpler (intrinsic) mode of frequency-locking can account for concurrent production of TD and TT gestures (as in "cop top").

- Before transition, both TT gesture (for [t]) and TD gesture (for [k]) are in 1:2 relation with LIP gesture for final [p] (and gesture for vowel).
- After transition, all gestures are in 1:1 relation.

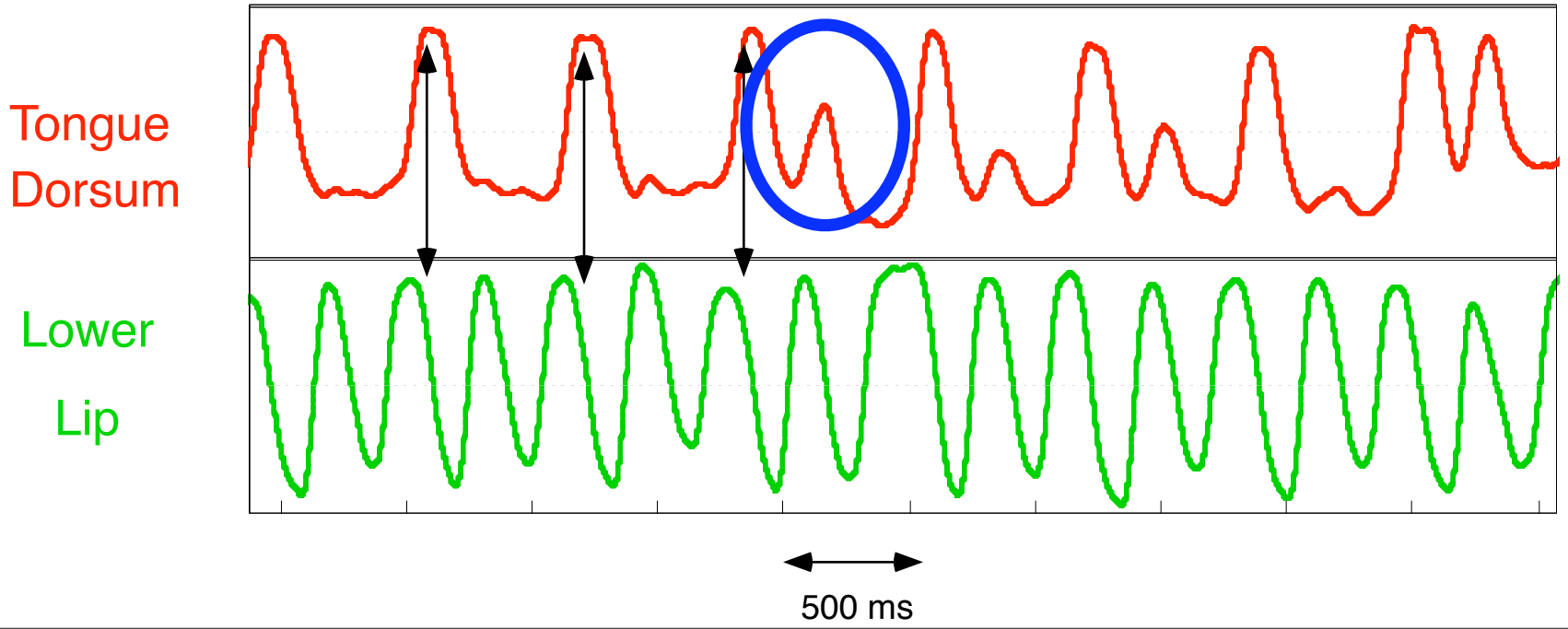
“cop cop”

- TD and Lips are frequency-locked in 1:1 relation.



“cop top”

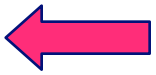
- TD and Lips are frequency-locked in 1:2 relation.
- TD intrusion error can be viewed as a spontaneous transition to a simpler mode of frequency-locking (1:1) between TD and Lip constrictors.



Non-repetitive tasks

- Can intrusion errors occur without overt repetition?
  - Would provide evidence that such errors occur as part of planning dynamics.
  - Parallel results using SLIP technique (Pouplier, 2007).
    - Subjects produce a phrase once after silently reading phrases that induce an error.

gore bail
goat ball
gone boat
bud goad



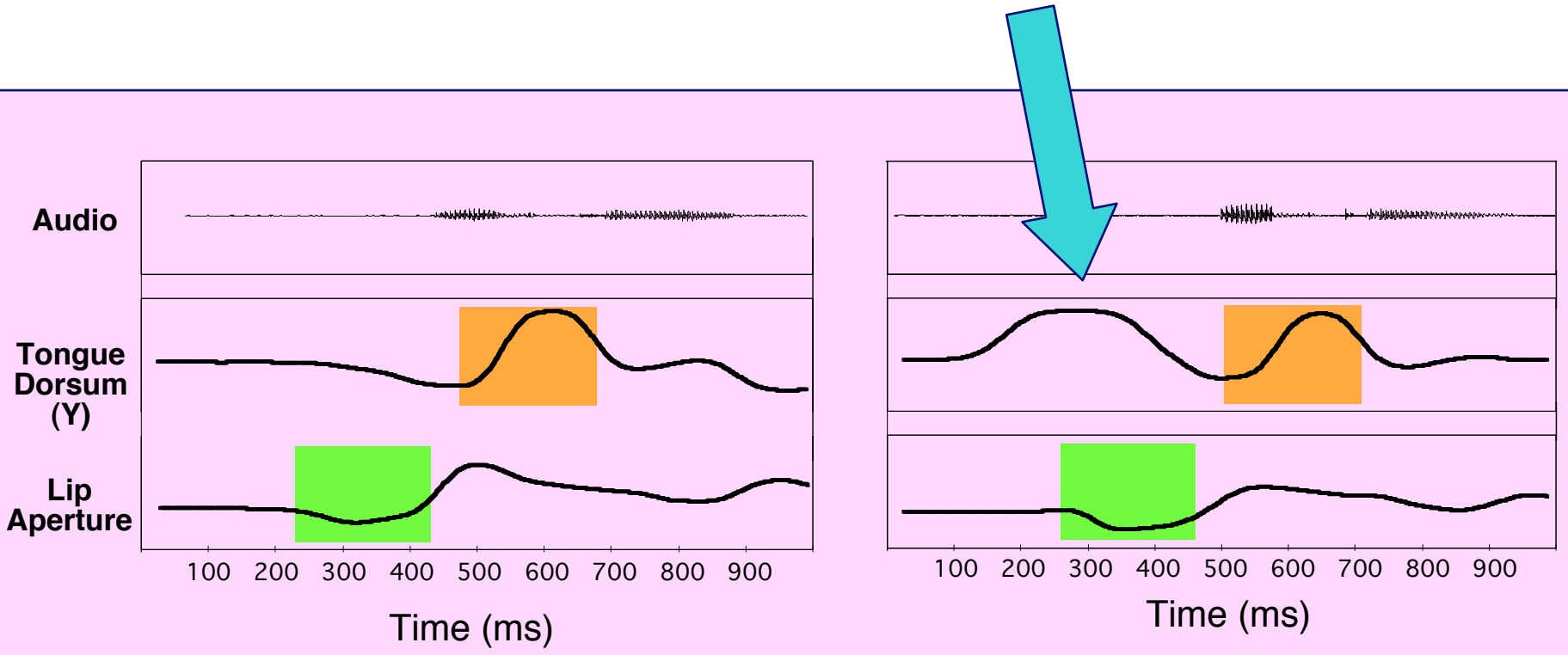
speak

- Results also show a dominance of intrusion errors.

Gestural Intrusion in SLIP

"bud goad"

Intrusion

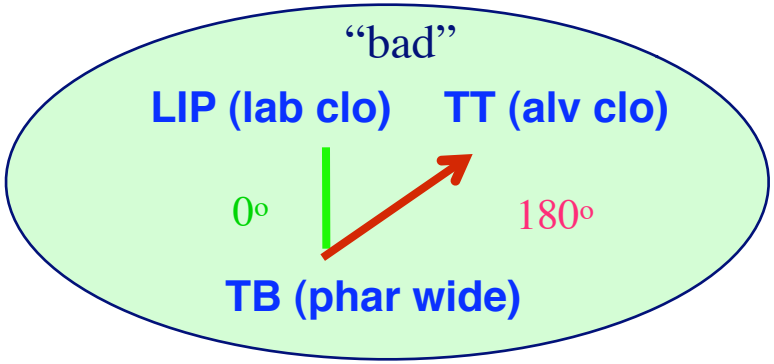


Stages of Task-Dynamic Speech Production Model

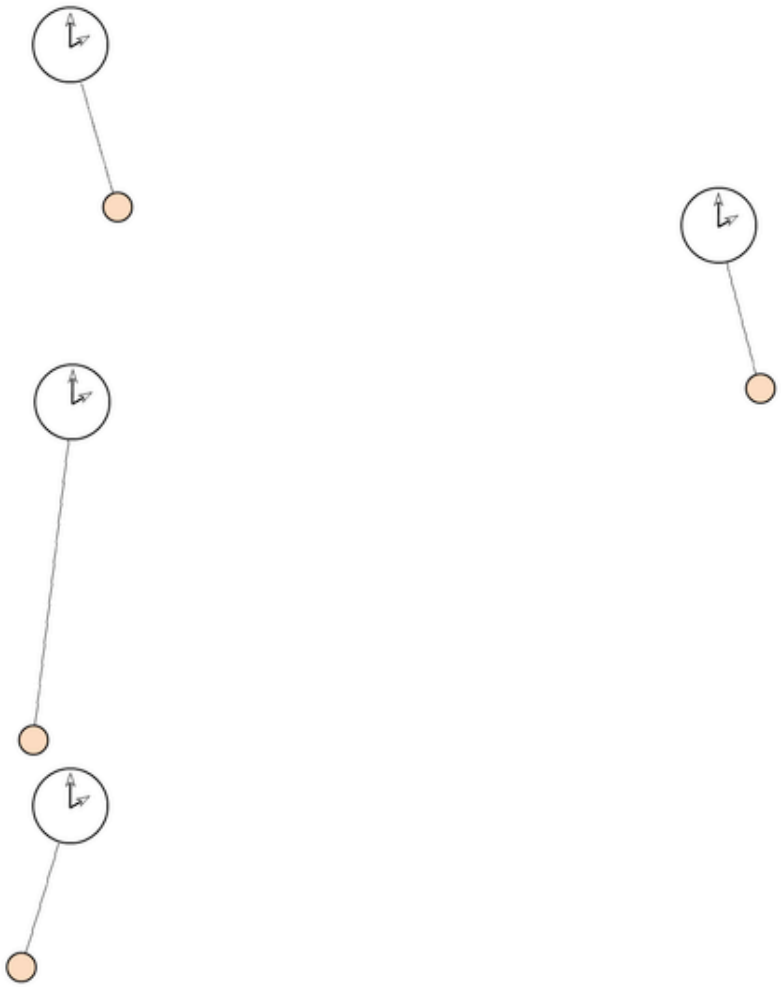
- Planning
  - Planning is be modeled as oscillator entrainment.
  - Produces temporal pattern of gestural activations from a phonological specification.
- Constriction Formation
  - Coordinated motion of articulators results from activations of invariantly specified gesture tasks.

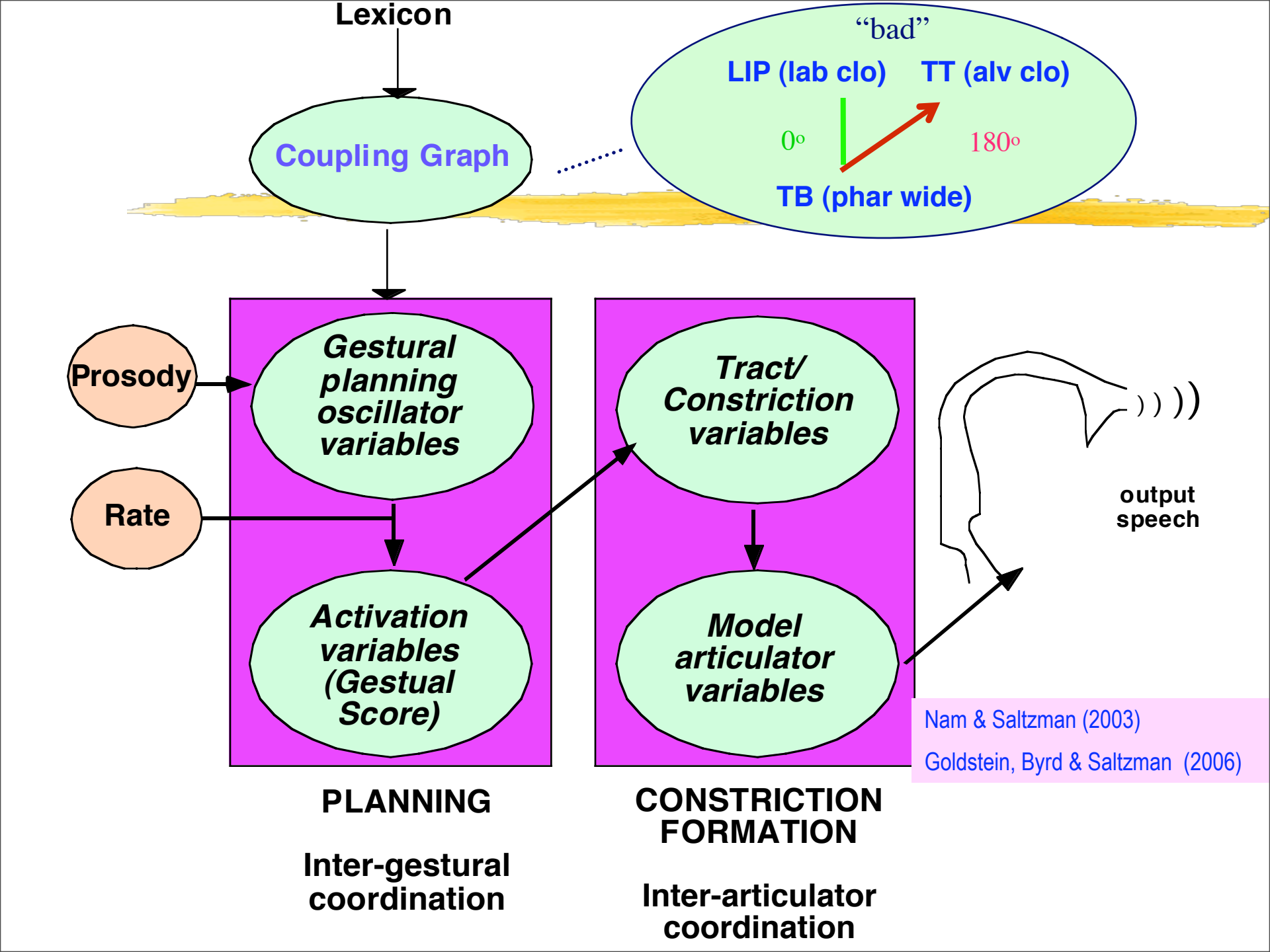
Planning model (Saltzman, Nam, Goldstein)

- Phonological input to planning is a coupling graph:
  - NODES: Specification of gestures and the associated planning oscillators
  - EDGES: coupling functions between pairs of planning oscillators.

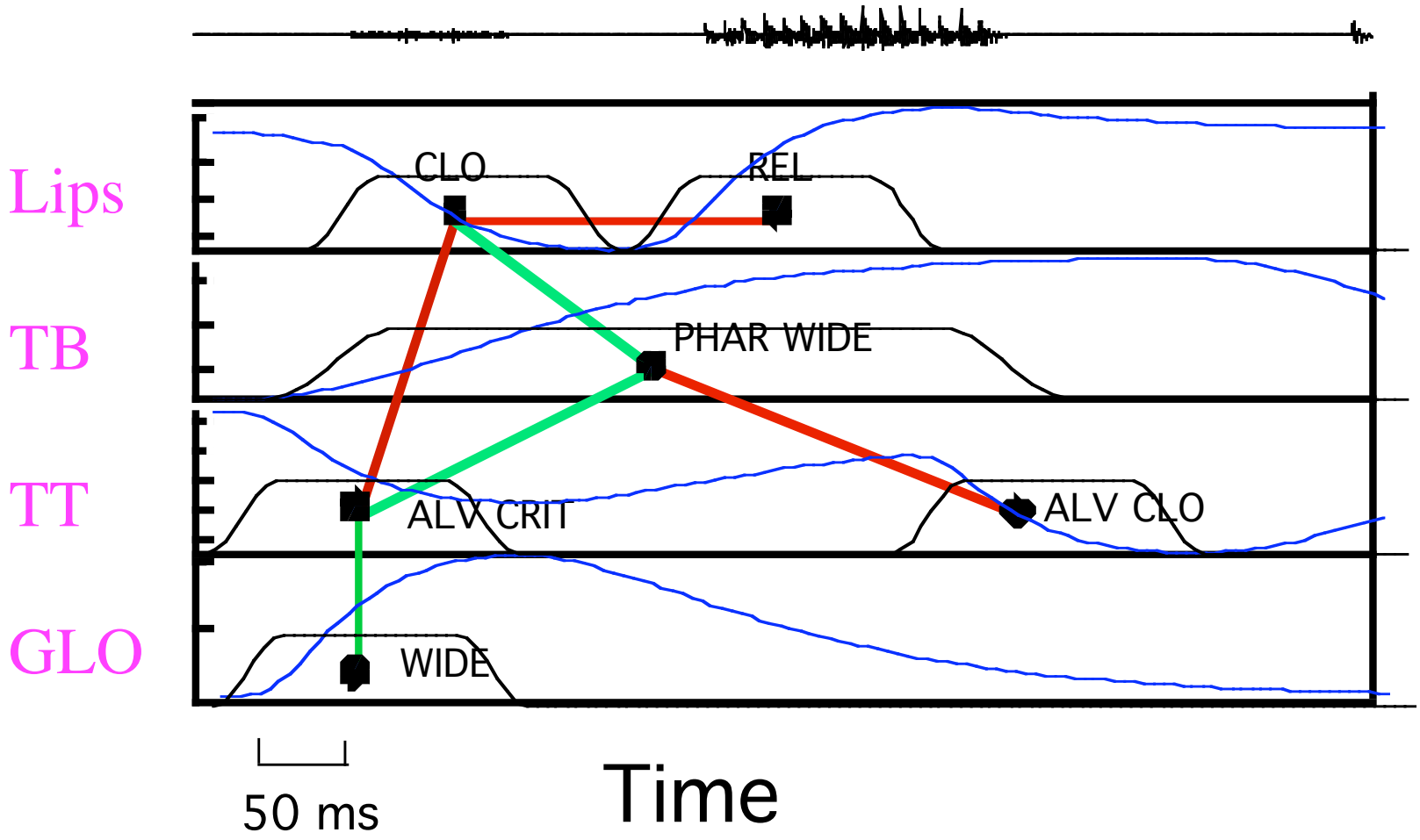


- At the beginning of planning process, oscillators are set into motion at random phases.
- Coupling forces specific to graph cause the oscillators to settle at stabilized relative phases (Saltzman & Byrd, 2000).
- Once stabilized, timing oscillators trigger the activation of their associated gesture(s).




Lexicon
Coupling Graph
"bad"
LIP (lab clo)
TT (alv clo)
0o
180o
TB (phar wide)
Prosody
Rate
Gestural planning oscillator variables
Activation variables (Gestual Score)
Tract/ Constriction variables
Model articulator variables
output speech
Nam & Saltzman (2003)
Goldstein, Byrd & Saltzman (2006)
PLANNING
Inter-gestural coordination
CONSTRICTION FORMATION
Inter-articulator coordination

Example: “spat”



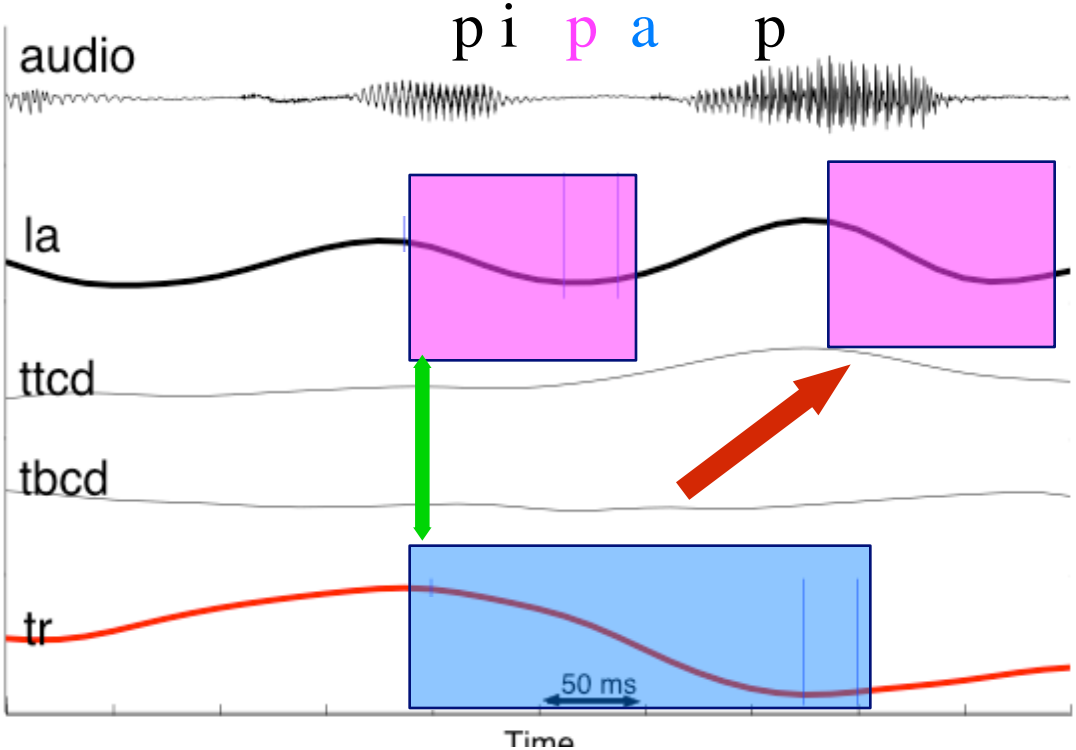
Syllable structure and modes in coupling graphs

- Certain relative phases values can identified as intrinsic modes of a system of coupled oscillators
- Hypothesis: these relative phases (modes)
  - are the ones employed in coupling graph
  - underlie syllable structure
- The hypothesis can account for
  1. universals of syllable structure
  2. acquisition of syllable structure
  3. planning time
  4. regularities of timing in CC clusters

Syllable structure and phase-locking modes

- If a basic consonant constriction (C) gesture and a vowel (V) constriction gesture are to be coordinated in a spontaneously available mode, there are just two possibilities:
  - in-phase
    - hypothesized for C-V (onset relation) most stable
  - anti-phase
    - hypothesized for V-C (coda relation)

Evidence for C-V and V-C modes



LIPS

TONGUE ROOT

Onset C and V gestures begin synchronously (Löfqvist & Gracco, 1999); Hypothesize that clocks are in-phase.

Coda C begins later than V; hypothesize that clocks are anti-phase.

Universality of CV structure

- All languages have CV syllables (e.g. Clements, 1990).
  - Not all languages have VC structures
  - This can be accounted for by the the fact that in-phase is the more accessible, more stable mode.

Combinatoriality

- Hypothesis: Combinatorial freedom of gestures is possible just where intergestural coordination exploits the most stable mode of coupling.
  - Onset C gestures combine freely with V gestures (in-phase coupling)
  - Coda C gestures are in a less stable mode with Vs, and therefore there is increased dependency between V and C.
  - Within-onset and within-coda coordination may employ non-intrinisically available modes.
    - specific couplings are learned
    - acquired late
    - typically small numbers of combinations

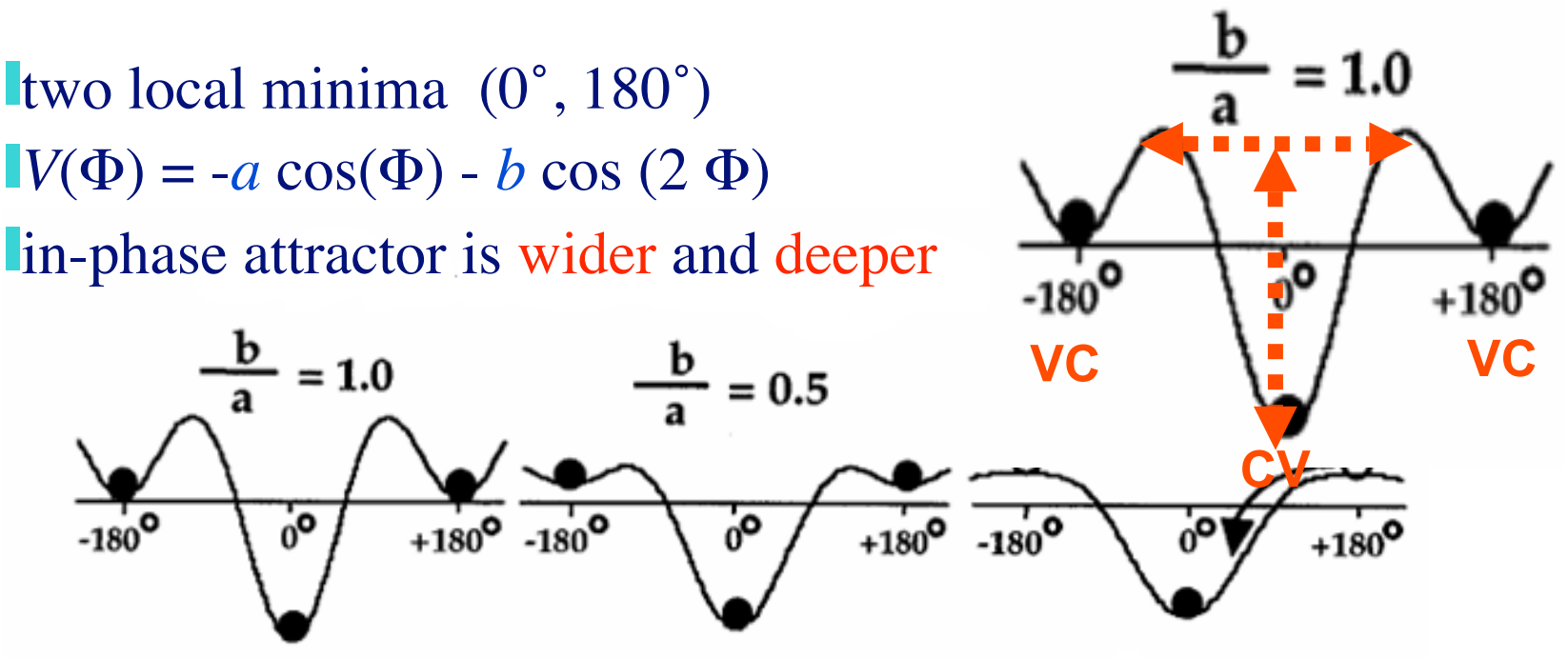
Acquisition of syllable structure

(Nam, Goldstein & Saltzman, in press)

- Infants develop onsets (CV) before codas (VC) in all languages.
  (e.g.Vihman & Ferguson, 1987; Fikkert, 1994)
- Lag in acquisition of codas is shorter in languages that make more frequent use of VC (Roark & Demuth 2000).
- These facts can be accounted for with a model that includes both:
  - Preference for in-phase mode
  - Attunement to C<->V phase in the ambient language

Haken-Kelso-Bunz (HKB), 1985 Coupling Model

- An intrinsic potential function has been proposed that can model the results of many experiments on bi-manual coordination
  - two local minima (0°, 180°)
  - $V(\Phi) = -a \cos(\Phi) - b \cos(2\Phi)$
  - in-phase attractor is wider and deeper

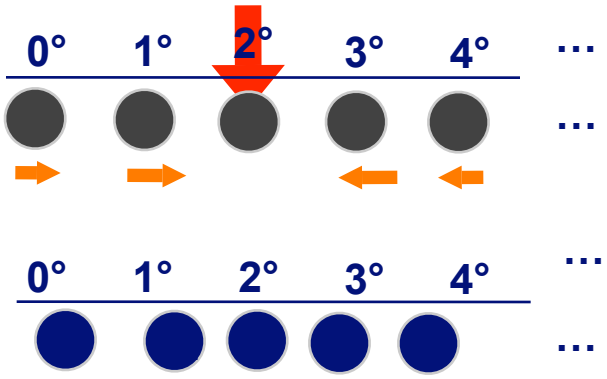


  - As b/a decreases, anti-phase minimum disappears
  - If b/a is a function of oscillaton frequency, spontaneous transitions to in-phase can be explained

Model of phase learning (Nam, 2007)

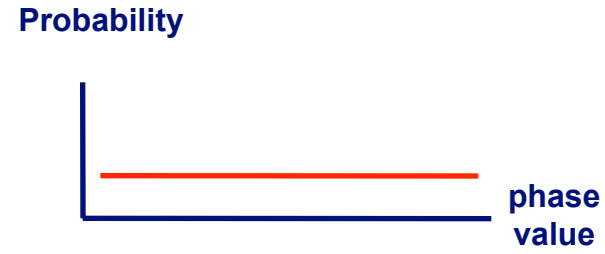
- "neural" units represent values of a phase continuum

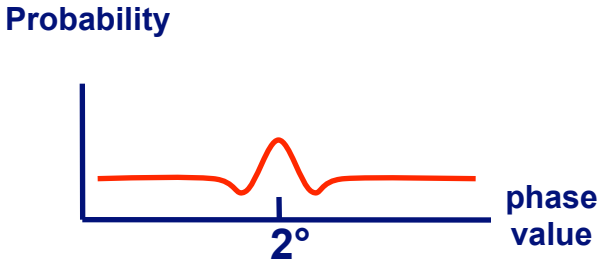
Selected value 2º matches



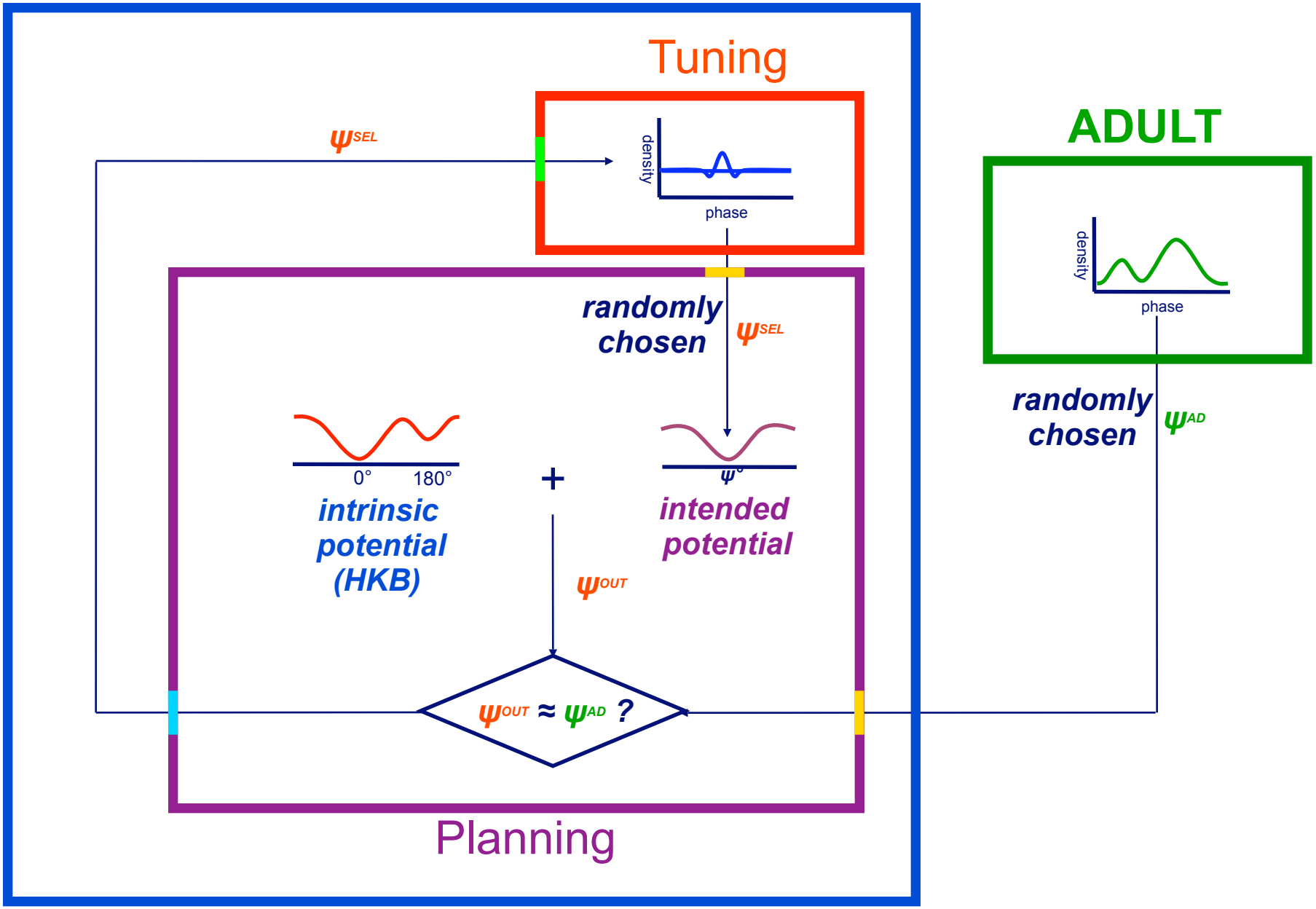
- neural units are slightly attracted to the phase value that has matched an adult utterance (*tuning or learning*)

- Units are *selected* at random from production.
- Distribution is flat at outset





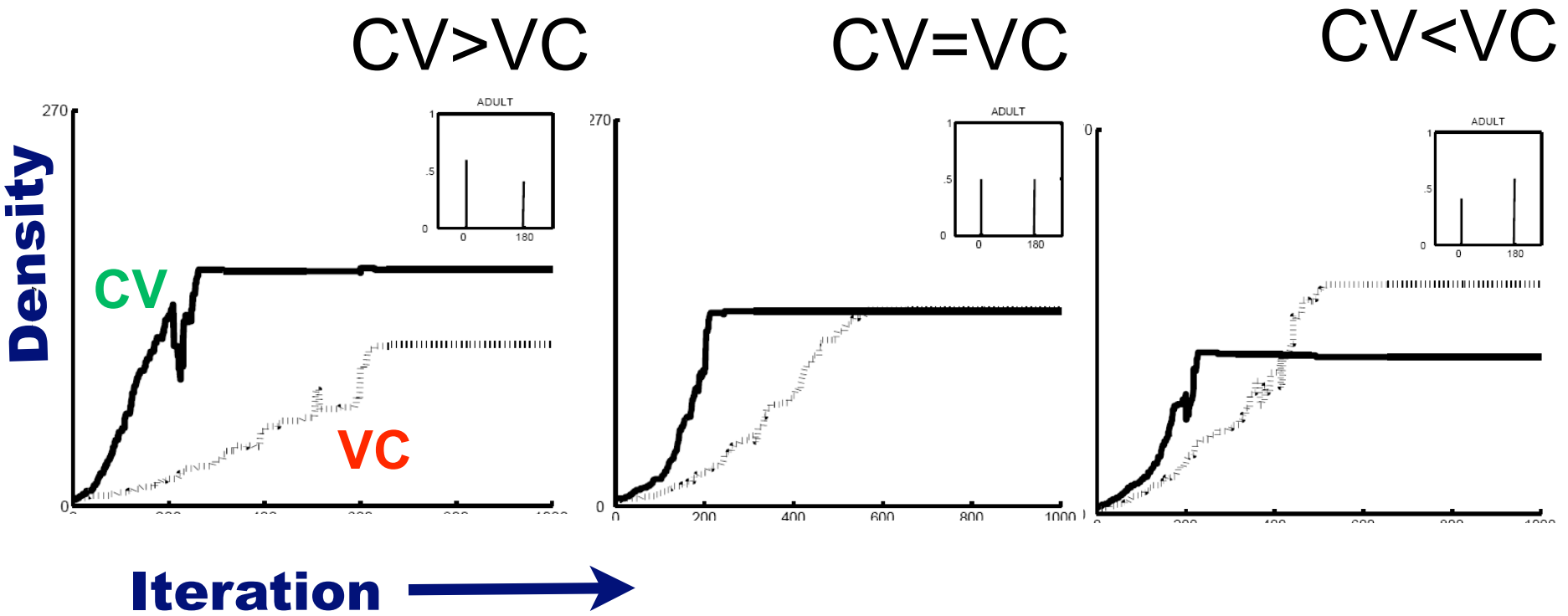
- Probability of selecting a *matched* phase value increases


CHILD
Tuning
ψSEL
density
phase
ADULT
density
phase
randomly chosen
ψSEL
randomly chosen
ψAD
0°
180°
intrinsic potential (HKB)
+
ψ°
intended potential
ψOUT
ψOUT ≈ ψAD ?
Planning

Simulation Conditions

- Adult Frequency modes
  - CV>VC
  - CV=VC
  - CV<VC

Results

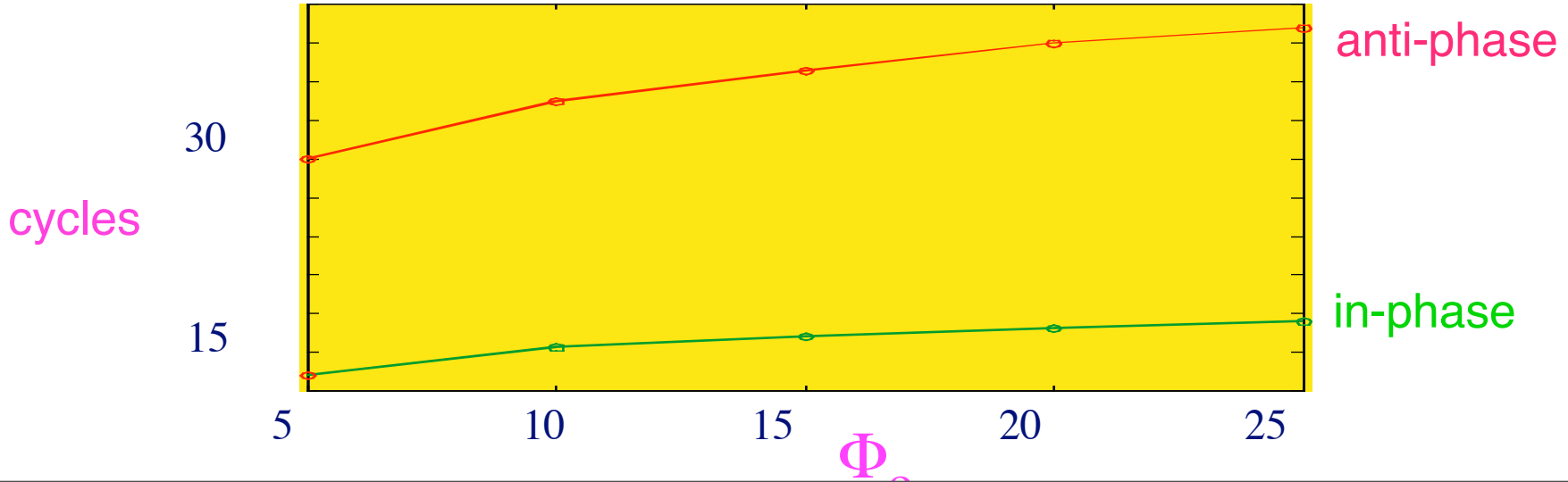

CV>VC
CV=VC
CV<VC
Density
CV
VC
ADULT
270
0
180
200
400
600
800
1000
Iteration

Summary

- For all adult “languages,” frequency modes corresponding onset consonants (CV) emerge earlier.
- Eventually, the child learns to match adult frequencies.
- Lag between emergence of CV and VC modes is influenced by adult frequencies

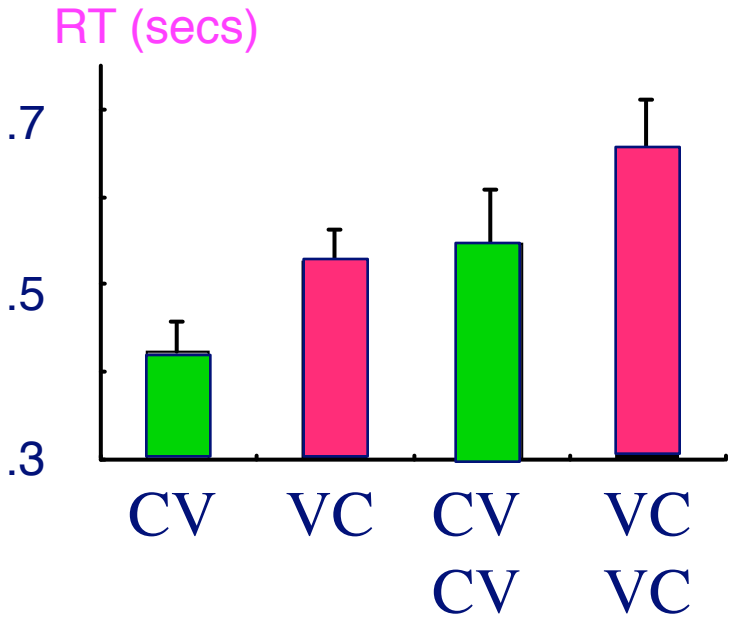
Settling Time: Simulations

- Because the in-phase attractor has a steeper well, relative phase should settle at its target more quickly.
- Test:
  - Initial $\Phi_o = 5, 10, 15, 20, 25$ degrees from target (0^o or 180^o)
  - 100 repetitions for each Φ_o with random choice of individual oscillator phases (ϕ_1, ϕ_2)



Planning time: Pilot Behavioral Experiment (Nam, 2006)

- Task:
  - Produce syllable(s) as fast as possible
  - segments of a syllable were verticall aligned
    (not a reading task)
- Syllable types:
  CV, VC, CVCV, VCVC
  - C and V randomly chosen among /p, t, k/ and /a, e, i, o, u/
- Measure:
  lag from stimulus to acoustic onset of response
- Subjects:
  - 2 English speakers
  - 2 Korean speakers



BUT:
- Frequency effect?
- Articulation?

Planning time experiments

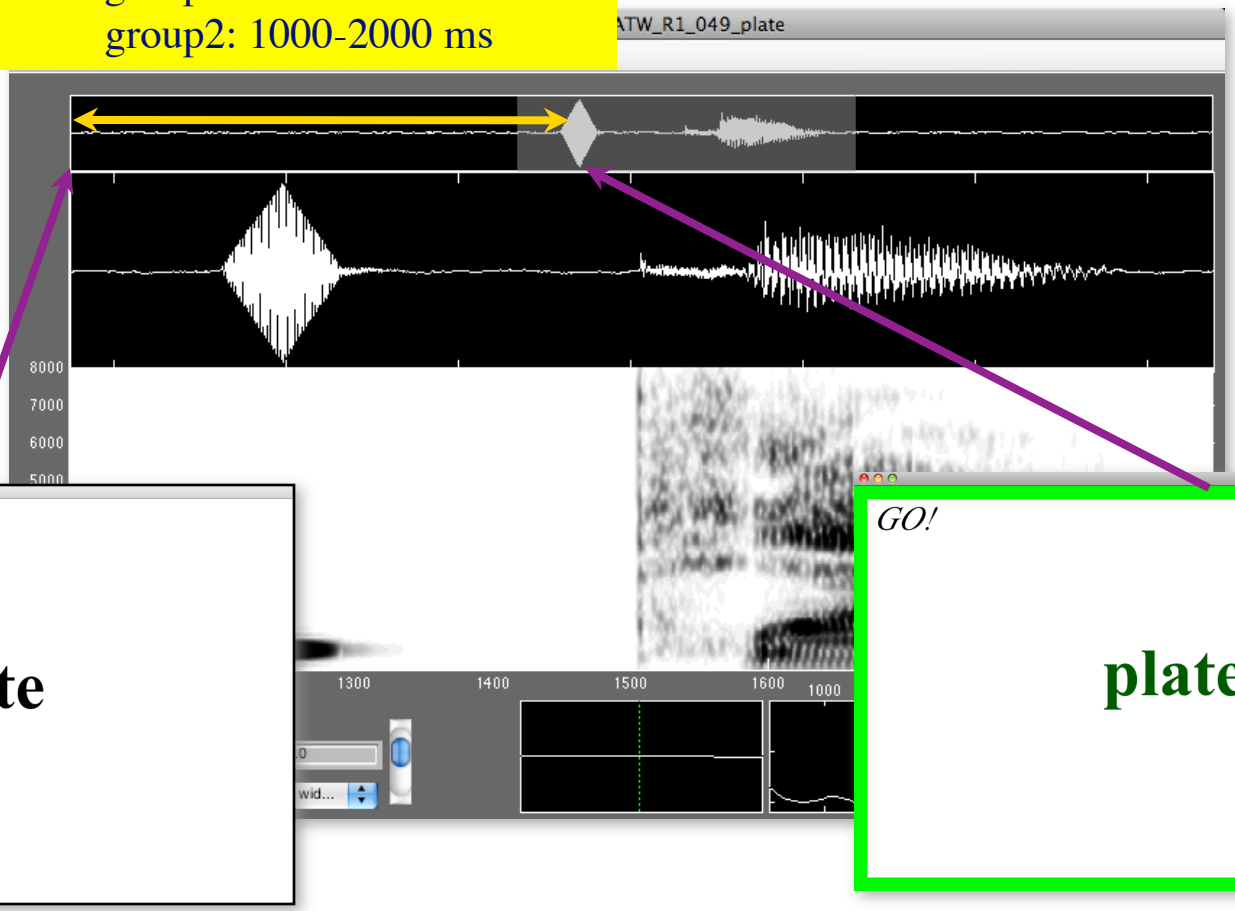
(Mooshammer et al, 2008)

- Acoustic-Only Response
- Tasks:
  - simple delayed naming
  - post-vocalic delayed naming
- Participants:
  - 20 American English (12 female, 8 male)
- Materials:
  - VC, CV
  - V: /eI/ (‘pay’) /i:/ (‘pea’)
  - C: /p, t, k, s, l/
  - Other structures (CVC, CCV, CCVC)

- EMMA
- Task: Post-vocalic naming
- Participants:
  - 4 American English (3 female, 1 male)
- Materials:
  - VC, CV
  - V: /eI/ (‘pay’) /i:/ (‘pea’)
  - C: /p, t, k, s, l/
  - Other structures (CVC, CCV, CCVC)

Simple delayed naming

Random delay:
group1: 1000-1600 ms
group2: 1000-2000 ms

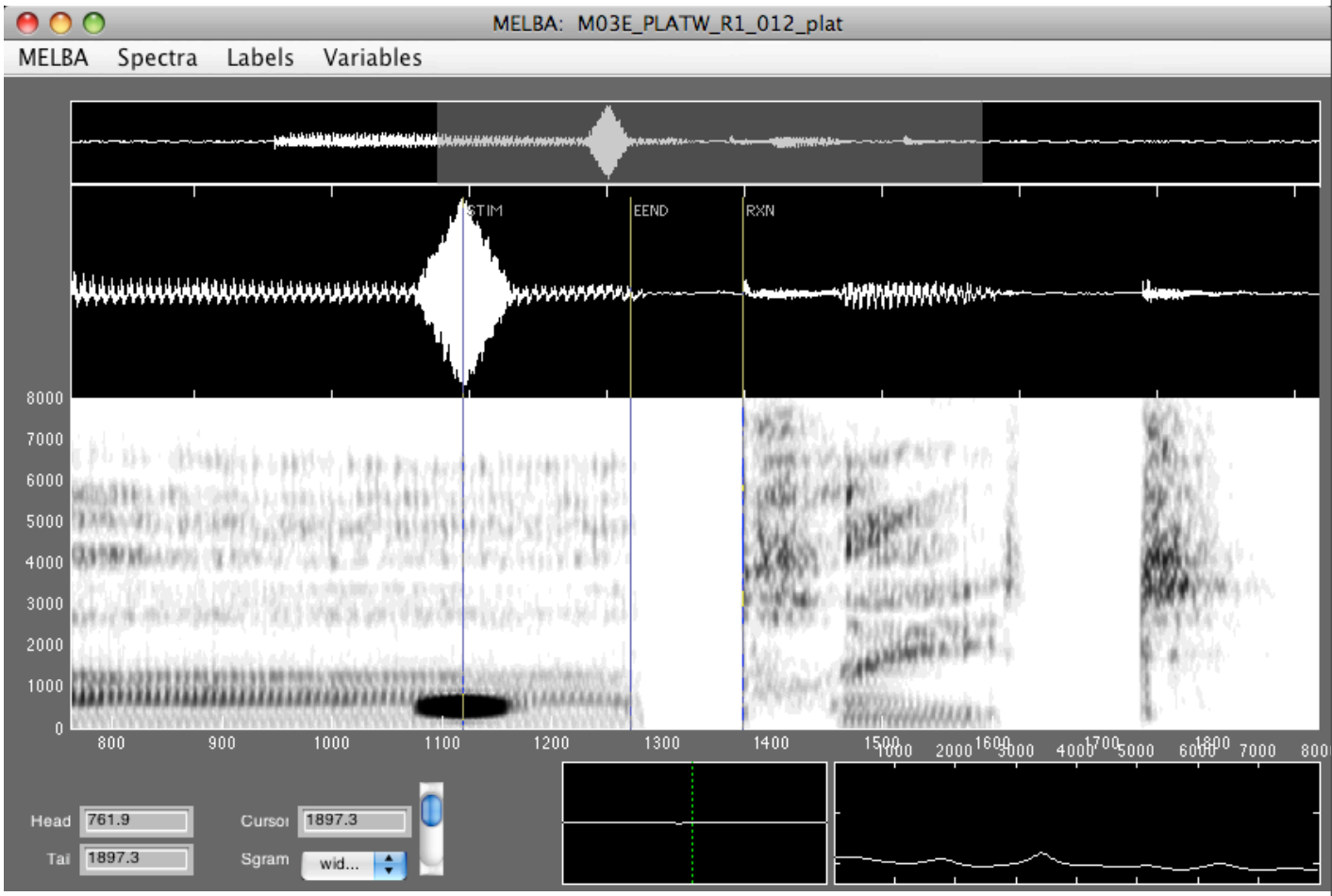


Postvocalic delayed naming

Instruction:

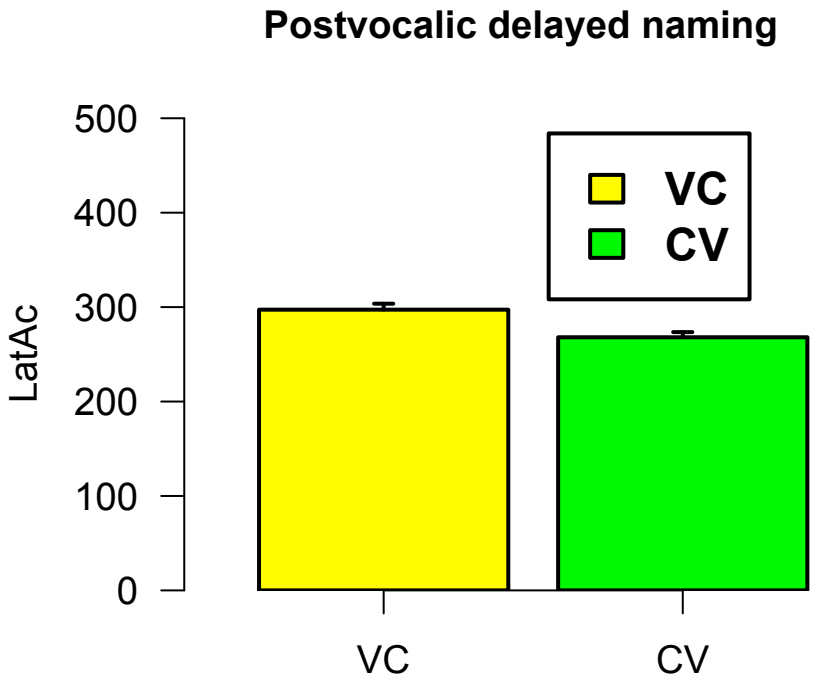
Get ready (say "uh")

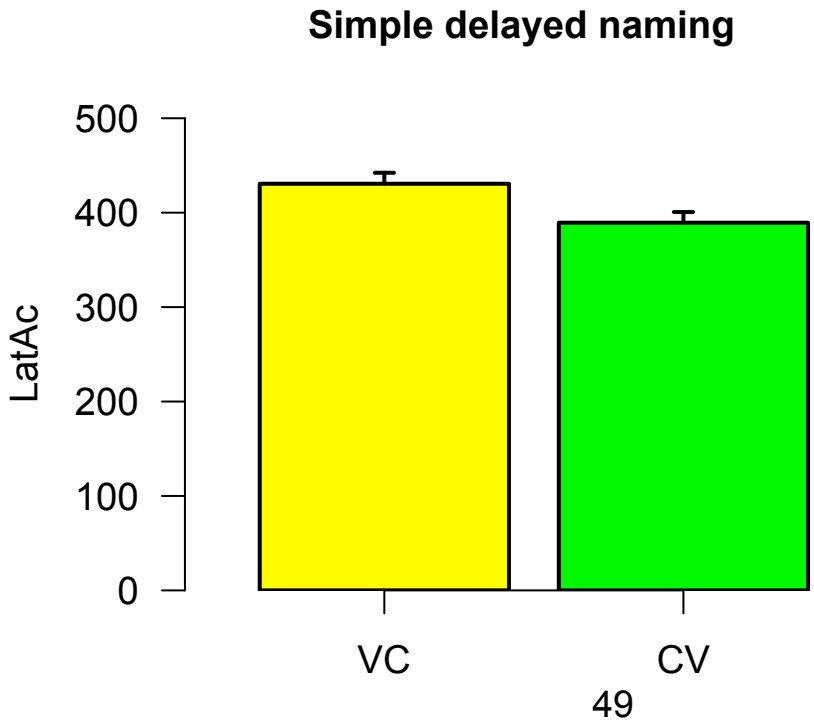
for detecting the onset of stops



Acoustic-only Results

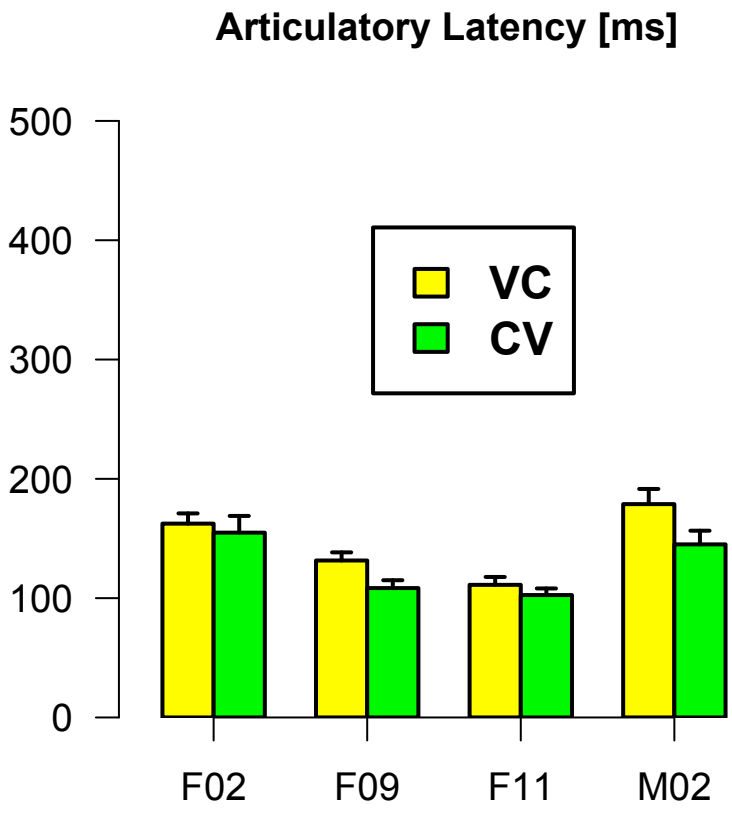
- CV significantly faster than VC

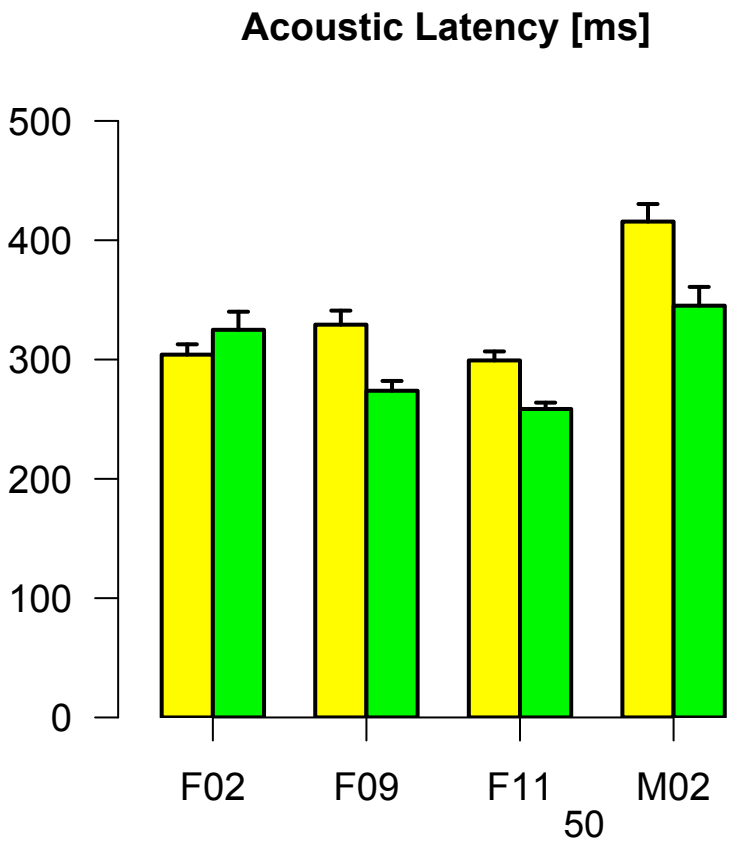




Articulatory Results

- CV significantly faster than VC (except F02)





Summary

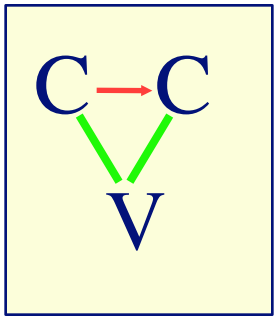
- Robust difference between CV and VC in planning time.
- Effect likely not due to frequency effects.

CC clusters in onset

- If onset is defined by an in-phase relation between C gesture and V, then all onset C gestures should be synchronous with V (and therefore with each other).
- Multiple constriction gestures in onset cluster (e.g., /spat/).
  - Gestures must be at least partially sequential to afford perceptual recoverability and to allow order contrasts (/spa/ vs /psa/)
  - What in the coupling graph identifies them as both in the onset?

Competitive coupling hypothesis (Browman & Goldstein, 2000)

- Specifications in the coupling graph can compete with one another
- C-V coupling
  - *All* C gestures in onset coupled in-phase with the V.
- C-C coupling
  - C gestures also coupled sequentially (anti-phase)

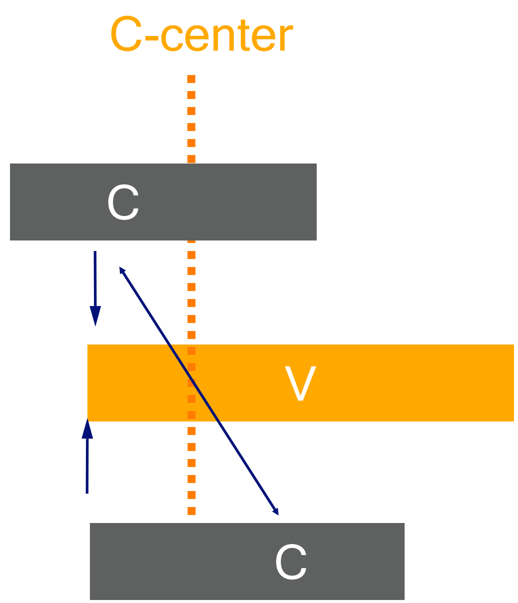


- Prediction: Observed coordination should reveal the presence of both couplings.

As Cs are added to an onset:

- Rightmost C (C_n) should shift later with respect to the vowel.
- Leftmost C (C_1) should shift earlier with respect to vowel.

Modeling shift with competitive coupling



results of competition

C-V phasing

C_1 shifts left

C_n shifts right

But C-V ...

Add an additional coordination

(C-C phasing)

C-V and C-C phasings

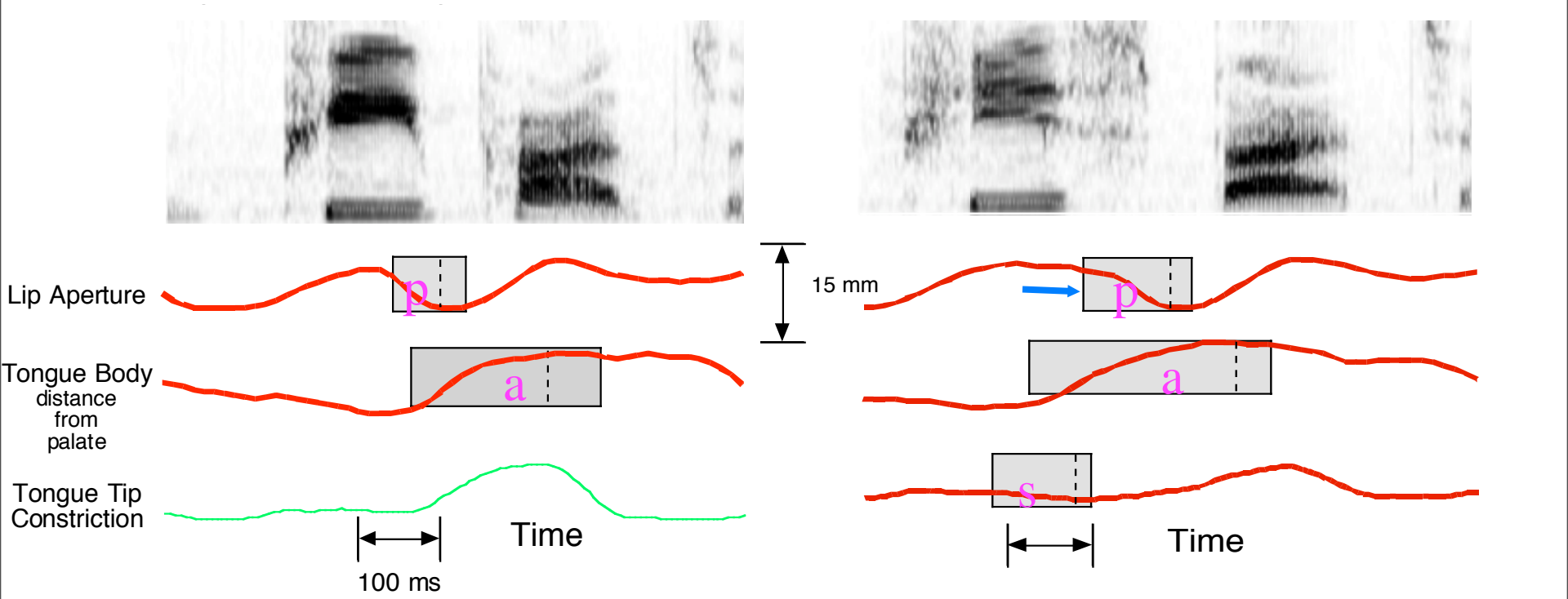
global C-center

in competition

Rightward shift of C_n

“pea pots”

“pea spots”



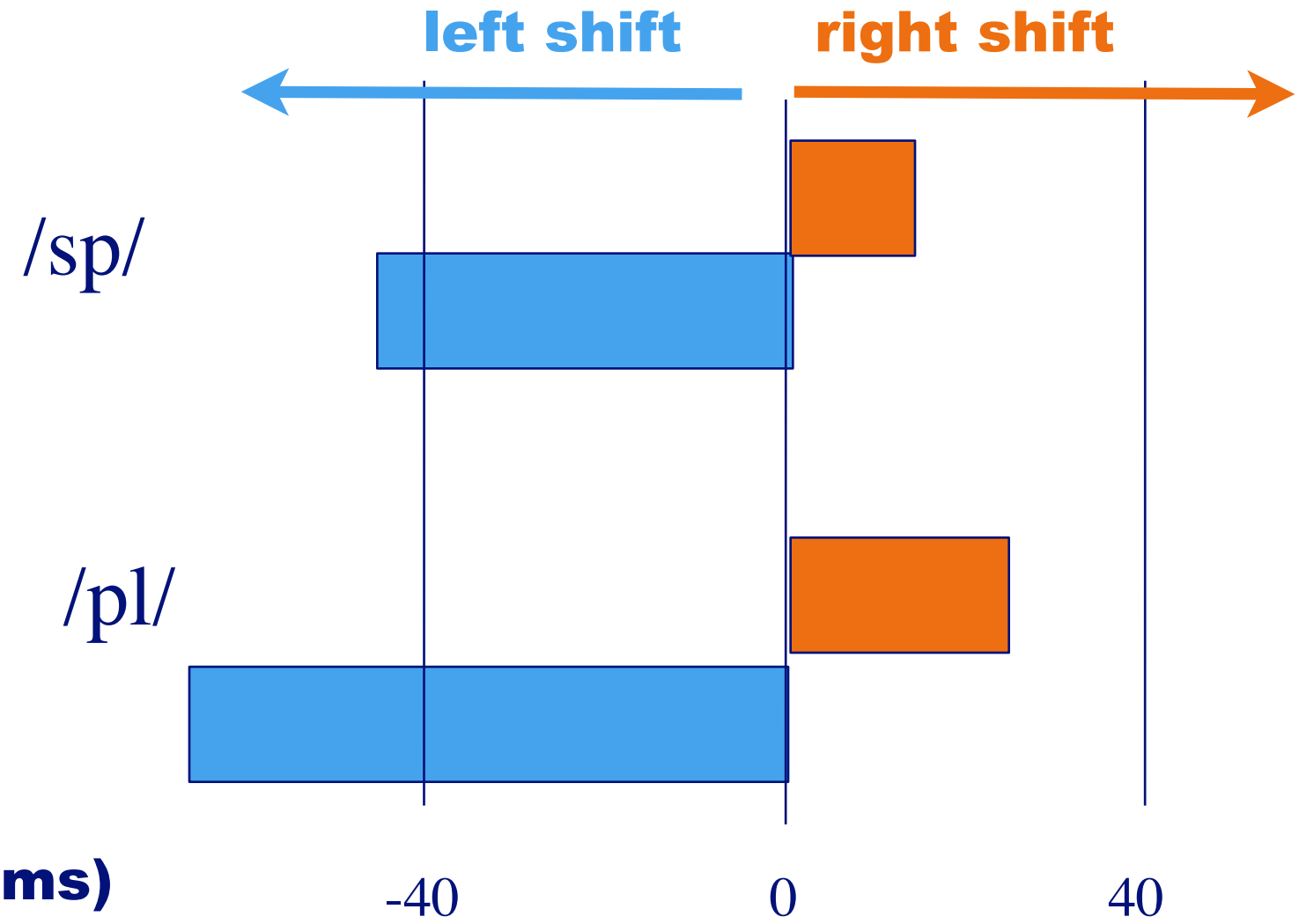
Symmetry of shift: English x-ray experiment

/sp/	/pl/
pa seats	pa peats
pa peats	pa leets
pa speets	pa pleets

- 6 Subjects
- 2 subjects varied accent patterns
  - A = accent on *Pa*
  - B = accent on *_eets*
- accent on *_eets*.

- Measure times of gesture target achievement
- Comparisons (e.g. for /sp/):
  - left shift = time from [s] to V in /s/ - time from [s] to V in /sp/
  - right shift = time from [p] to V in /p/ - time from [p] to V in /sp/
- Comparable for /pl/

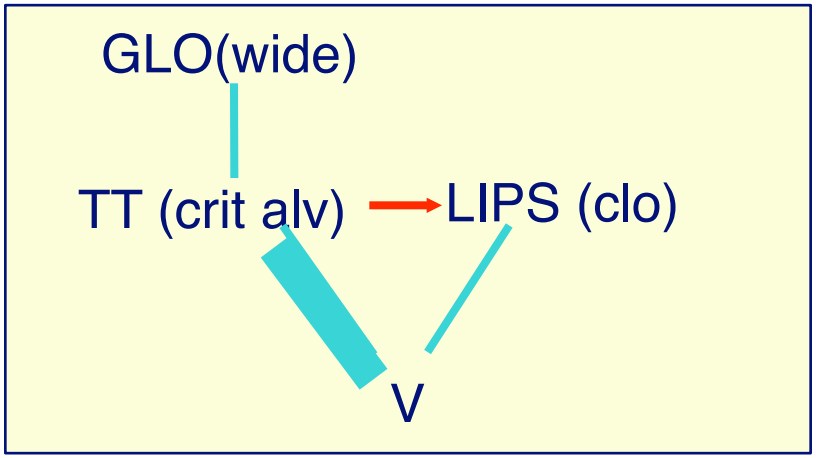
Results: Means



Results: Summary & Model

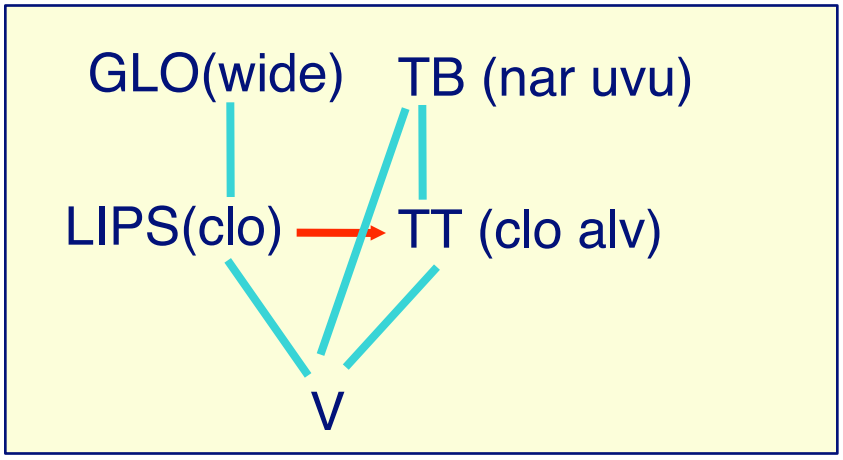
- /sp/
  - Slightly more leftward shift overall (~20 ms.)
  - Not systematic across subjects
- /pl/
  - Larger asymmetry for leftward shift (~60 ms.)
  - Systematic for every subject
- Coupling Graph Model
  - systematic, cross-subject effects should be represented in graph topology
  - individual differences should be modeled by tuning coupling strength of in-phase links.

Coupling Graph models



/sp/

- model produces no asymmetry.
- speaker differences can be modeled though coupling strength



/pl/

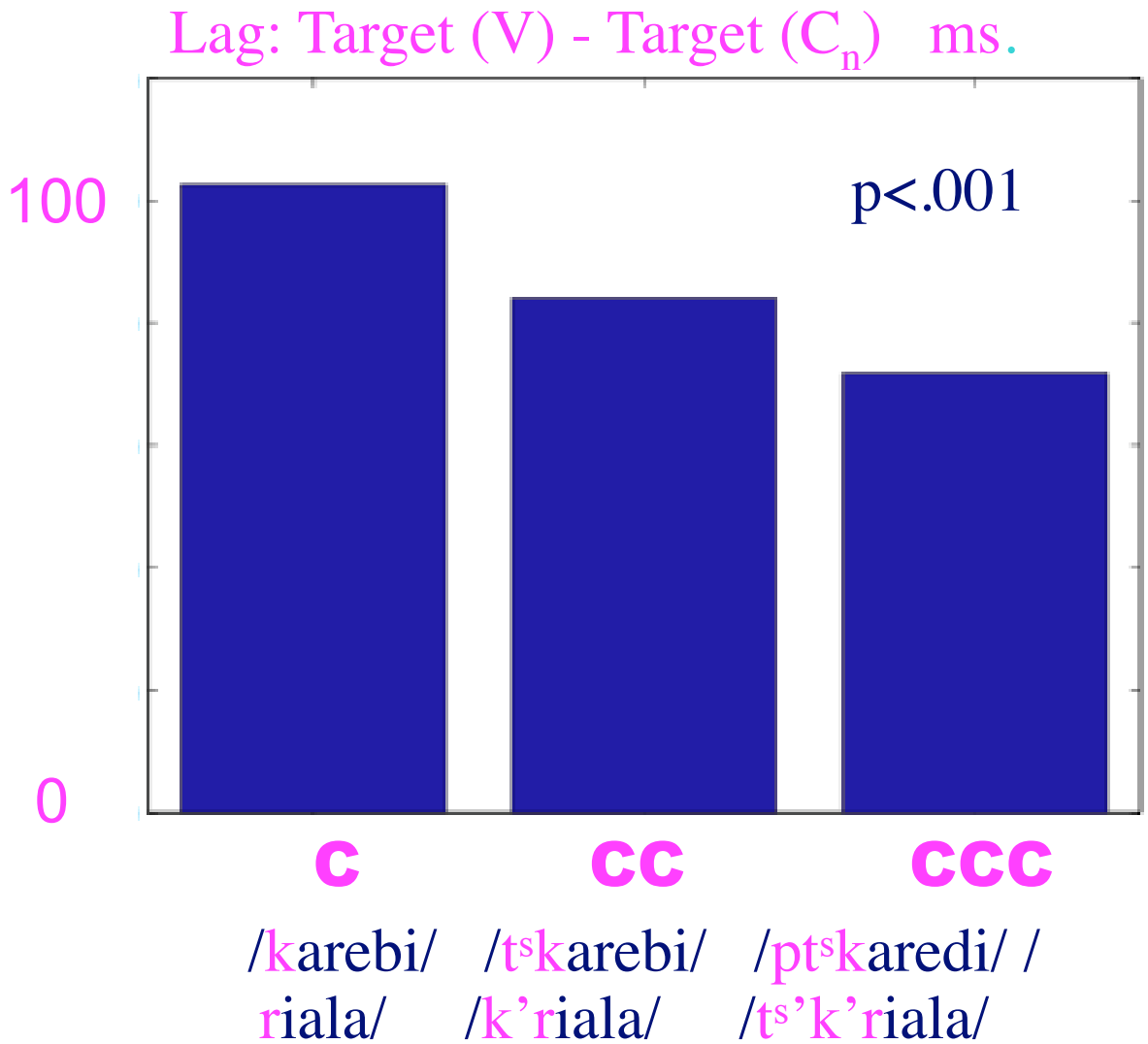
- Additional TB coupling with V gives advantage to the the in-phase coupling of /l/ with V.
- Model results show asymmetrical shifts: left shift > right shift

Rightward shift of C_n as diagnostic for onset?

- If a consonant sequence is syllabified as part of an onset, then it should exhibit rightward shift.
- Georgian and Tashlhiyt Berber are languages in which words can begin with sequences of 3 obstruents.
- Languages differ as in syllabification of such words:
  - Georgian Cs are complex onsets
  - Berber only allows a single C in onset, other Cs constitute nuclei of additional syllables.
- Do Georgian and Berber differ in rightward shift?

Georgian: Rightward shift

EMMA data (w/Chitoran)



/p/ Lip Aperture

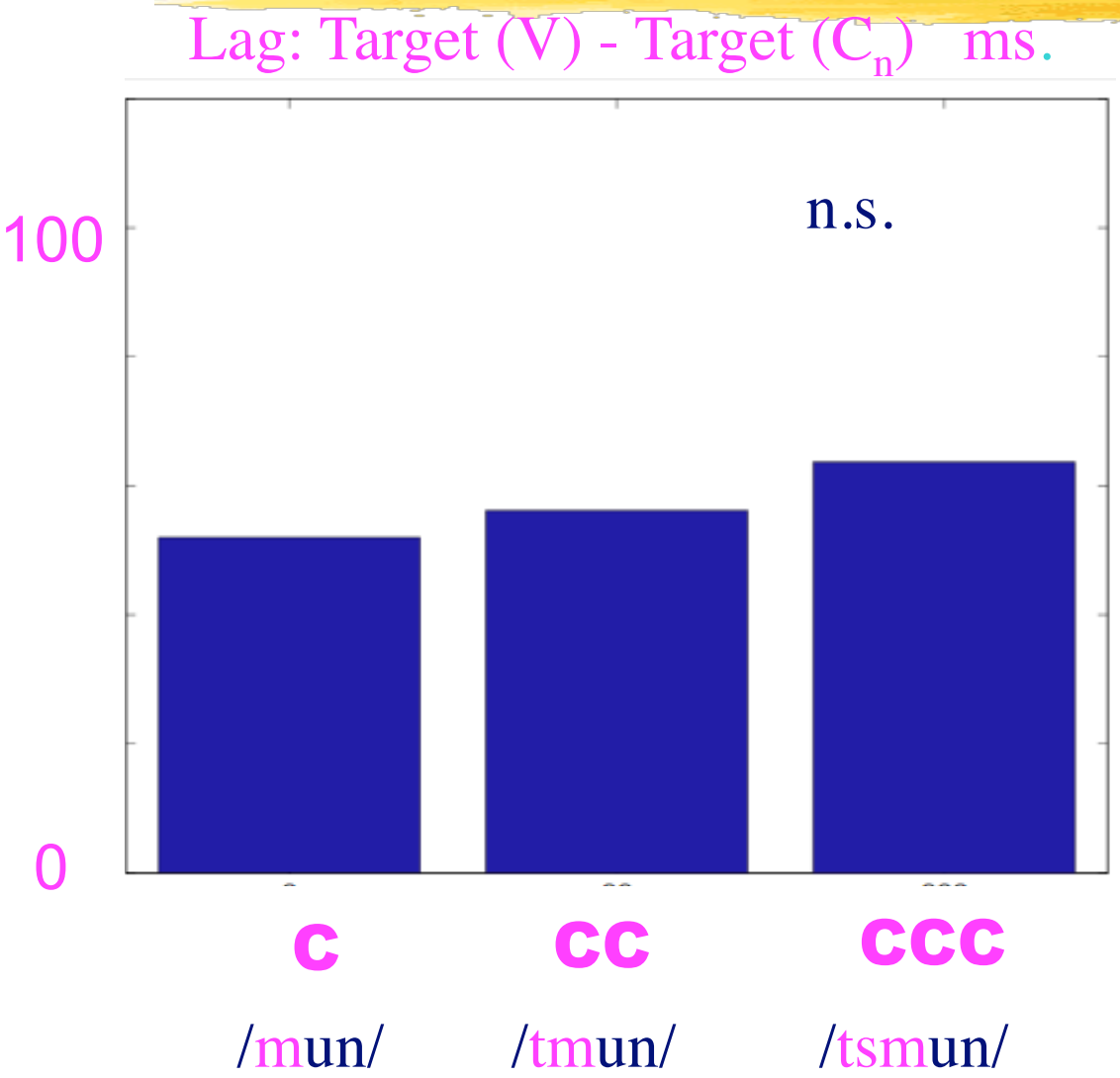
/t/ Tongue Tip Constriction Degree

/k/ Tongue Dorsum Constriction Degree

/V/ Tongue Body Constriction Degree

2 speakers

Tashlhiyht Berber EMMA data (w/Selkirk)



/m/ Lip Aperture

/t/ Tongue Tip Constriction Degree

/s/ Tongue Tip Constriction Degree

/V/ Tongue Body Constriction Degree

2 speakers

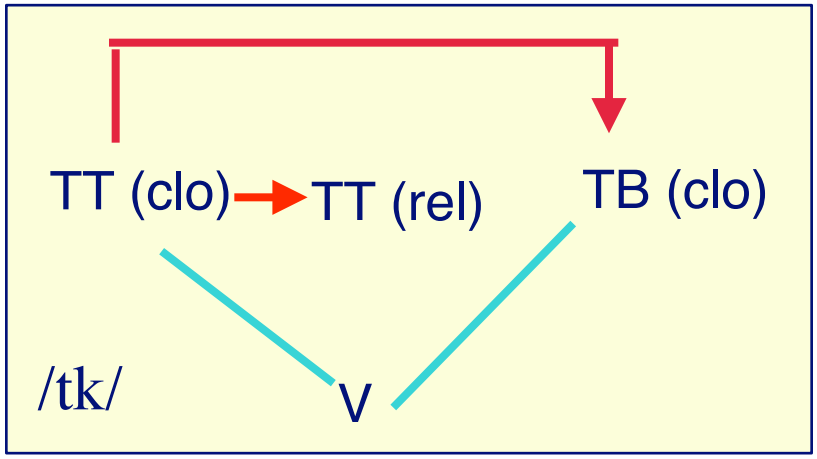
Summary

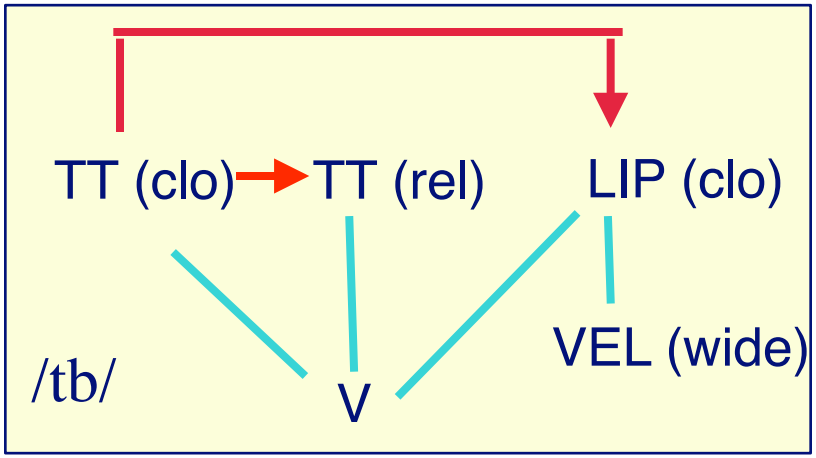
- Differences between Georgian and Berber provide tentative support for the modeling of onsets as competitive coupling structures.
- The rightward shift could be a useful diagnostic measure of syllabification.
- Results support the (controversial) syllabification of Tashlhiyt Berber.

Georgian Cluster types

- front-to-back (/t^{s}/k)< back-to-front (/t'b/)
  - greater lag in back-to-front (Chitoran et al, 2002)
  - back-to-front fail to exhibit rightward shift
- Model:
  - Make use of the fact that C releases must be actively controlled, and can appear in graph with separate oscillator (Nam, in press).

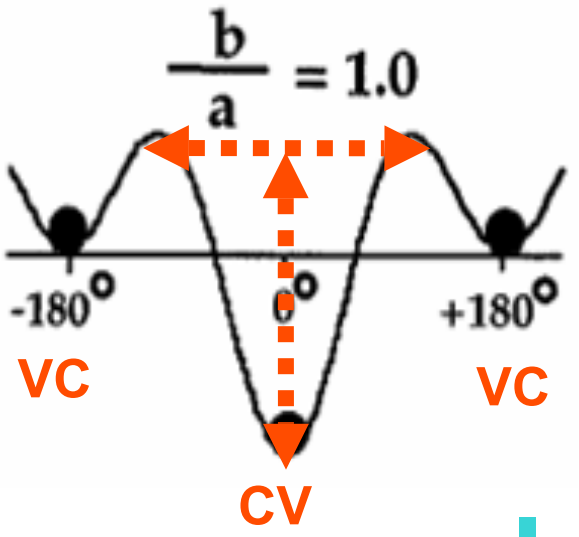
Alternative graphs: greater/lesser overlap

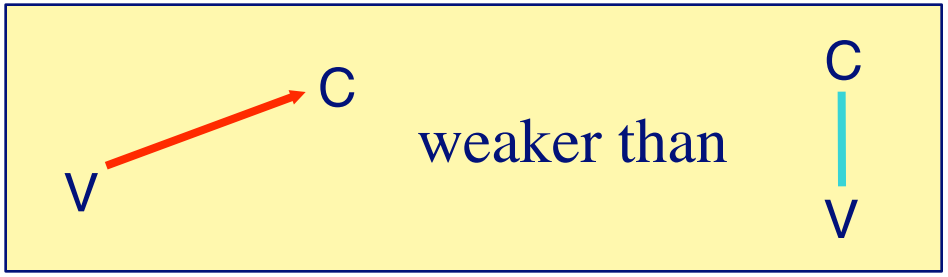




- Model produces about 25 ms differences in lag.
- Right magnitude for Georgian data
- No competition, so no rightward shift predicted.

Onset vs. Coda Clusters





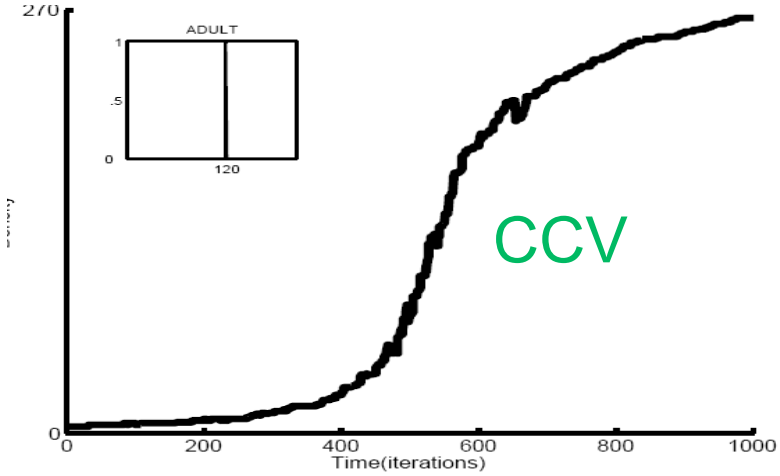
- Consequences:
  1. Easier to learn to produce CC clusters in Coda than in Onset
  2. In VC_1C_2, C_2 may not be coupled to V.

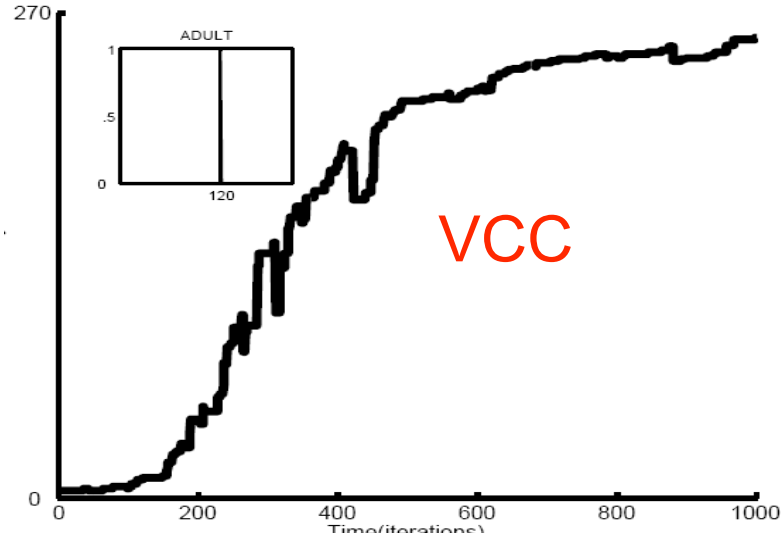
Paradox: Acquisition of CC

- In many languages, consonant clusters can be acquired in coda before onset (opposite result from single Cs).
  - English, Dutch, Spanish, German, Telegu
- This is predicted by phase learning model
  - Weaker V-C coupling in Coda makes it easier to learn to produce C-C sequencing

Extension of Phase Learning Model to Clusters

- Add 2nd C to learners experience.
- Coupling graph has to be learned:
  - C-C
  - C-V (or V-C)
- Development of C-C mode is faster in coda than in onset:
  - less strong competing synchronization

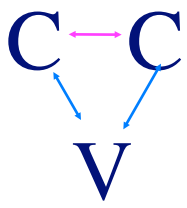


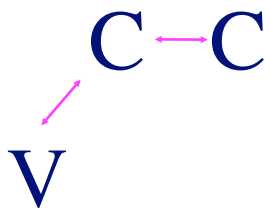


CC in onset vs. coda: possible coupling graph differences

- Hypothesis: No competitive coupling in coda for English

Onset





Coda

- Can account for differences between onset and coda in:
  - timing
  - syllable weight
  - variability in timing

Conclusion: syllable structure embodied in coupling modes

- Can provide a basis for an account of both macroscopic and microscopic properties of phonological structure:
  - hierarchical/combinatorial structure of syllables
  - acquisition of syllable types
  - generalizations of intergestural timing and its relation to syllable structure
  - planning time

Current model directions

- Coupling of gesture oscillators to rhythmic oscillators
  - foot
  - phrase
- Local modulation gestures associated with stress and accent