Coordination of speech gestures and syllable structure

Syllabification

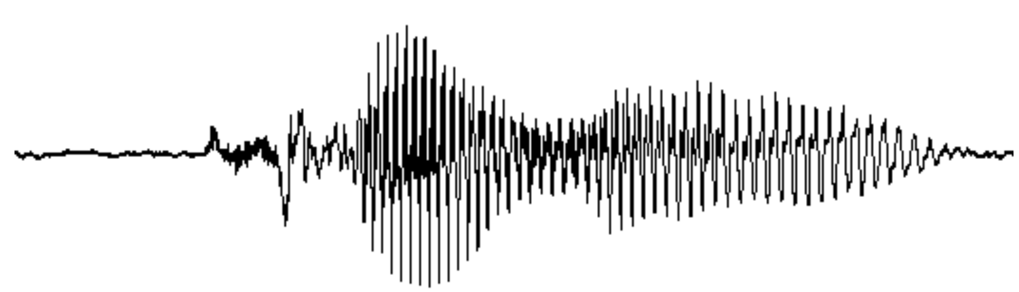
- As Ladefoged notes, English speakers (and speakers of other languages) generally agree about the number of syllables in a word. Some problematic cases:

- (a) "prism, mysticism"
  (b) "peel, seal, pail,"
  (c) "fear, fire, hour"
  "hire" vs. "higher"
  (d) "mediate, heavier, neolithic"

- How can syllables be defined phonetically?
  - Sonority Peaks
  - Jaw cycles
  - Gestural Coordination

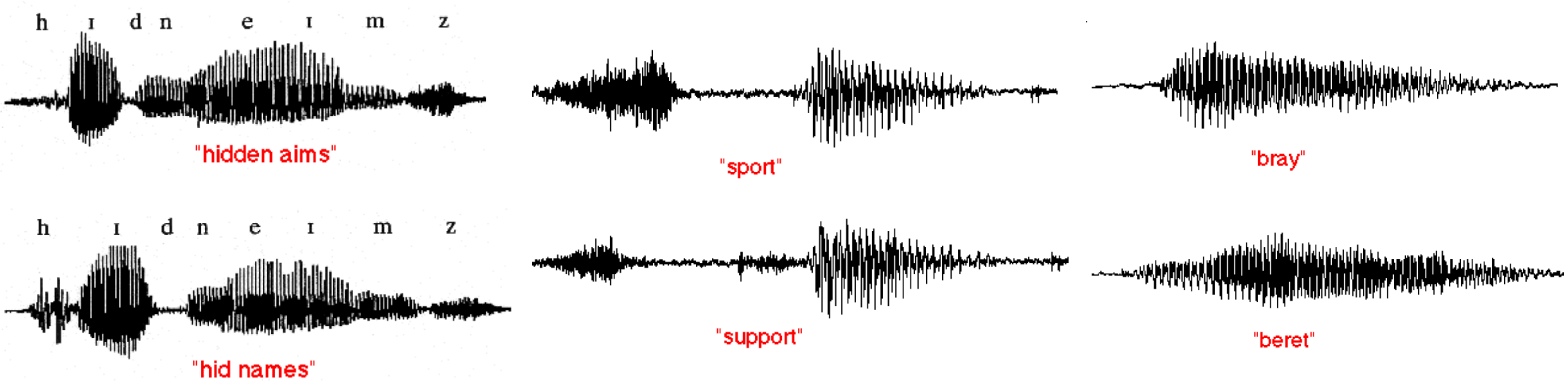
Sonority Peaks

- The number of syllables corresponds to the number of distinct amplitude peaks in the acoustic signal. Why?
- Each peak corresponds to the nucleus of a syllable.
- Sonority principle can account for relatively clear cases.
- It could also explain the disagreements in (1). The "extra" syllable corresponds to a local peak of amplitude that is greater what precedes it, but is not as high as a typical nucleus. E.g,



“prism”

Problems for Sonority

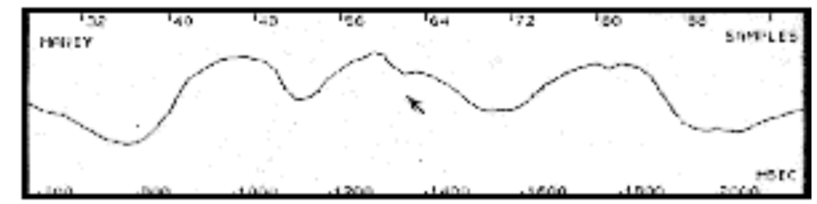


Jaw Cycles

- Frame-Content Theory (McNeilage & Davis)
    - Speech develops in the infant initially through the oscillation of the jaw.
    - Each cycle of the jaw oscillation corresponds to a syllable and is called by M&D ‘the frame.’
    - The individual consonants and vowels are “the content” and develop later.
    - For adults, the nucleus corresponds to the downward phase of the jaw cycle, the onset and coda correspond to the upward phase.
- In careful speech, jaw behavior differentiates "support" from "sport."
- However, for faster, more casual speech, the "extra" jaw deflection for "support" is absent.

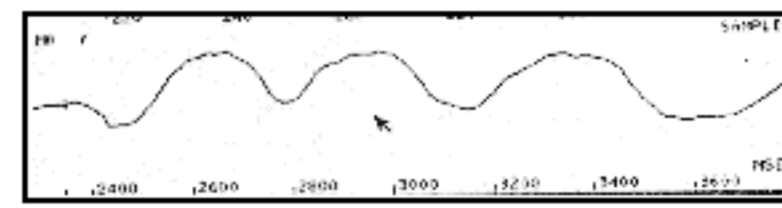
slow rate:

Jaw Height



"I say support today"

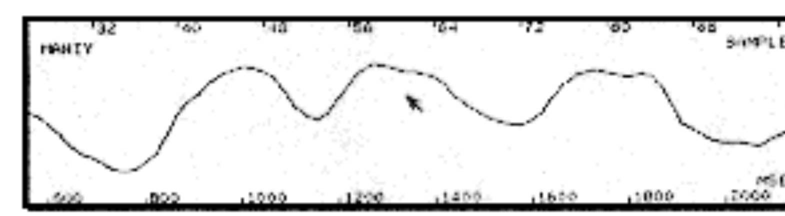
Jaw Height



"I say sport today"

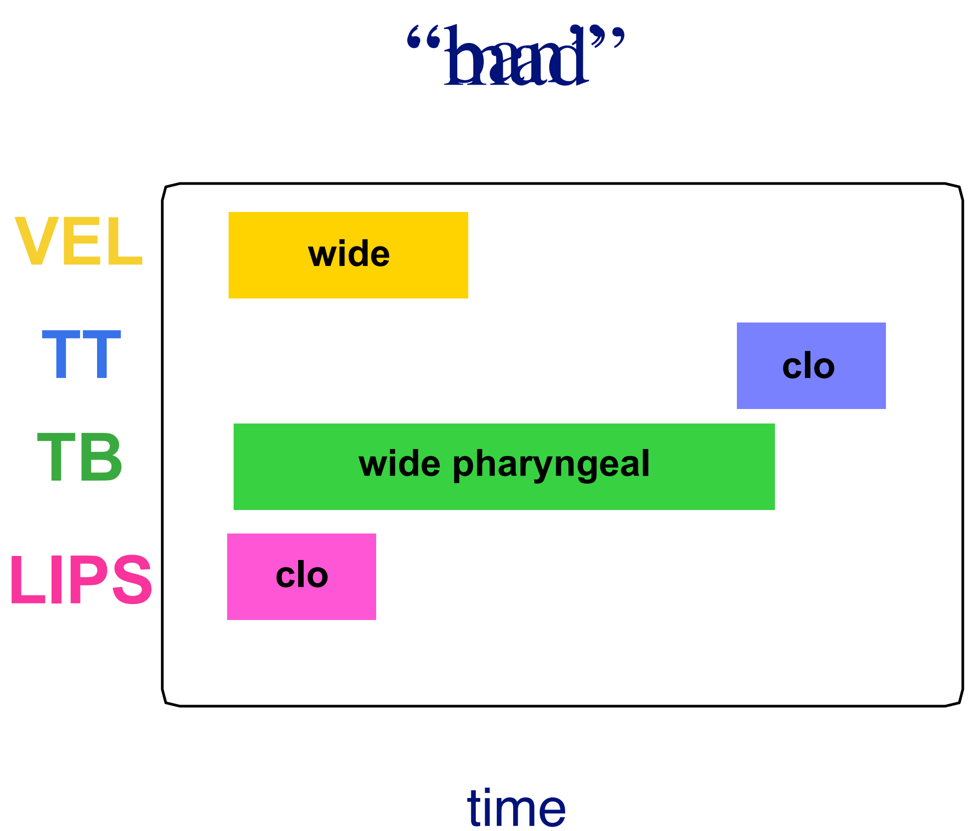
faster rate:

Jaw Height



"I say support today"

Organization of speech into syllables: How are gestures coordinated in time?



- Relative timing of gestures carries information.
- How is appropriate relative timing maintained?
    - Timing needs to be systematic to preserve information
      and flexible to vary with rate and prosody and speaker
- Dynamical systems have these properties
    - The dynamical systems corresponding individual gestures are fixed during their activation intervals and context-independent, but the resulting movements vary flexibly as a function of context.
    - What kind of dynamical system causes the gestures to be activated and deactivated in time?

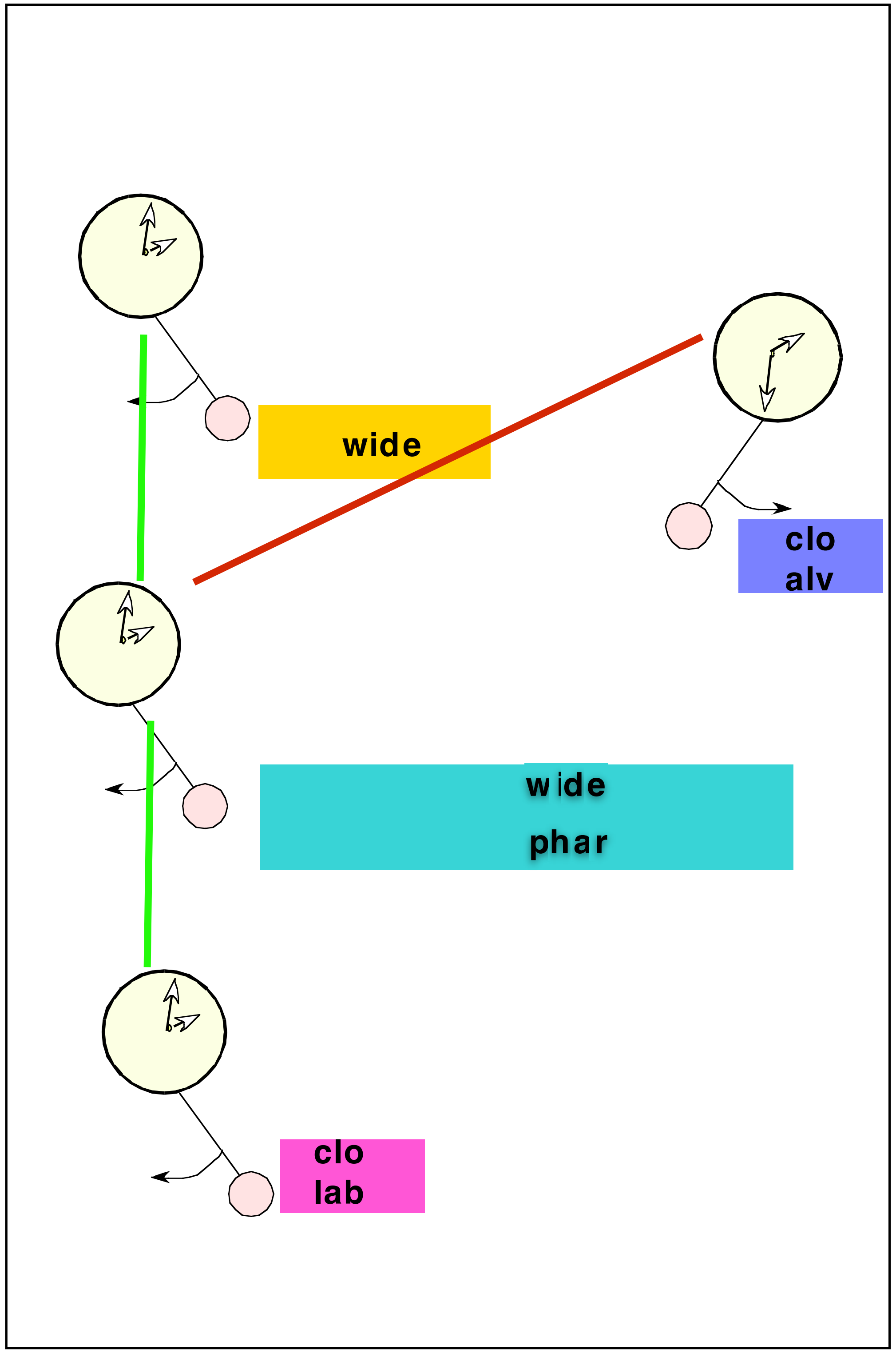
Dynamics of planning intergestural timing:

- Each gesture is associated with a planning oscillator, or clock, responsible for triggering that gesture's activation.

- A clock is a dynamical system: a different kind than the simple point attractor that controls individual gestures; it is a periodic attractor.

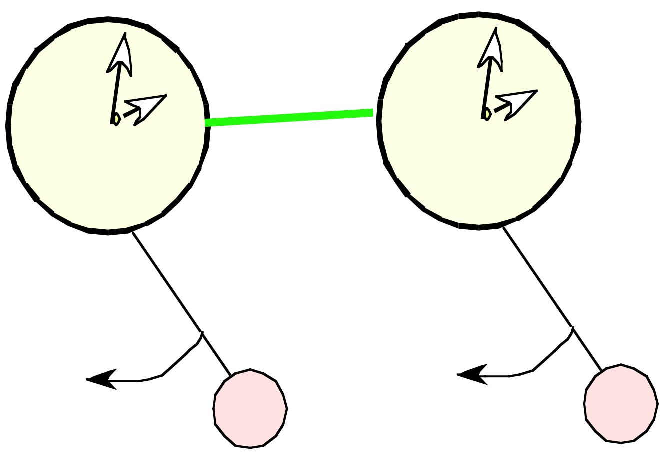
- Relative phase of clocks (and therefore time of triggering) is controlled by coupling the clocks to one another.

- Coupling means that the dynamical equation of one clock includes not only its current state, but the state of the clock it is coupled to.



Why clocks and coupling?

- Clocks are oscillatory dynamical systems.
  They exhibit entrainment:
  They spontaneously synchronize with one another



Huygens' clocks

Demo: Bahraminasab

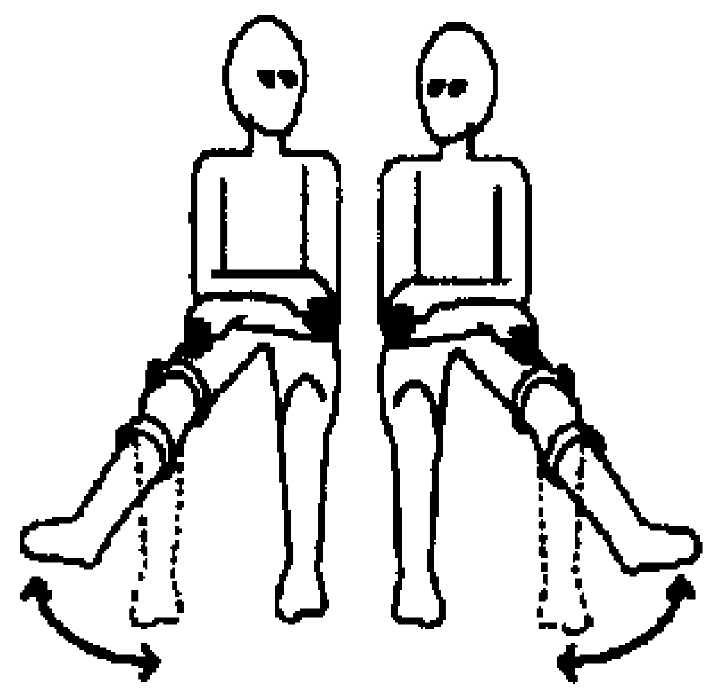
http://www.youtube.com/watch?v=W1TMZASCR-I

Generality of Entrainment

- Applies to living systems, including humans.

http://www.youtube.com/watch?v=4FNoIDgNE6o

- Entrainment of clocks within an individual or across individuals



Turvey, 1990

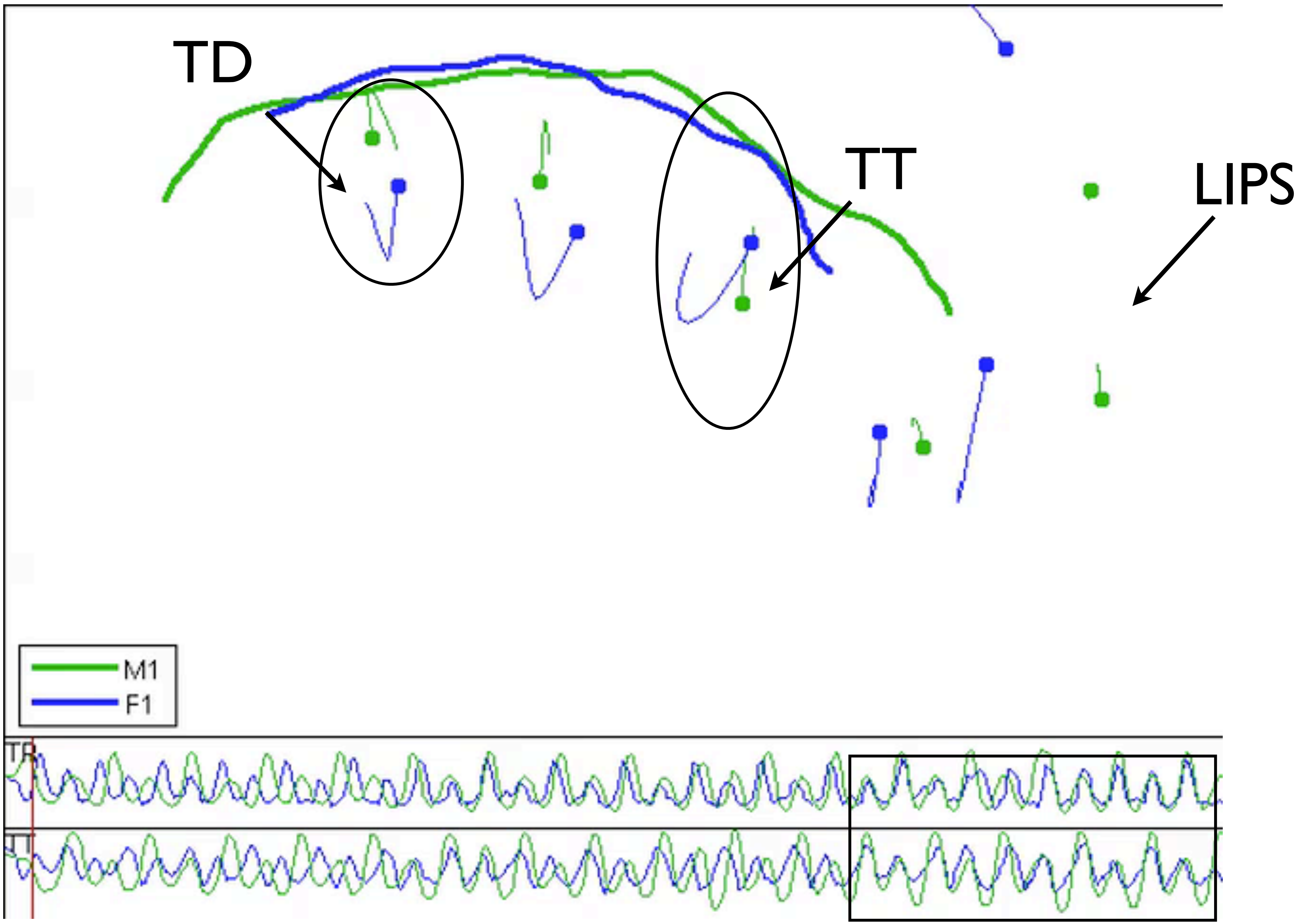
- Coupling doesn’t have to be mechanical.

It can be informational (Saltzman, 1995).

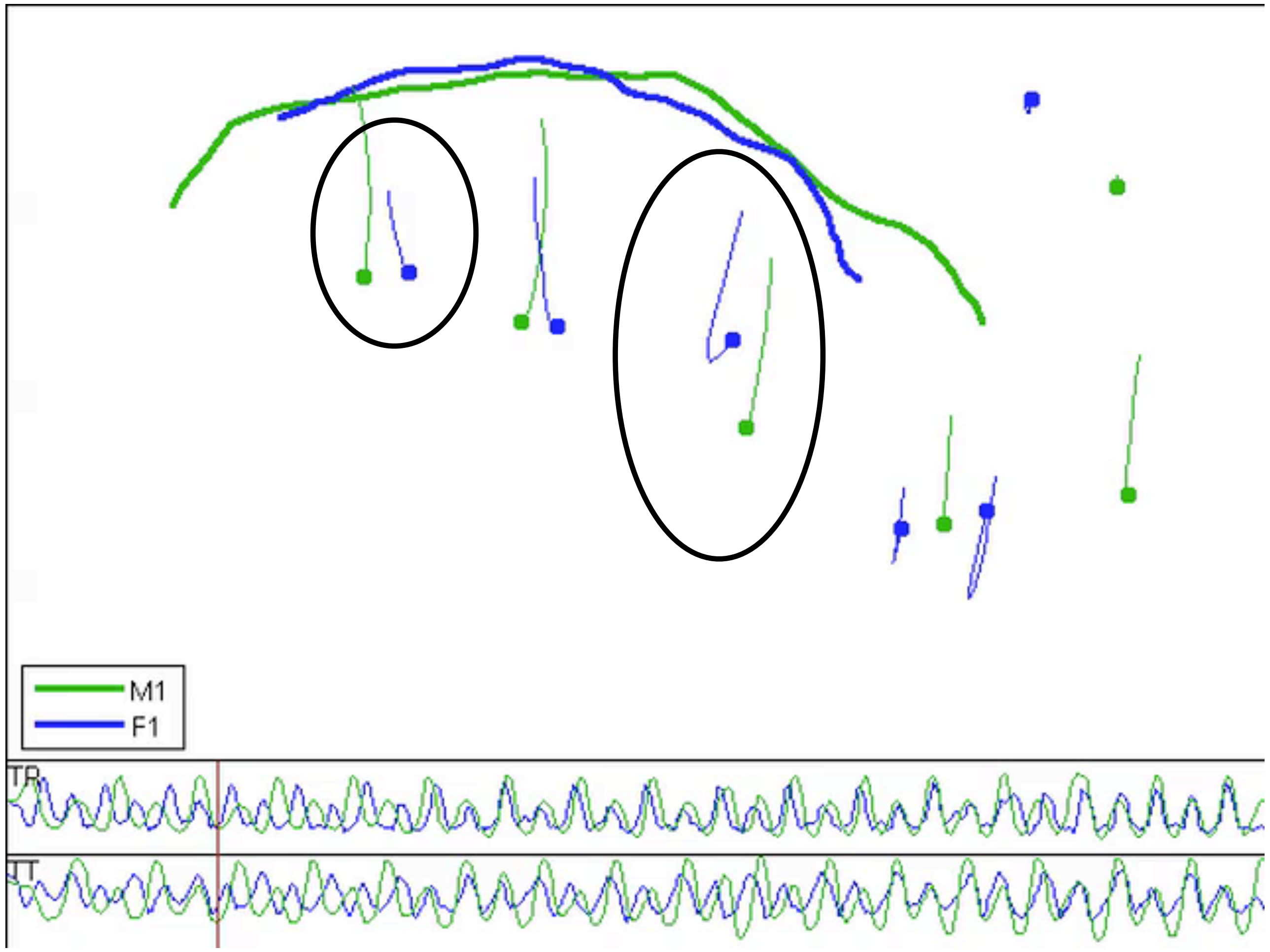
Speech entrainment across talkers

- Articulator kinematics from two talkers simultaneously.
  - Talkers (1 M, 1F) sat 2 m apart facing one another
    - M: "cop top cop top..."
      F ; "top cop top cop..."

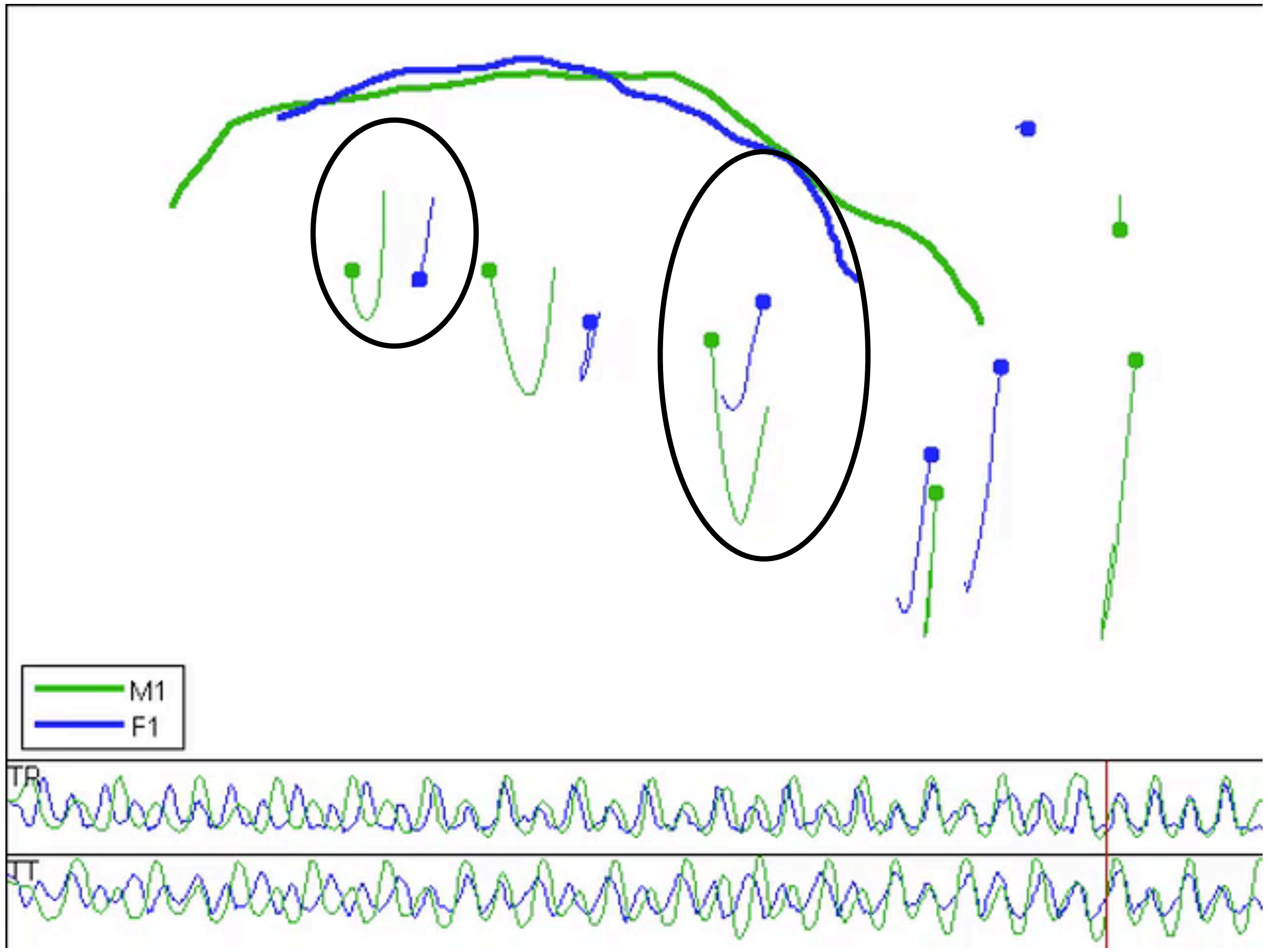
Tiede, Kroos, Bundgaard-Nielsen, Gilbert, Attina, Kasiopa, Vatikiotis-Bateson, Best (2010)


TD
TT
LIPS
M1
F1
TR
TT

BEFORE Entrainment



AFTER Entrainment



Modes of coupling

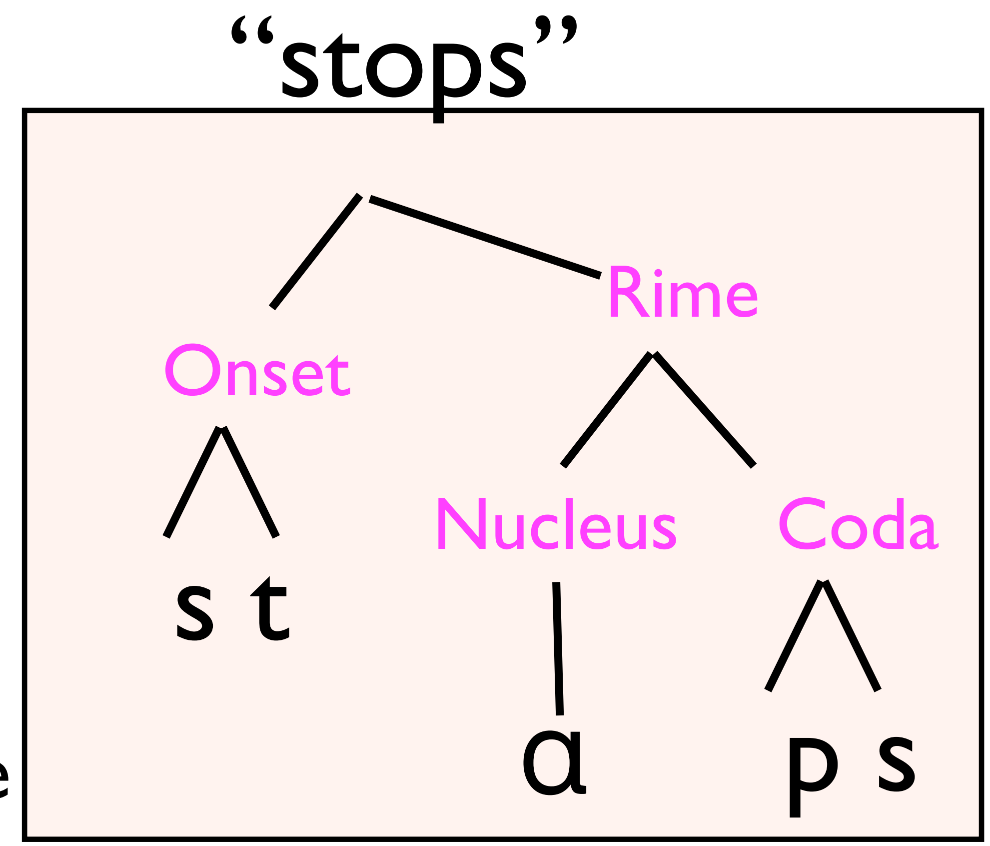
- Systems of coupled oscillators exhibit discrete modes of synchronization:
  - frequency-locking
  - phase-locking

- These modes have been shown to underlie the coordination of movements of multiple limbs in human action. (e.g., Turvey, 1990; Kelso, 1995).

- The same modes can be used to coordinate speech actions and form an account of syllable structure and account for generalizations about syllable structure.

Syllable Structure: Combinations of segments and their gestures

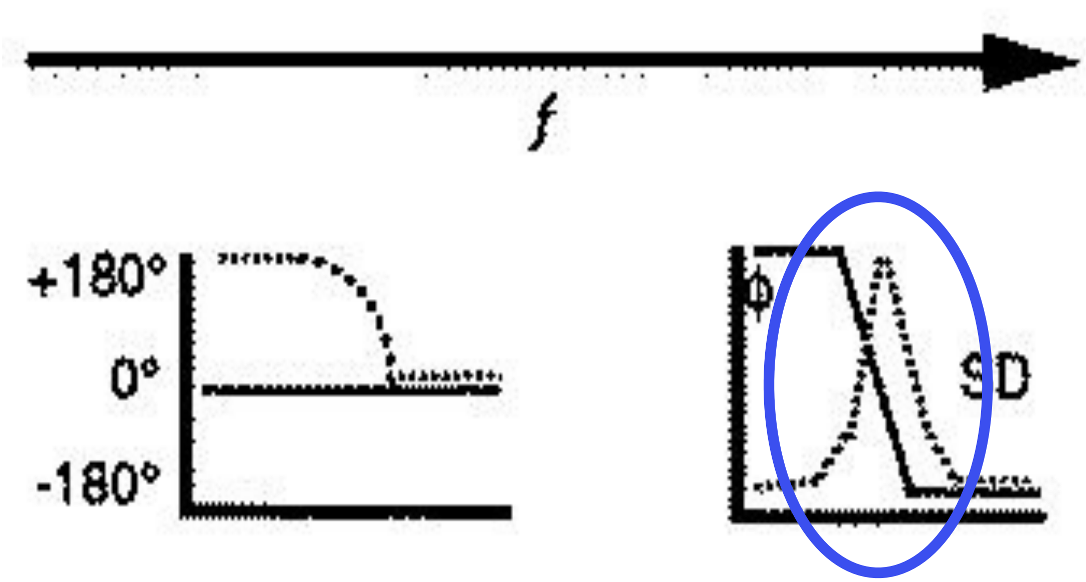
- Two types of units:
  - Consonant (C)
  - Vowel (V)
- Cs and Vs form syllables.
- Combinatorial generalizations:
  - Universality: CV syllables only universal type
  - Freedom:
    Onsets and Rime combine relatively freely in languages.
    Combinations within Onset and within Rime can be more constrained.
  - Acquisition: CV combinations acquired by child earlier than VC
  - Weight:
    Onsets rarely contribute to weight. Codas frequently do



Synchronization modes for limb coordination: phase-locking

- Two relative phase modes (or attractors) are *spontaneously* available (require no learning) Haken, Kelso & Bunz, 1985
  - 0˚ (in phase) most stable, accessible
  - 180˚ (anti-phase)

- Oscillation frequency (rate) is a control parameter:
  - Spontaneous transitions to most stable mode (0˚) as frequency increases.
  - Fluctuations in phase during transition interval.

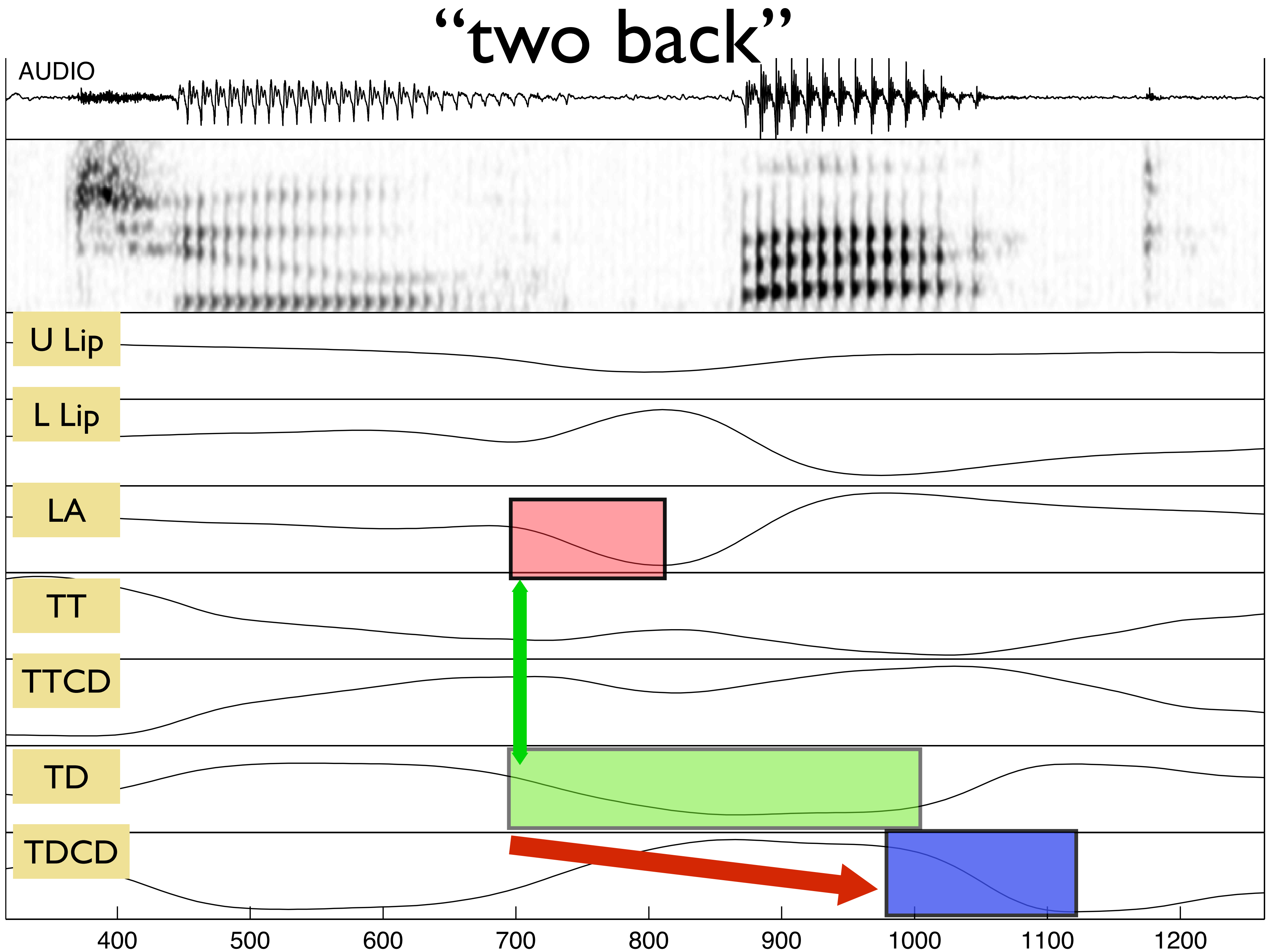


Turvey, 1990

Modes & Syllable Structure

- 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) simplest, most stable, accessible
  - anti-phase
    - hypothesized for V-C (coda)

Evidence for C-V and V-C modes



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

Combinatorial Freedom

- Combination is free where the coupling mode is maximally accessible without learning (in-phase).
- Combinations are most restricted where learning is required.

Freedom

CV: onset-rime	in-phase	no learning
VC: nucleus-coda	anti-phase	no learning, but less accessible
CC: within onset,coda	other	particular combinations must be learned

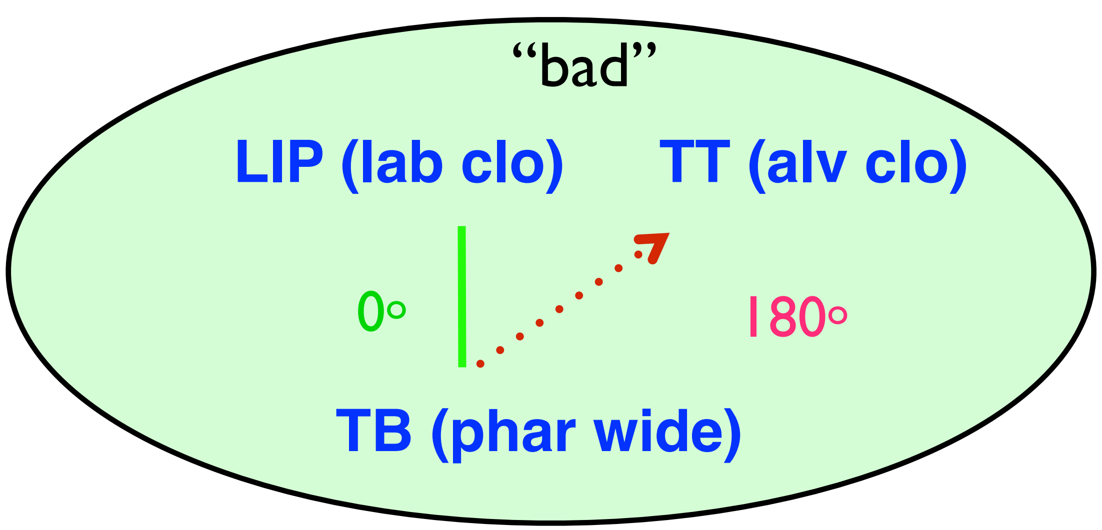
Accessibility

Acquisition of syllable structure
(Nam, Goldstein & Saltzman, 2009)

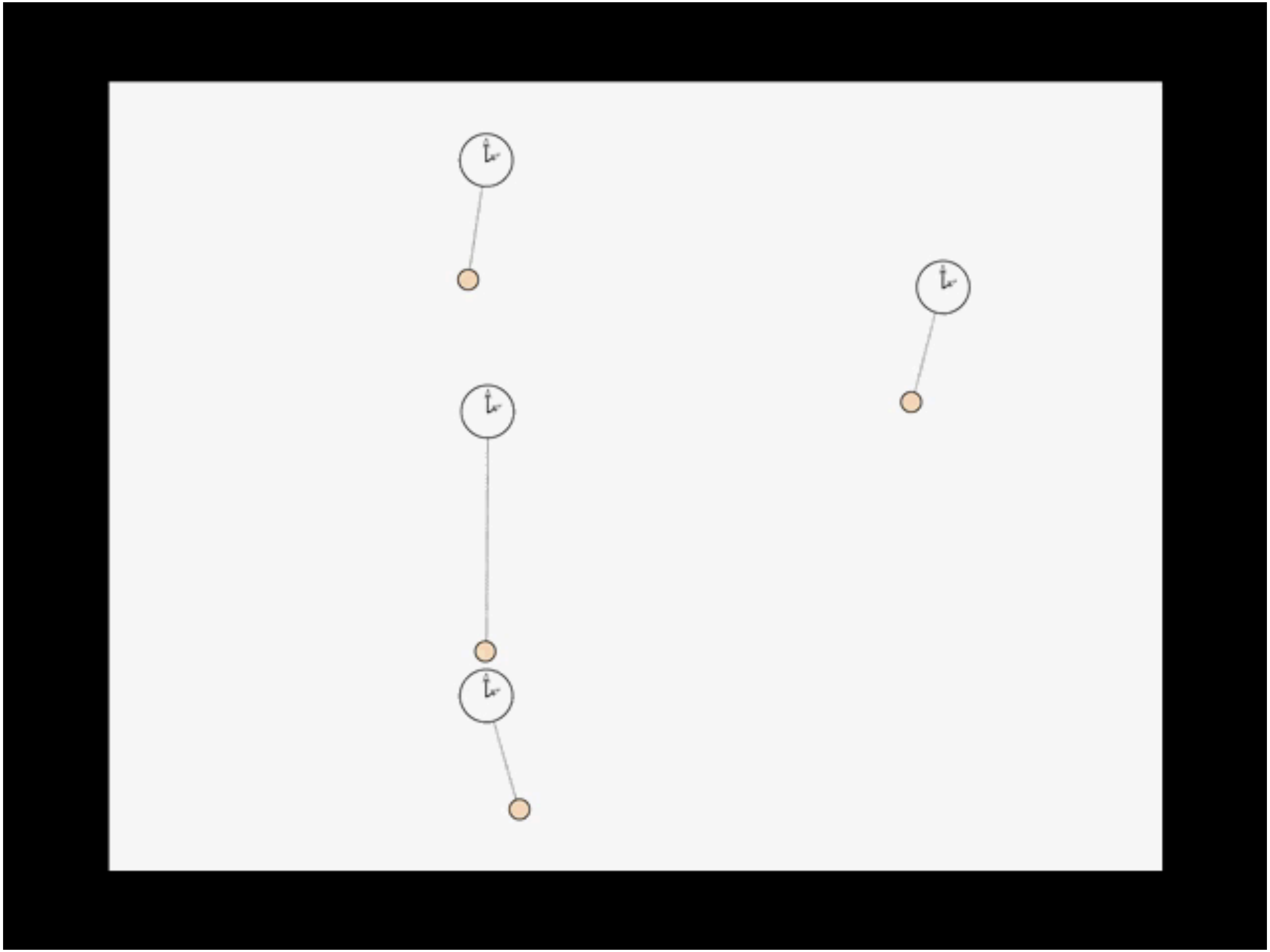
- Infants develop onsets (CV) before codas (VC) in all languages. (e.g.Vihman & Ferguson, 1987; Fikkert, 1994)
- Lag in acquisition of codas is shorter in languages that make more frequent use of VC (Roark & Demuth 2000).
- These facts are predicted by a model of a learning agent that includes both: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

Planning model (Goldstein, Byrd, Saltzman, 2006)

- 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).



Timing and lg-specific syllabification: CC clusters in onset

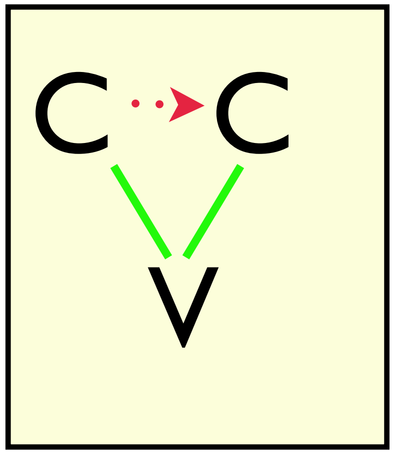
- If onset is defined by an in-phase relation between C gesture and V, then all onset C gestures should be synchronous with V (and therefore with each other).

- Multiple constriction gestures in onset cluster (e.g., “spats”).
  - Gestures must be at least partially sequential to afford perceptual recoverability.
  - Some languages, (e,g., French) contrast /sp/ and /ps/
  - What in the coupling graph identifies them as both in the onset??

Competitive coupling hypothesis

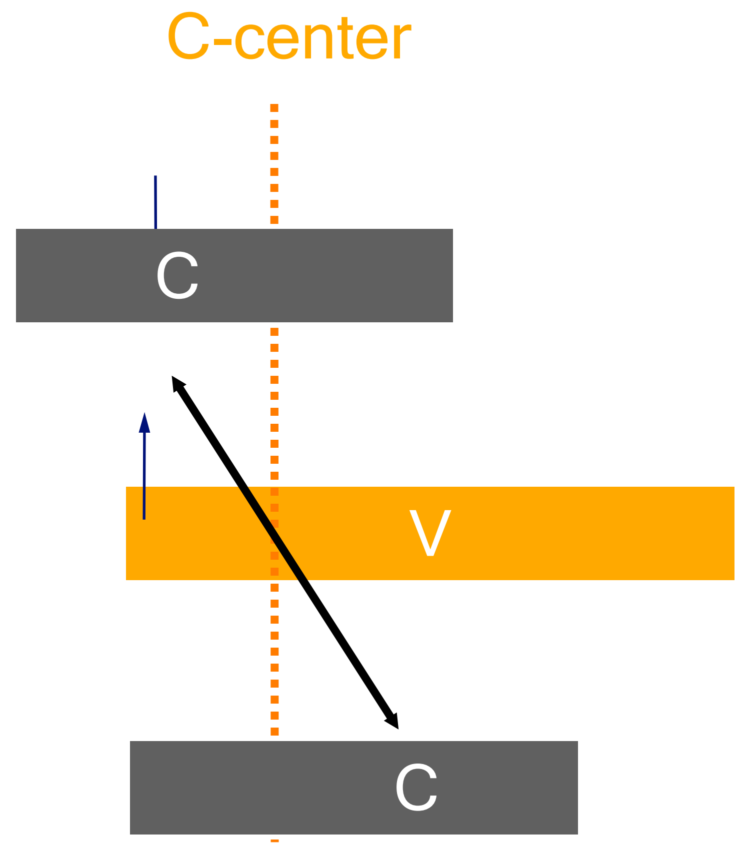
(Browman & Goldstein, 2000)

- 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 (eccentric)
- These cannot both be realized, so coupling is competitive

Consequences of competitive coupling



results of competition

C-V phasing

C_1 shifts left

C_n shifts right

But C-V phasing is preserved

Add an additional coordination

(C-C phasing)

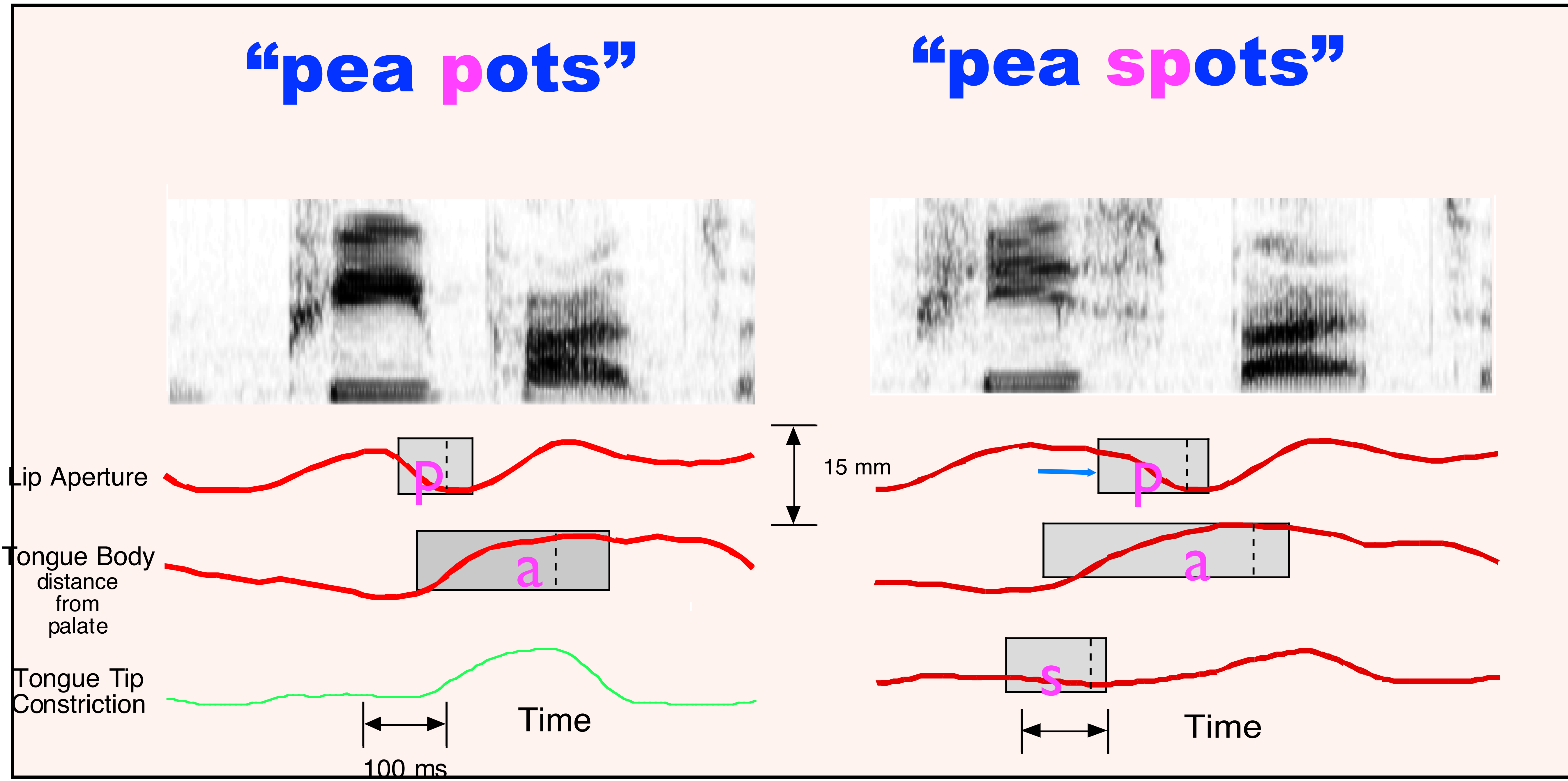
C-V and C-C phasings

in competition

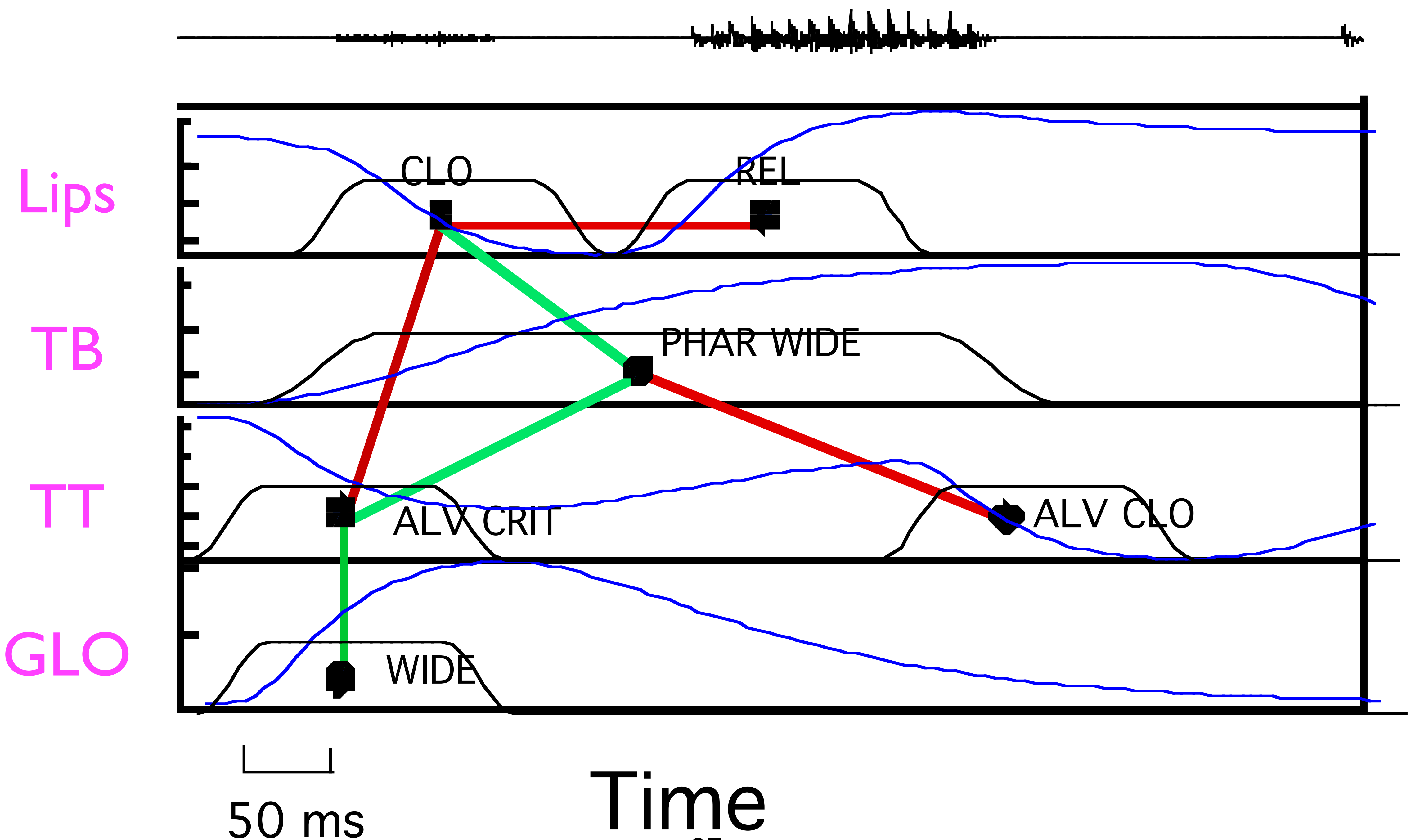
global C-center

- 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.

Evidence for Rightward shift of C_n



Example: “spat”



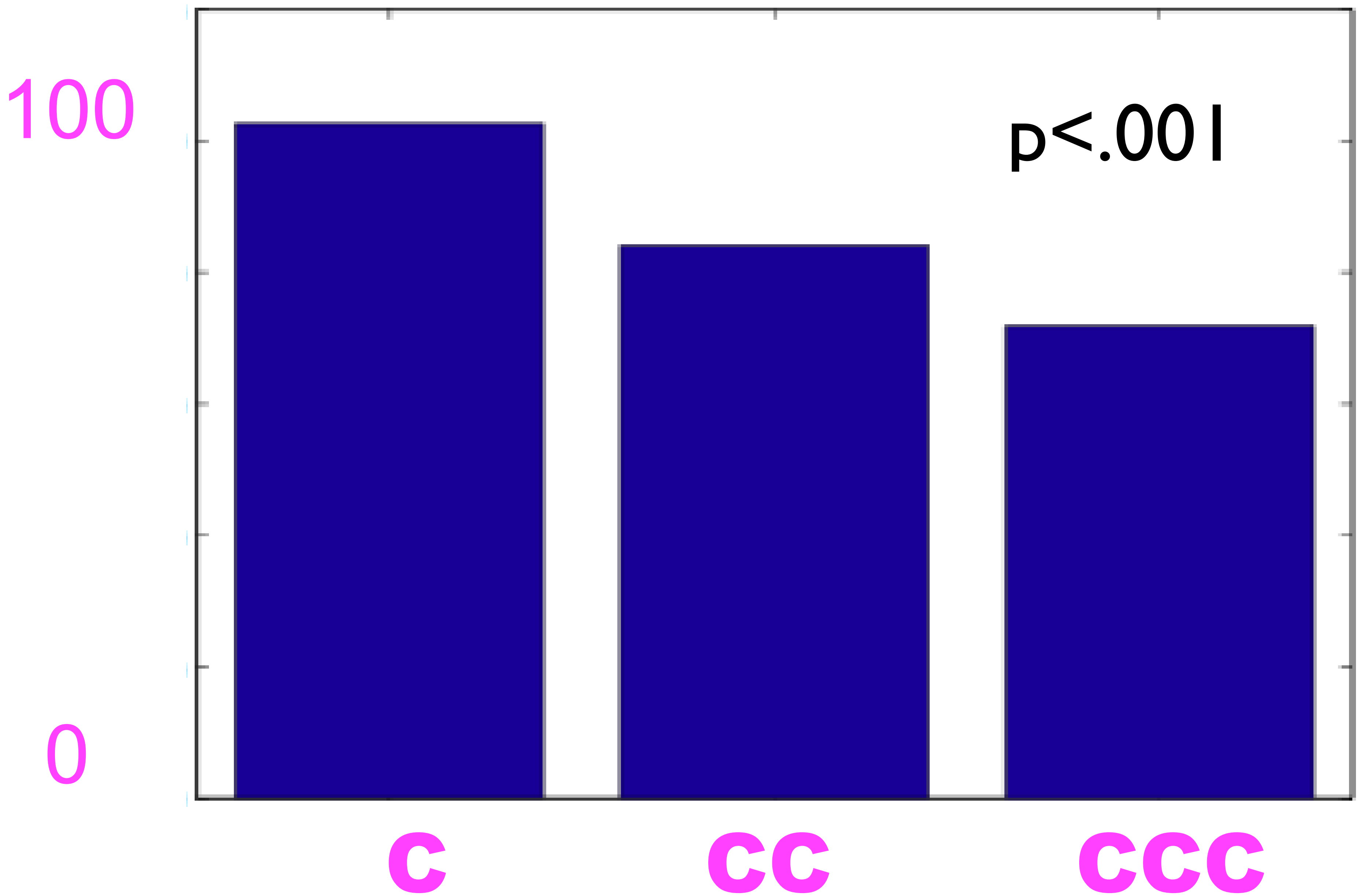
Rightward shift of C_n as diagnostic for complex onset?

- 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.
- But they 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.
- Georgian exhibits rightward shift, but Berber does not.

Georgian: Rightward shift EMA data

(Goldstein, Chitoran, Selkirk, 2007)

Lag: Target (V) - Target (C_n)



/**k**arebi/ /**r**iala/

/**t^s^k**arebi/ /**k'r**iala/

/**pt^s^k**aredi/ /**t^s'k'r**iala/

/p/ Lip Aperture

/t/ Tongue Tip Constriction Degree

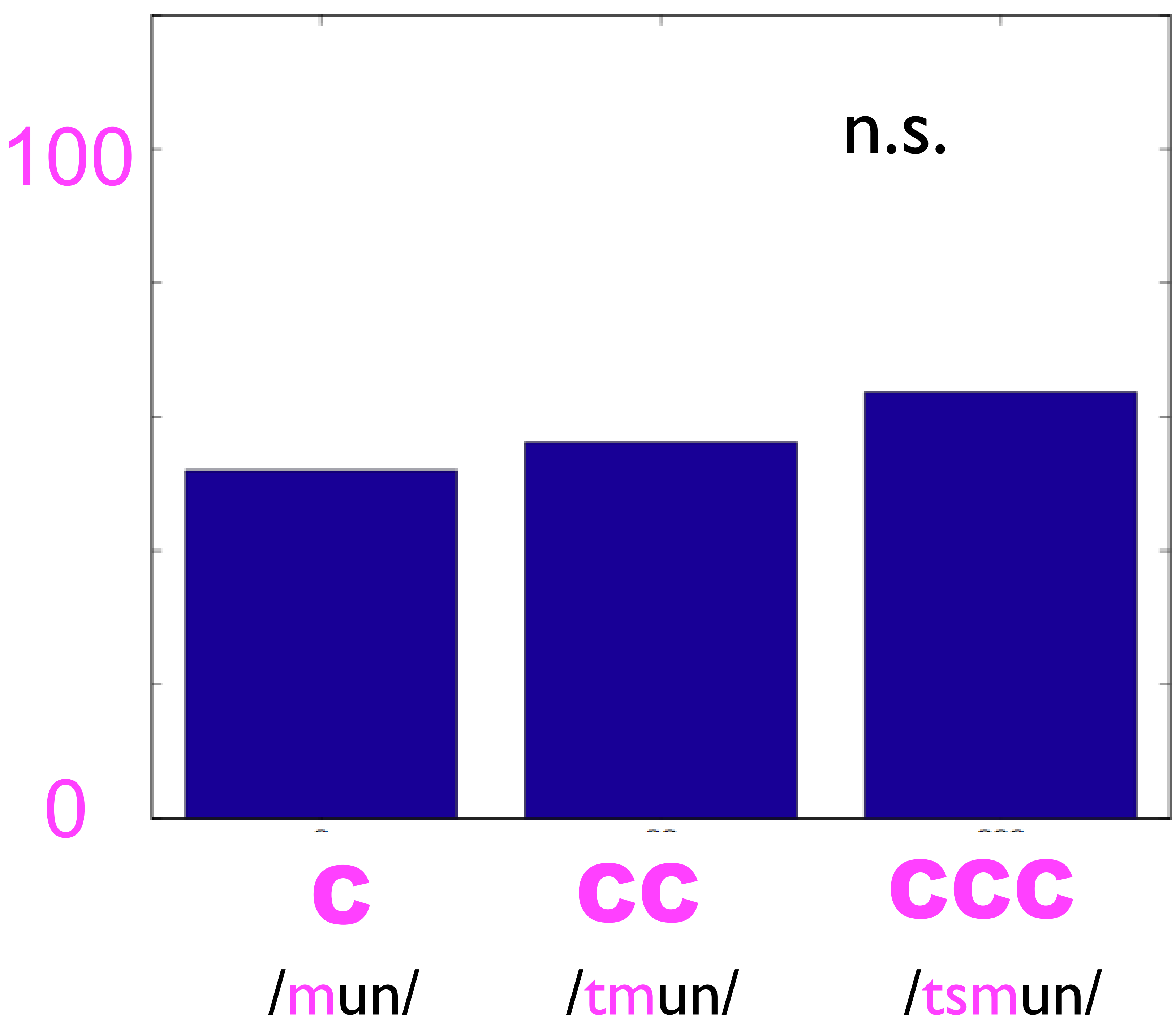
/k/ Tongue Dorsum Constriction Degree

/V/ Tongue Body Constriction Degree

2 speakers

Tashlhiyht Berber: Rightward shift EMA data

Lag: Target (V) - Target (C_n) ms.



/m/ Lip Aperture

/t/ Tongue Tip Constriction Degree

/s/ Tongue Tip Constriction Degree

/V/ Tongue Body Constriction Degree

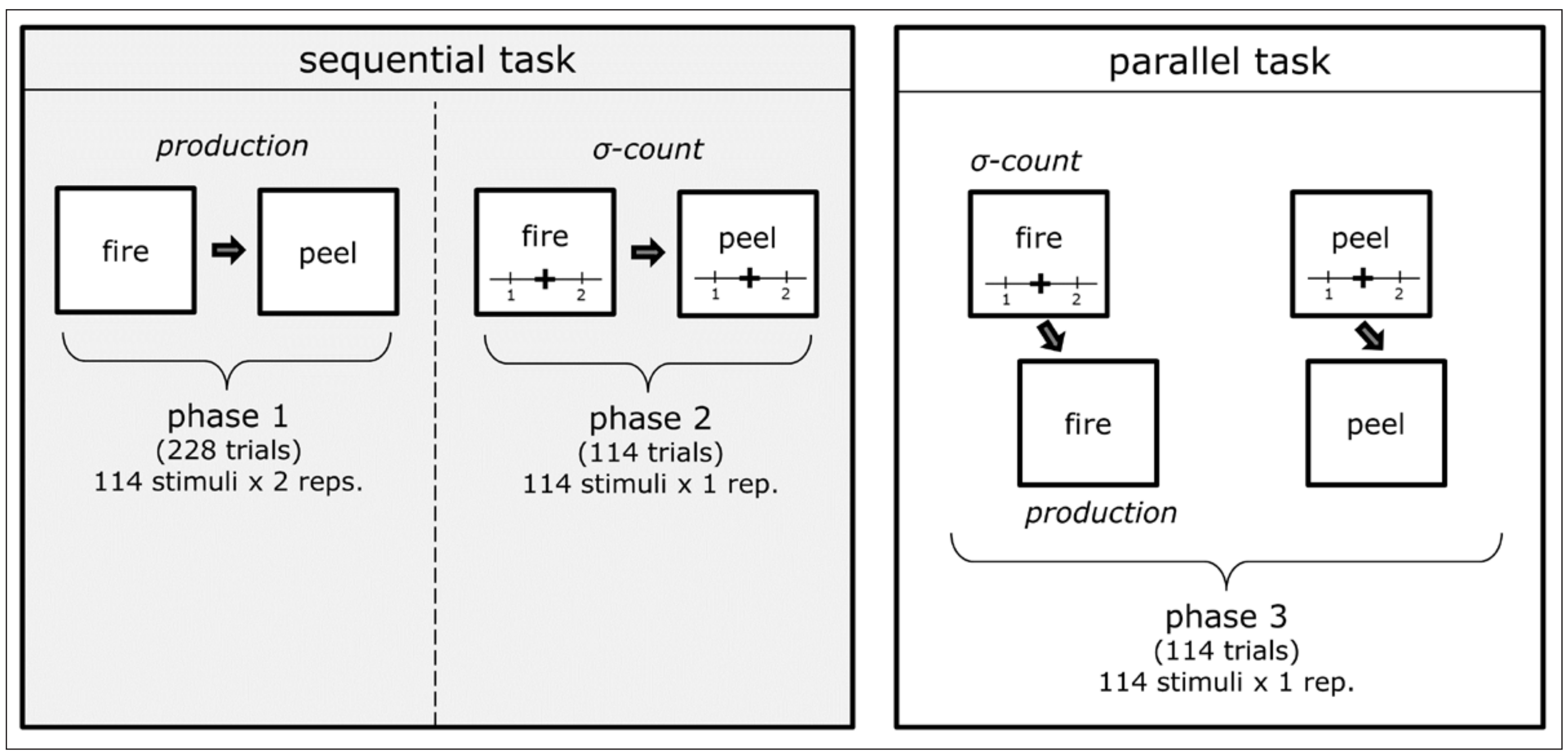
2 speakers

Evidence for complex onsets

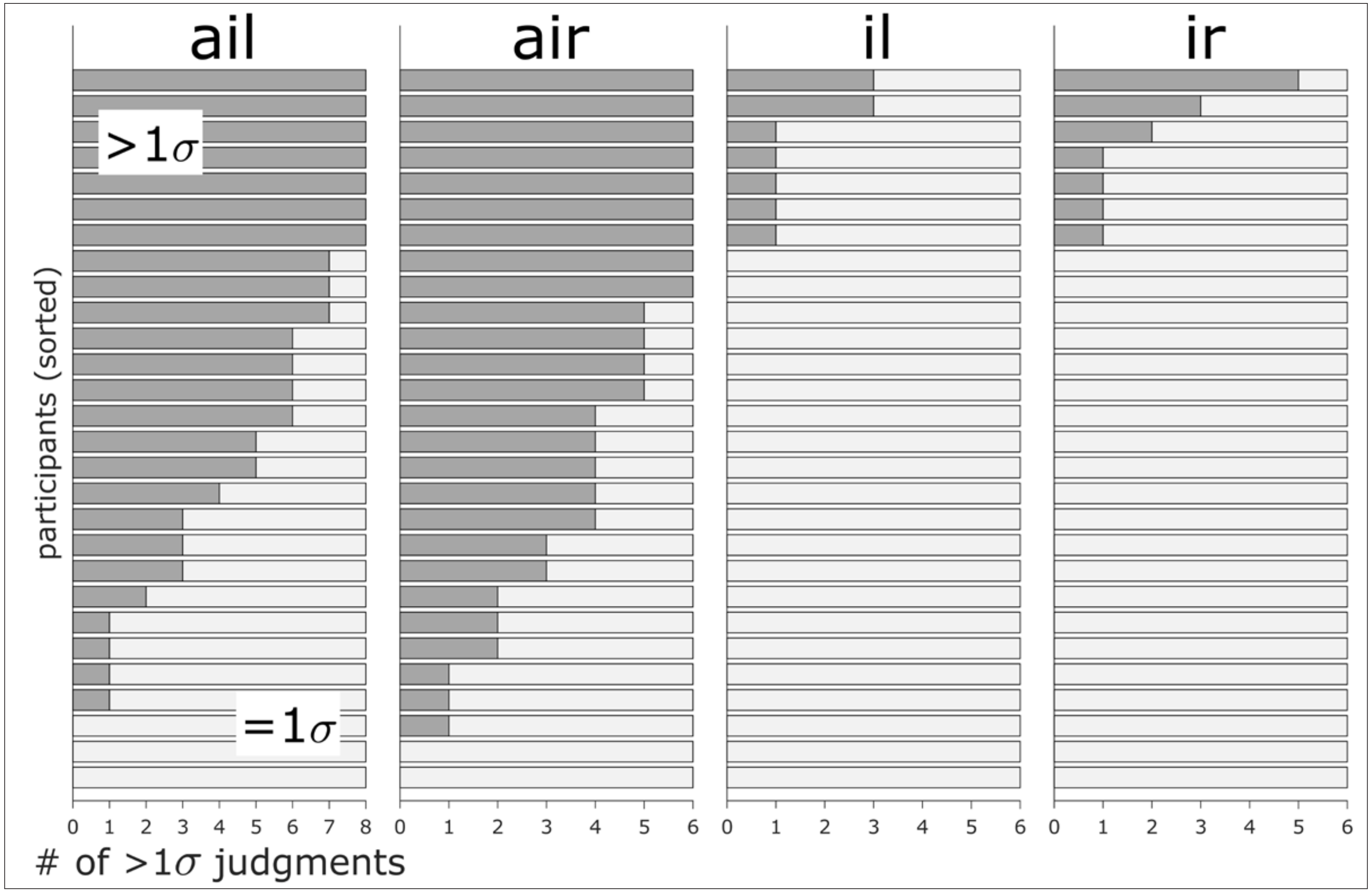
- Differences between Georgian and Berber provide tentative support that c-center can be used as diagnostic for complex onset structures.
- Support from other languages:
  - Italian: /sC/ vs. /Cl/ (Hermes, 2013)
  - Moroccan Arabic (Shaw et al. 2010)
  - Romanian (Marin 2012)
  - Chinese (tones & consonants) (Gao, 2008)

Relation of ambiguous syllabification to gestural coordination (Tilsen & Cohn, 2016)

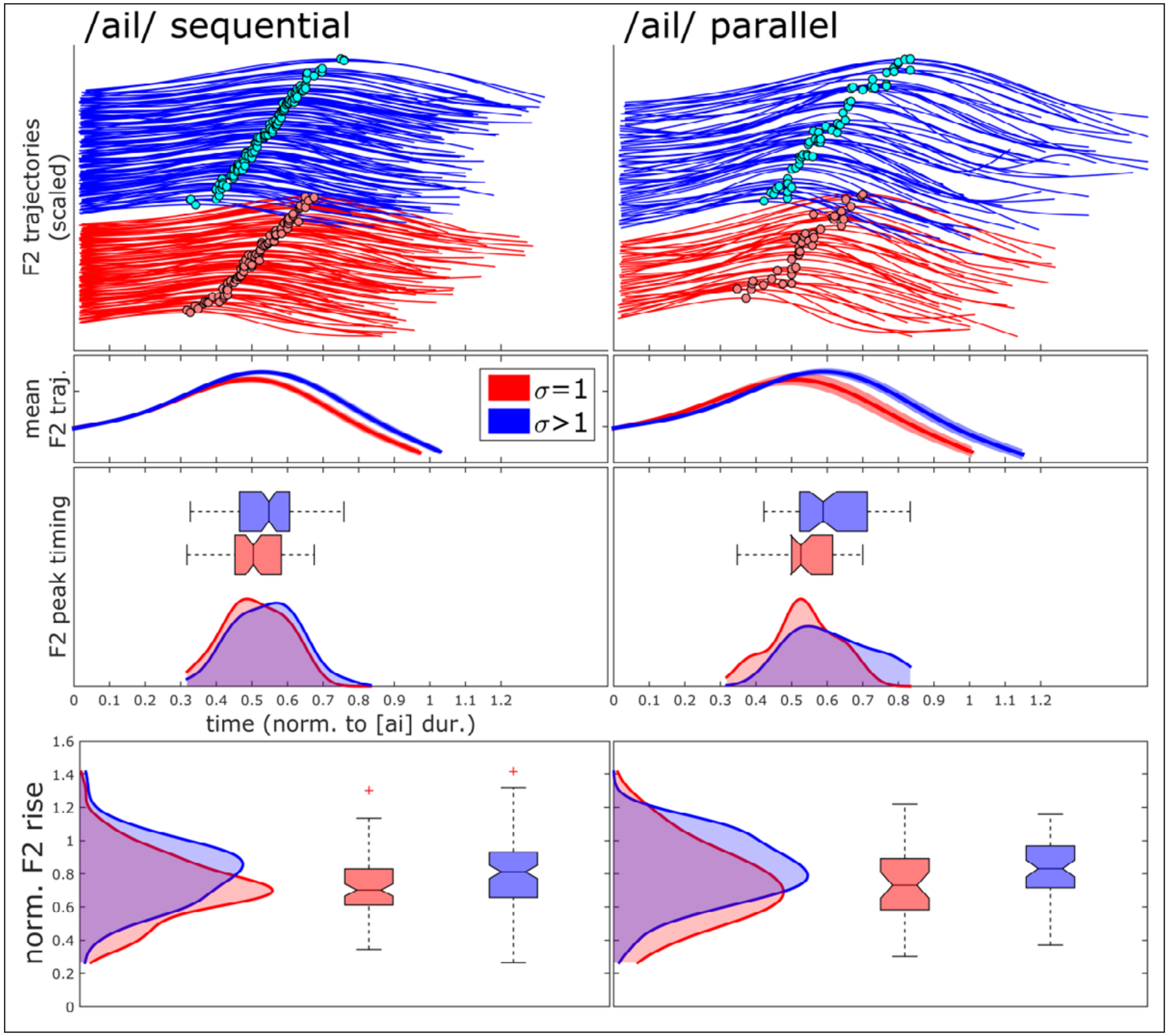
- Participants produced words like "peel, fire"and then gave syllable count judgements (using slider, they could choose value >1 but <2.



Syllable judgments

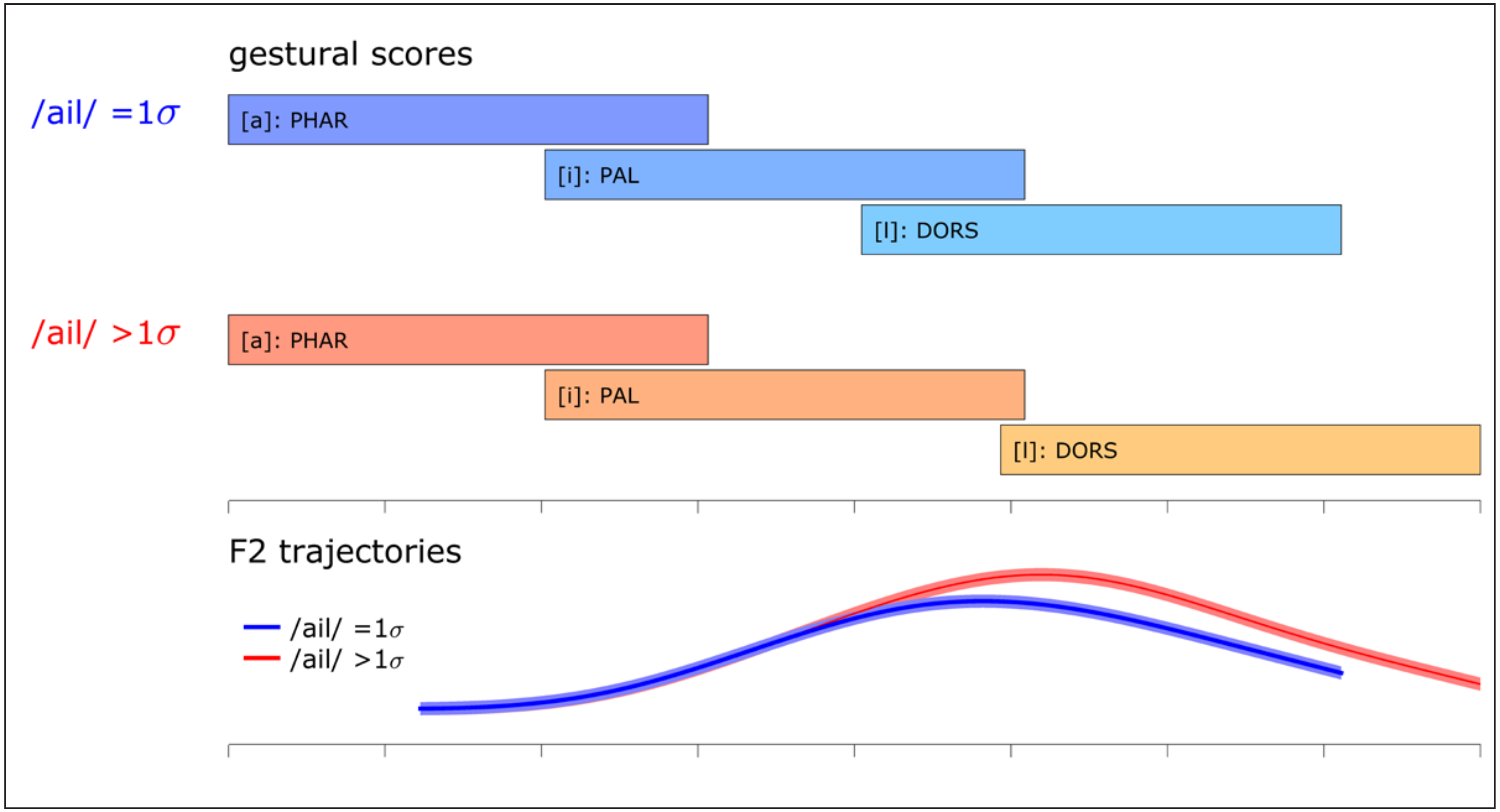


Relation of Production to Syllable Count



F2 peak timing	**Δ%**	***t* (df) =**	***p*-value =**
/ail/	6%	2.2 (267)	0.030
/air/	14%	2.7 (165)	0.008
F2 rise			
/ail/	6%	9.9 (257)	0.002
/air/	9%	1.1 (160)	0.005

Formant differences: Hypothesized gestural timing

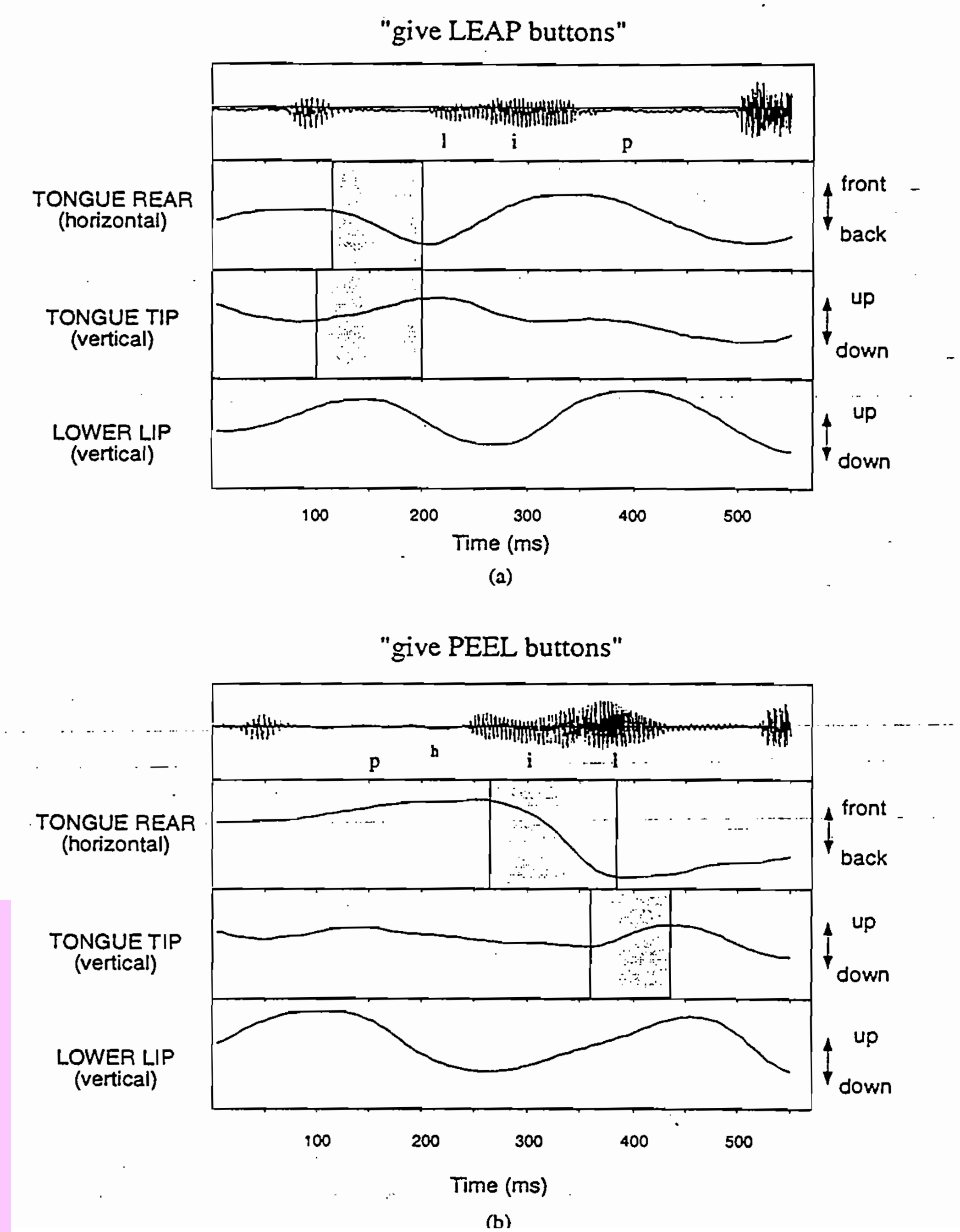


Dorsal gesture of /l/ overlaps vowel when it is in coda

Coupling graphs are presumably different

Velarization of /l/

- English /l/ is described as “dark” or “velarized” in coda, and “brighter” not velarized in onset.
- The gestures in the two positions are in fact very similar, but the timing is different.
- In coda, the retraction of the TB occurs first and contributes to the “velarized” percept.
- Pattern in very similar to that for nasals.



Principle 2: Coordination in onset vs coda in English
Onset: all gestures composing a C begin synchronously
Coda: gestures composing a C can be sequential, with wider constriction leading

leap

TB (l)

TT (l)

TB (V)

TB (l)

?

peel lσ TT (l)

TB (V)

Non-vocalic nuclei?

C and V gesture valences

- C and V gestures are differentiated by
  - degree of constriction (V is wider)
  - dynamic stiffness (V takes longer to get to target)
  - activation interval (V still active after C released)
- Nature of these differences is such that C and V gestures can be triggered synchronously and still be both be recoverable by listeners (Mattingly, 1981).
- These gestural properties, together with the stability of in-phase coupling gives rise to valence of C and V gestures -- they combine freely with each other in C-V structures.

What gestures can serve as syllable nuclei?

Voiceless vowel-less syllables:

phonetic and phonological evidence from Tashlhiyt Berber

Rachid Ridouane
LPP - UMR 7018,
CNRS – Sorbonne Nouvelle Paris III
19, rue des Bernardins 75005 Paris

Is there a language in which any segment can be syllabic?

- One such language is **Tashlhiyt Berber** (TB) (Dell & Elmedlaoui 1985, Prince & Smolensky 1993, Zec 1995, Clements *1997*).

Examples

- Voiced fricative : /tsbGt/ [ts.b**G**t] « you paint »
- Voiceless fricative : /t-sti/ [t**s**.ti] « she chose »
- Voiced stop : /t-g°ra/ [t**g°**.ra] « she took »
- Voiceless stop : /tkmi/ [t**k**.mi] « she

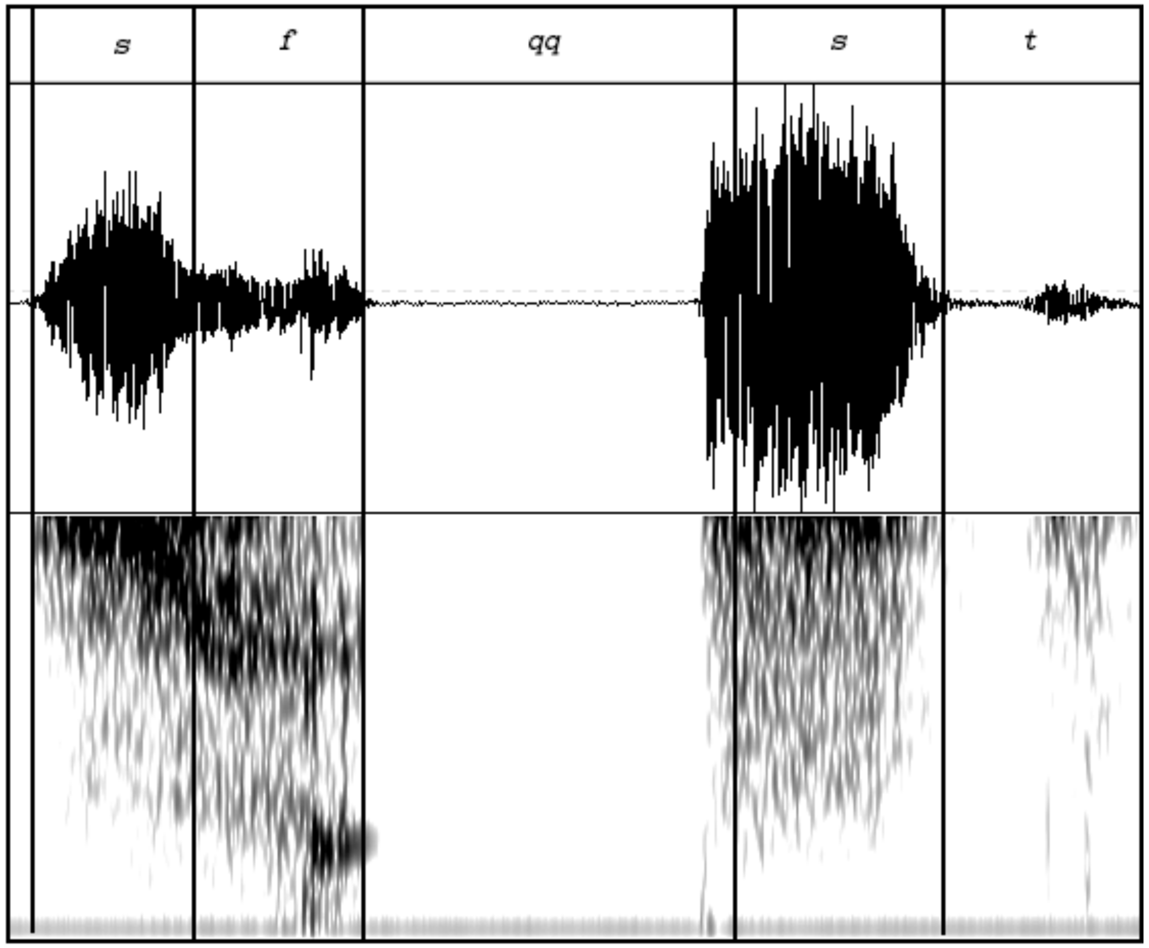
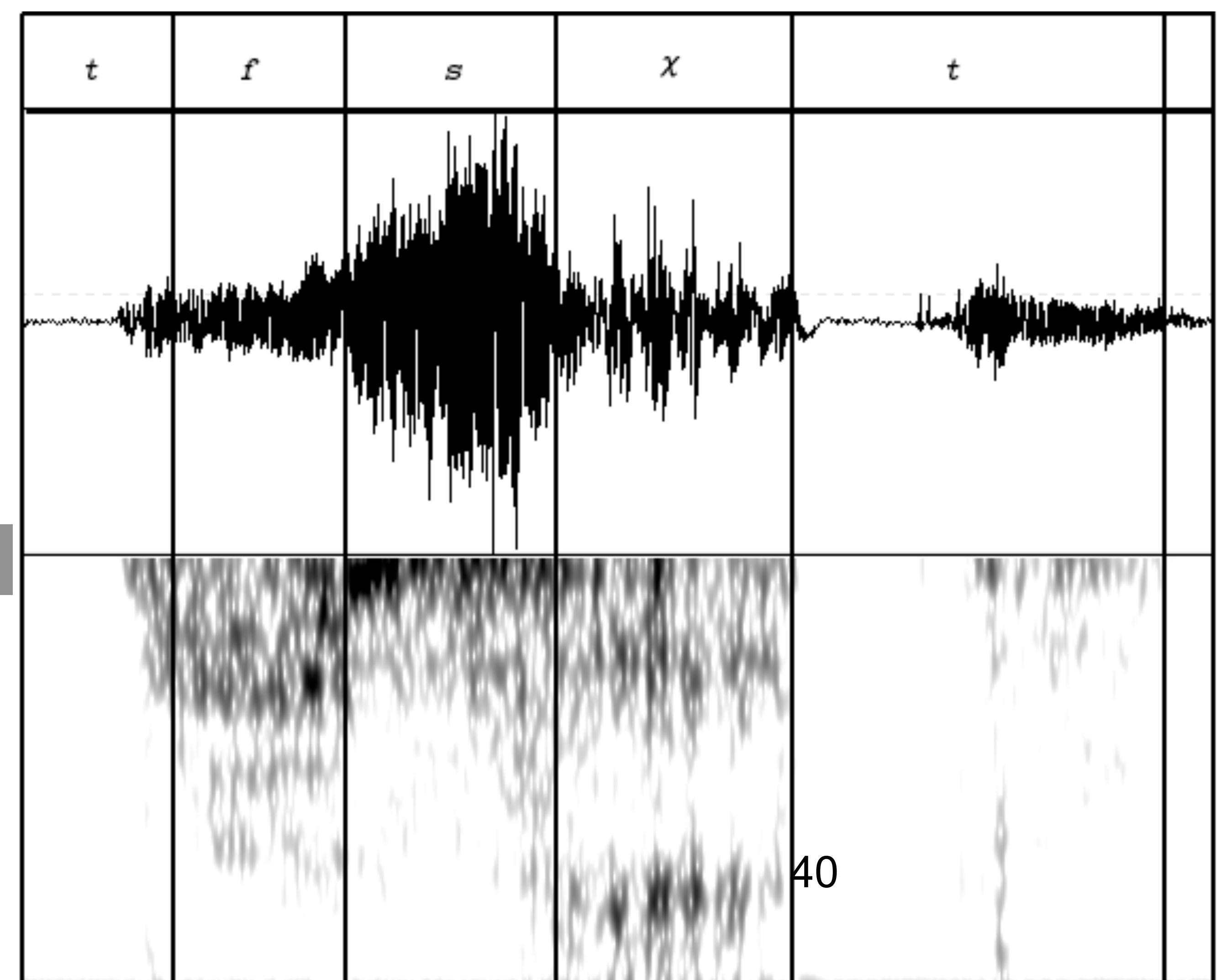


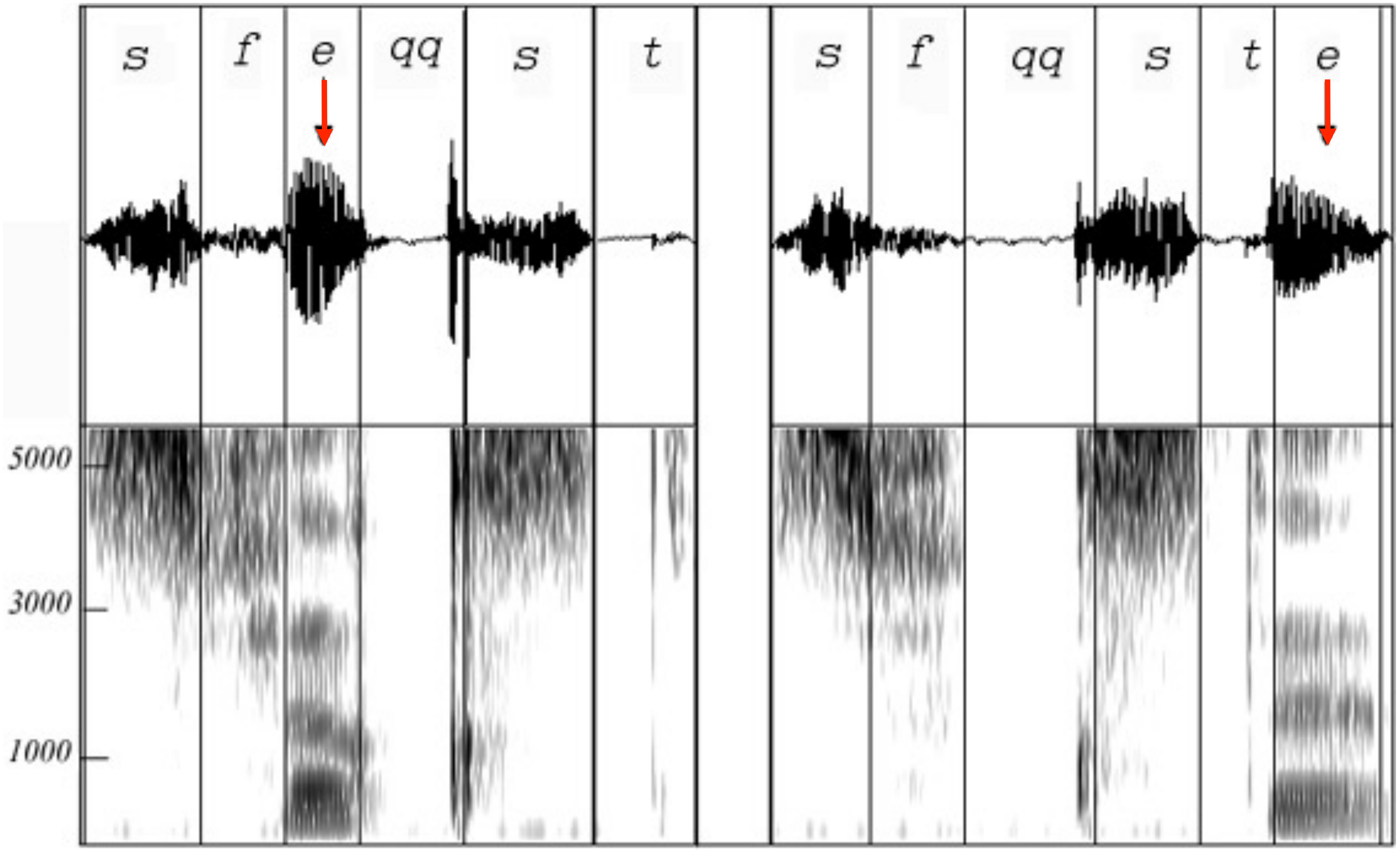
Figure 1. Audio signal and spectro of one repetition of [sfqqst] by R_R

*Figure 2. Audio signal and spectro of one repetition of [tfsxt] by A*_R



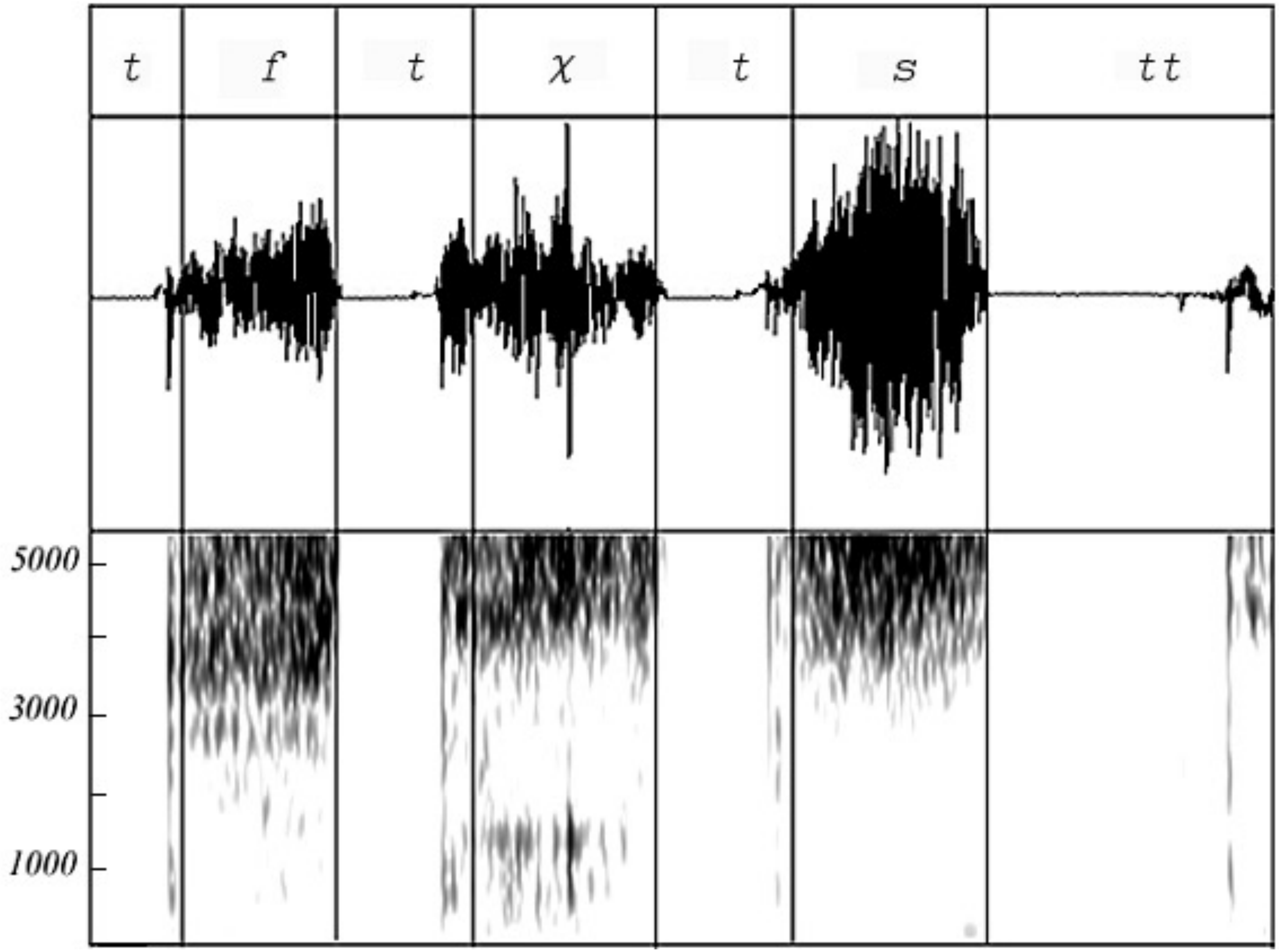
Schwa Unstable

- within one same subject and one same form.



Acoustic waveform and spectrogram of two repetitions of sfqqst « irritate him » by E_M one with internal schwa (left) and one with final schwa (right).

- This same subject may realize long voiceless sequences with no vowel at all. This is the case for instance for the items *fqqs, tfktstt* and *tftXtstt*.



Acoustic waveform and spectrogram of one repetition of [tftXtstt] « fadeaway » by E_M.

Evidence for syllabication: Versification

- TB versification distinguishes between heavy and light syllables (Jouad 1983).

- In TB poetry it is common for all the lines of a piece to be sung to the same tune.

- If a tune is comprised of *n* successive notes, a text with more than *n* syllables cannot be sung to it.

- (See Dell & Elmedlaoui 2002 and the references therein)

Sequences sung to a tune

I present below the parsing of three lines of *Immi nna* "my dear mother" sung by **Iznzarn** *(1970)*

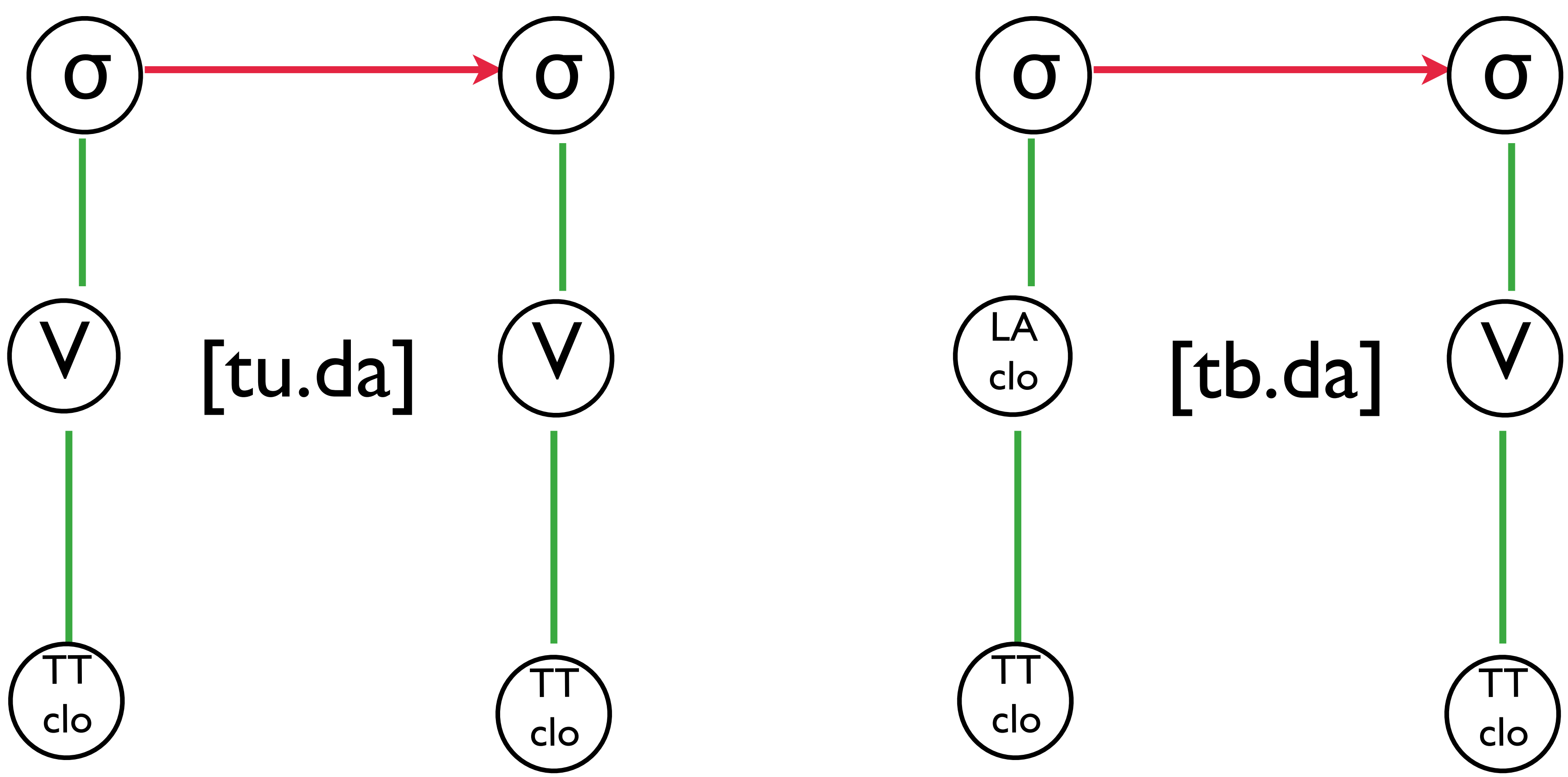
1	2	3	4	5	6	7	8	9	10	11	12	13
L	L	L	L	L	H	H	L	L	L	L	L	L
im	mi	n	na	kf	rab	biG	ta	sa	wa	la	wu	li
tz*	da	tz	dm	ta	g°md	Glh	ma	wa	la	sm	mi	di
ad	da	ln	sa	sm	mid	wat	sm	ms	tr	fl	la	Gi

* This syllable and the following three are emphatic

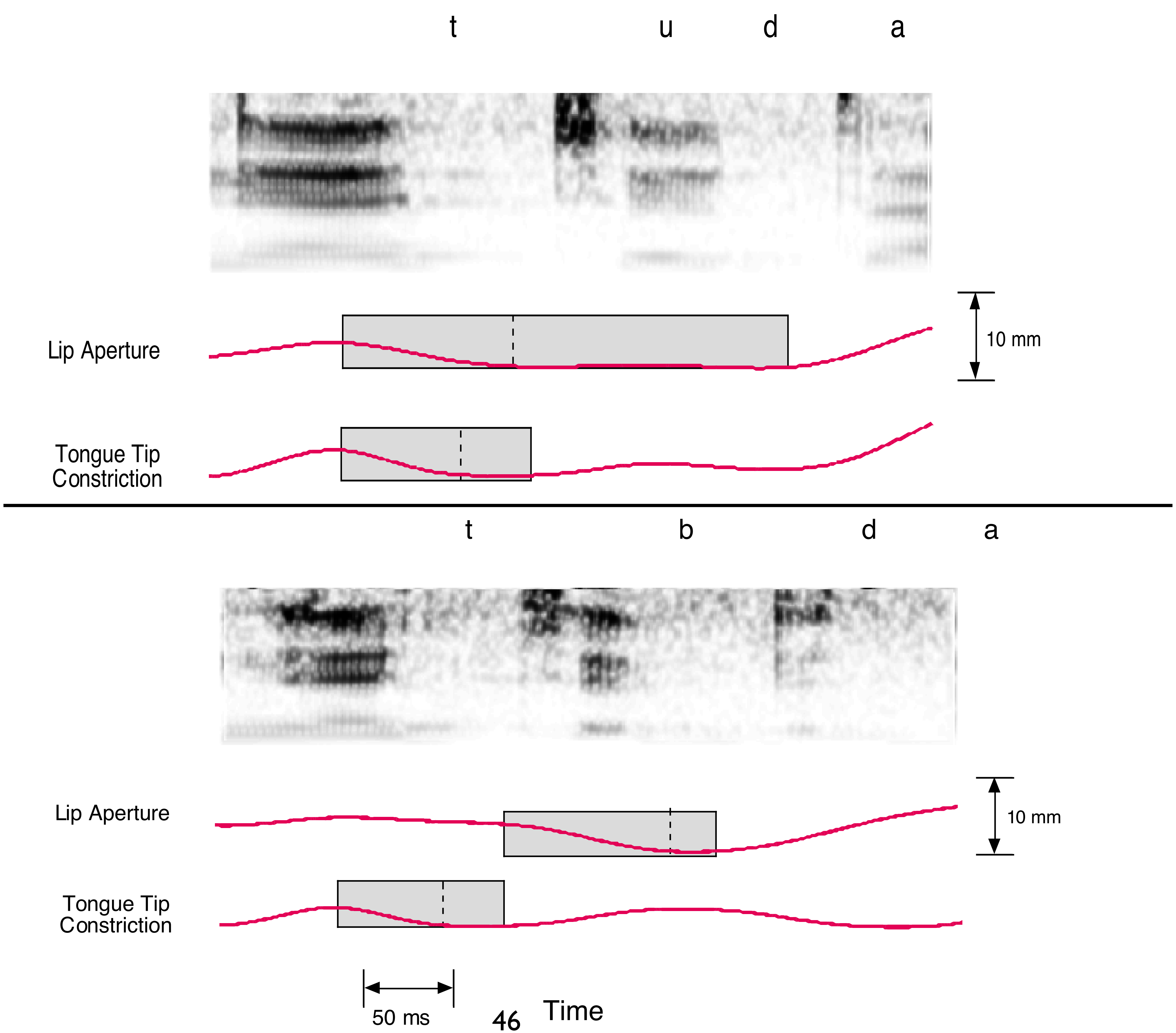
In-phase C-C syllables in Berber?

- Any segment may appear appear as nucleus in Tashlhiyt Berber (Dell & Elmedlaoui, 1985).
    - [tu.da] ‘suffice‘ [tb.da] ‘begin’

- Expected graphs?

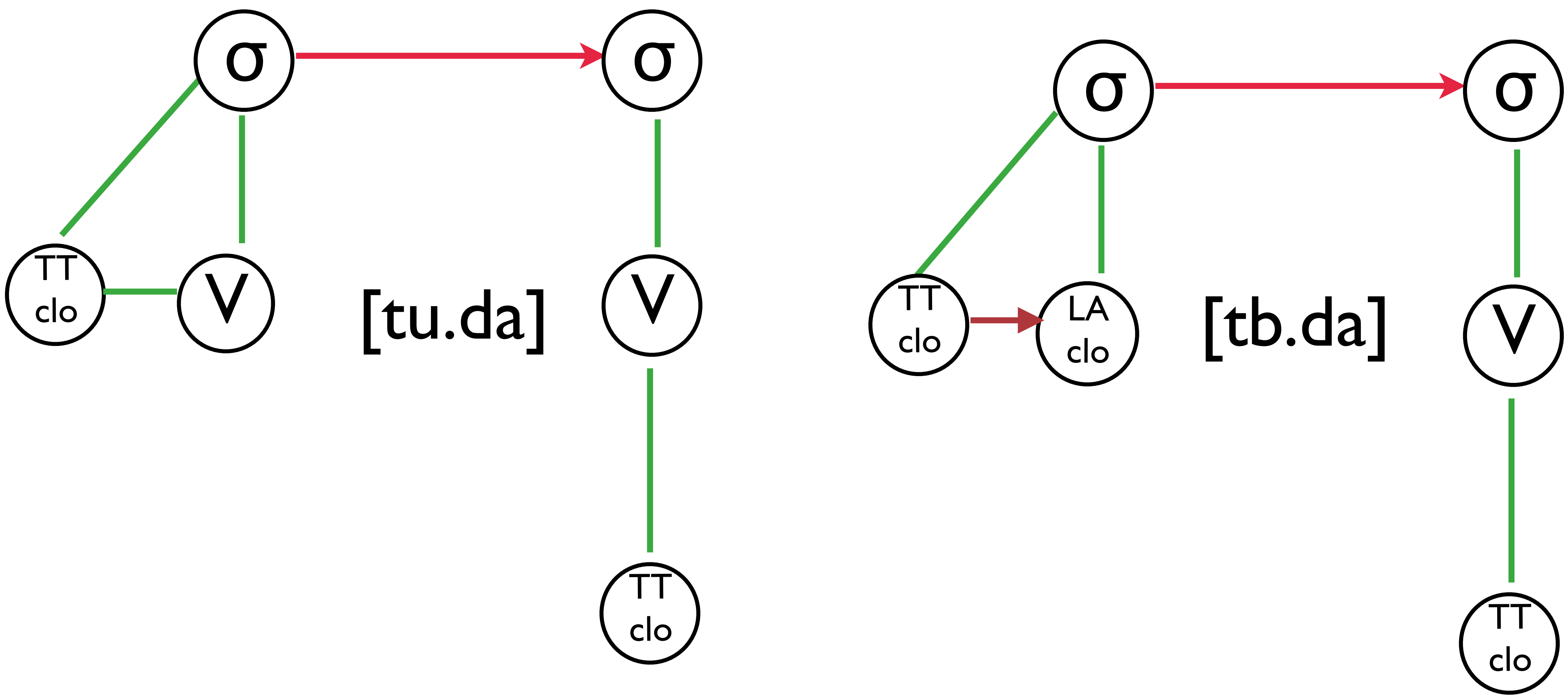


CC vs. CV syllables



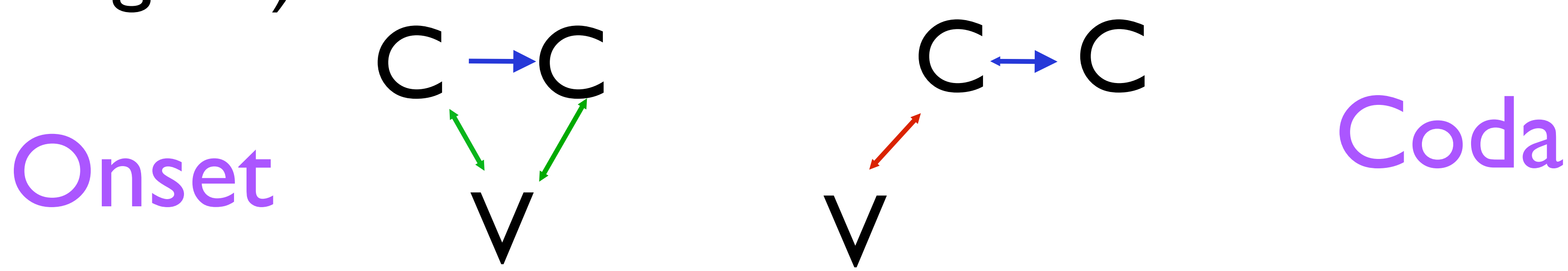
CC syllables

- C gestures are not in-phase
- If they were, they might not be able contrast in order
- Alternative graphs: Note abstract syllable oscillator



CC in onset vs. coda: possible coupling graph differences and weight

- Hypothesis: No competitive coupling in coda (for English)



Why?

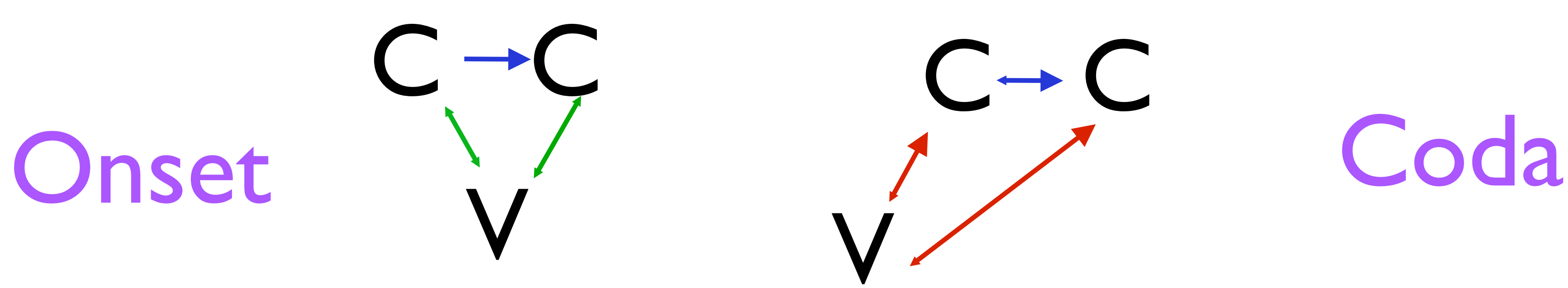
- Perhaps the weaker anti-phase coupling doesn’t attract the (more remote) as strongly as does the in-phase coupling of onsets.

Onset-Coda asymmetry in weight

(Nam, 2007)

- Onset Cs typically do not contribute to syllable weight.
- Coda Cs may or may not depending on the language

- If weight is related to (syllable) duration, then proposed coupling structures can account for the difference between onset and coda consonants in weight.
- With synchronous onset coupling, effect of rightward shift is that adding a C to onset does not increase syllable duration by the duration of the C (more like one-half the C duration).

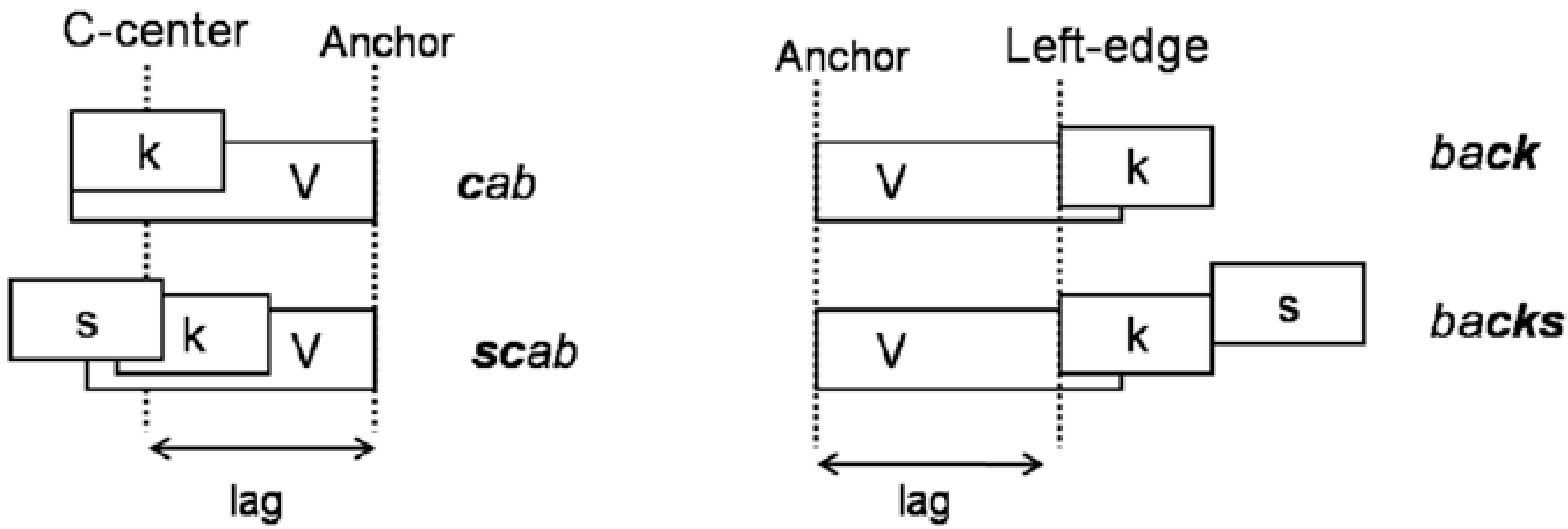


Languages in which coda Cs do not bear weight are predicted to show competitive coda coupling (e.g., Malayalam, Broselow et al. (1997)

C-Center in coda?

(Marin & Pouplier, 2010)

- Predictions of coupling asymmetry model:



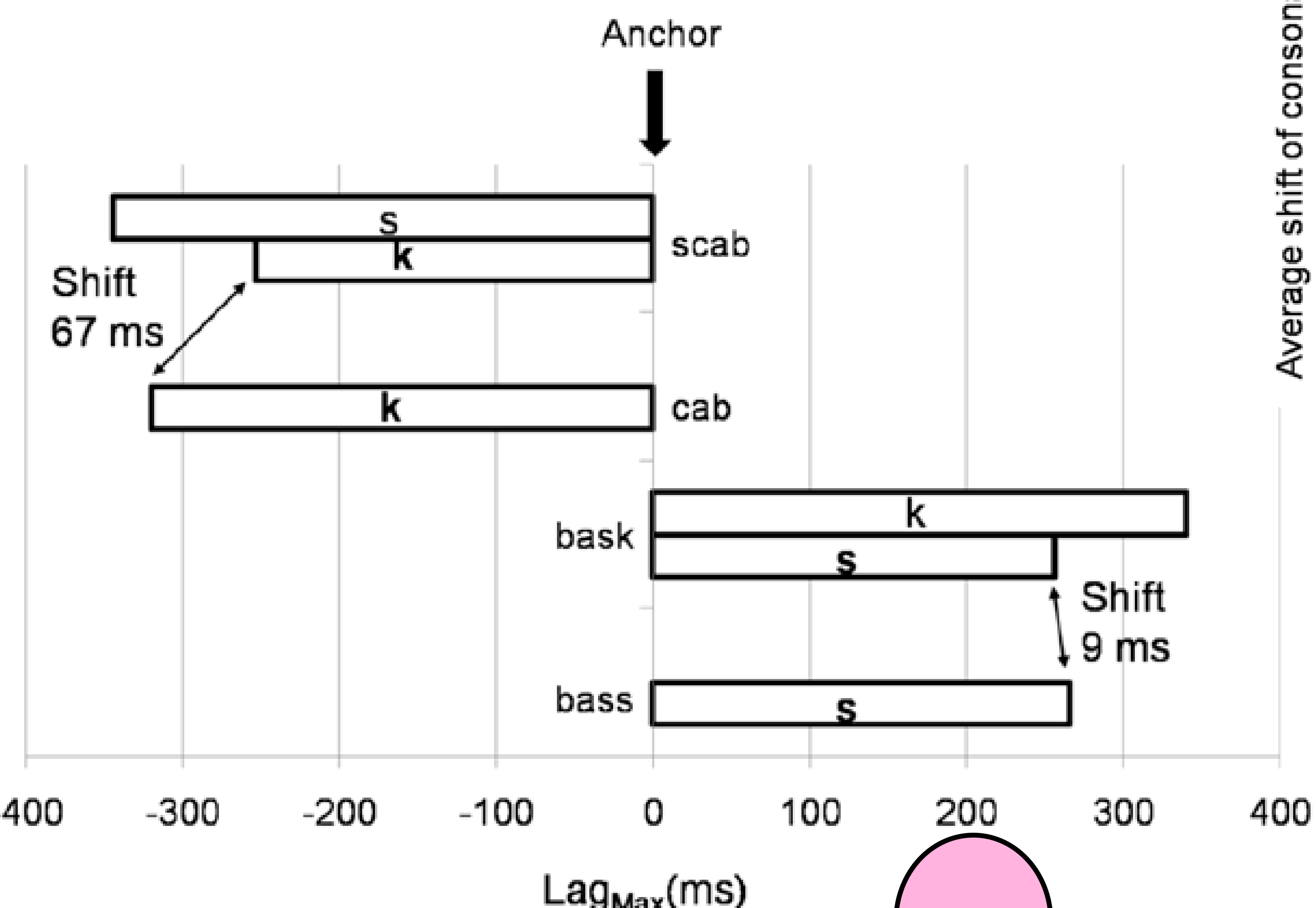
Onsets:

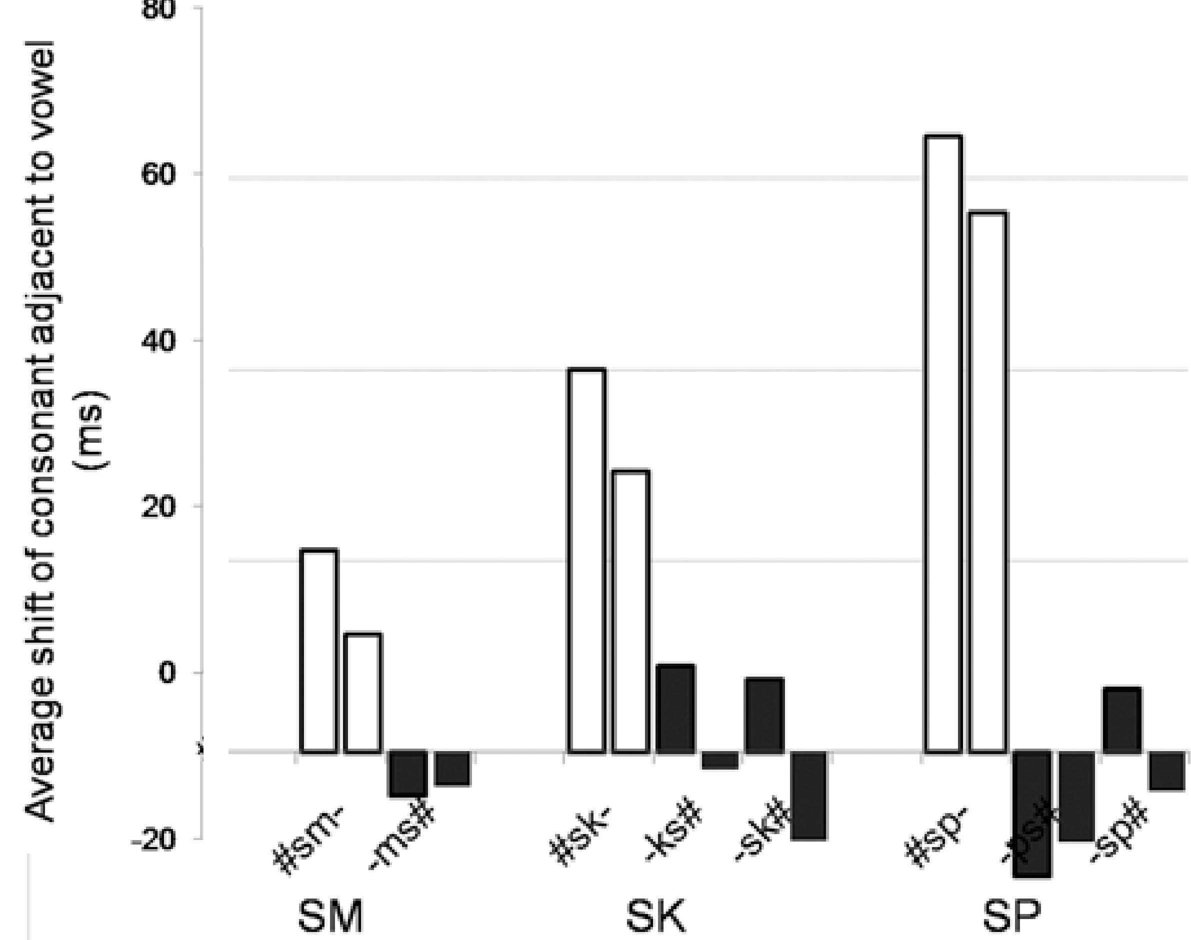
- c-center stability
- rightward shift

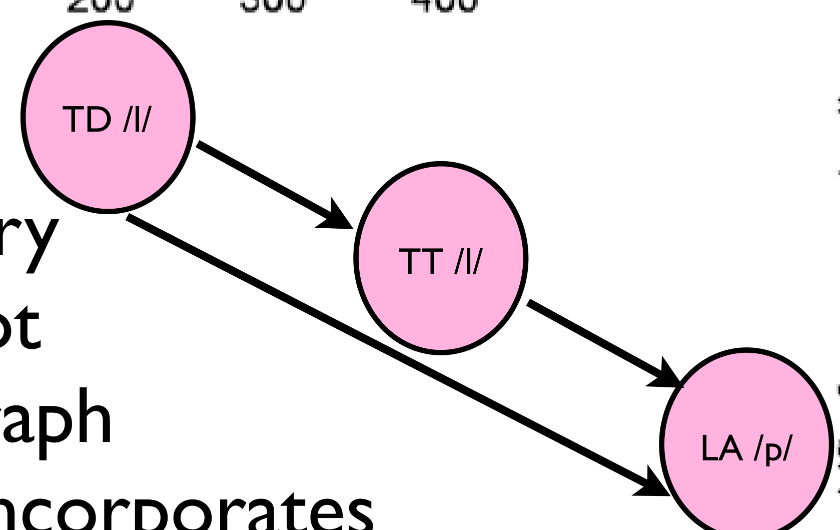
Codas:

- left edge stability
- no leftward shift

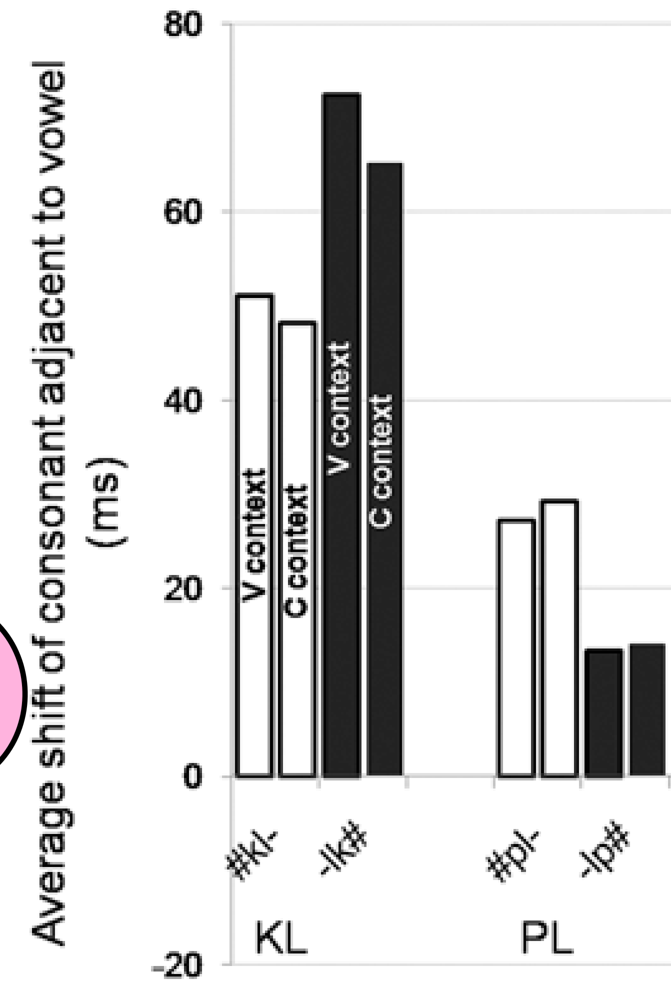
Results: shift







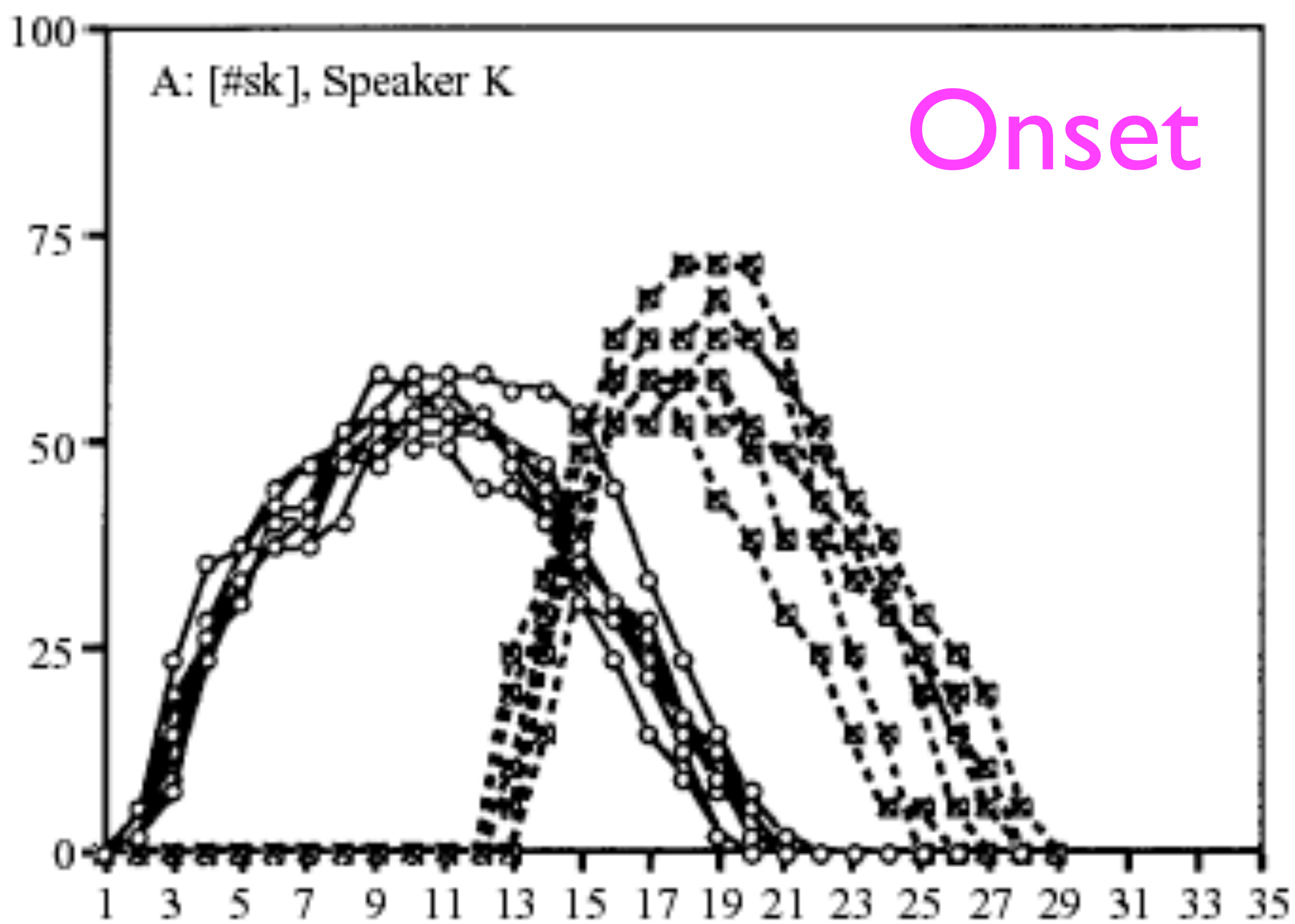
- /s-stop/ support asymmetry
- clusters with liquids do not
  (can be explained by a graph topology that explicitly incorporates multiple gestures for /l/)/

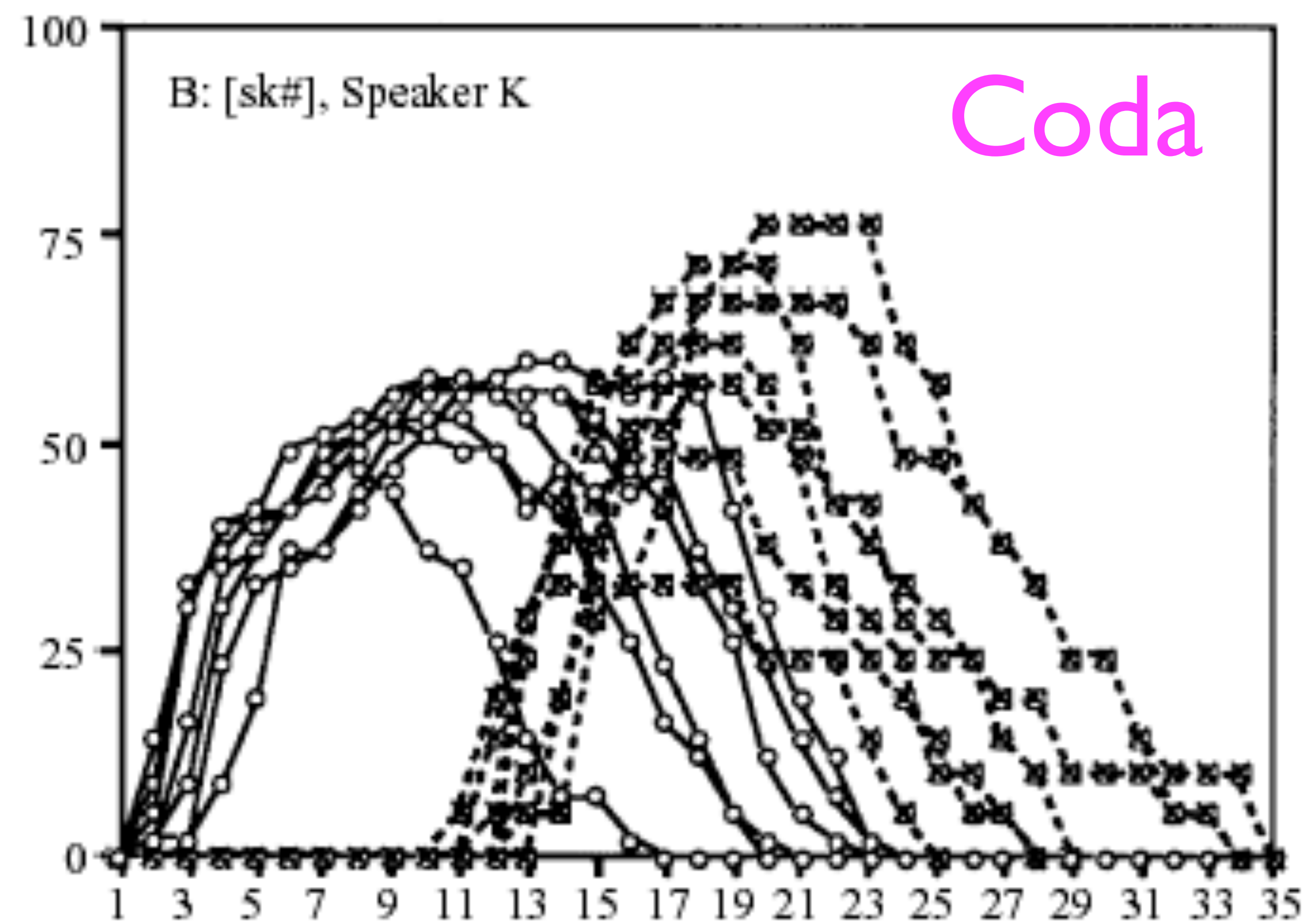


Timing stability in onsets vs. codas

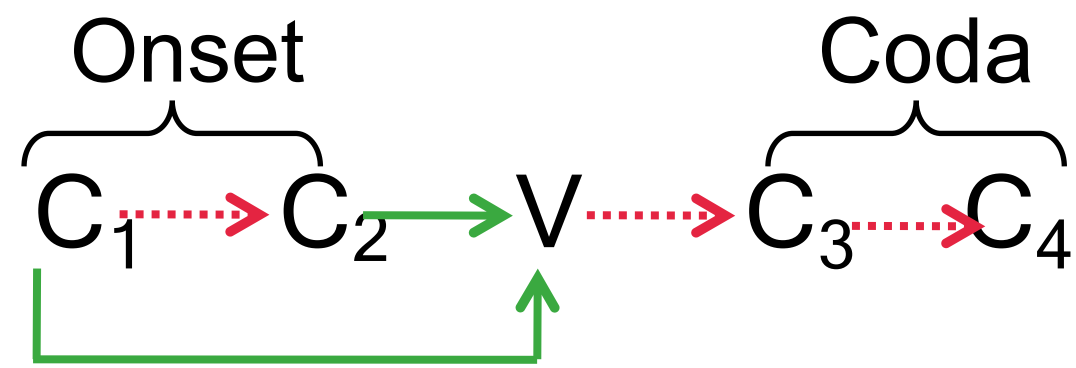
Timing between C gestures is more stable in onset clusters than in coda clusters (Byrd, 1996).

/sk/





Simulation
(Nam, 2007)



- Hypothesis: Loop topology of onsets adds stability (multiple paths) compared to chain.
- Add noise to simulations
  - Noise source: ξ_i (t) = Gaussian, zero mean, unit variance
  - st.dev. of noise ("strength"), β, varied across conditions
- Result: *Greater* steady-state relative phase stability (*lower standard deviation*, σ_{ss}) for clusters in onsets than codas