Grammar and Representation Learning for Opaque Harmony Processes

Society for Computation in Linguistics
February 8, 2022

Caitlin Smith
University of California Davis

Joint work with Charlie O'Hara
University of Michigan

UC DAVIS

UNIVERSITY OF MICHIGAN

Nzebi Height Harmony

(Guthrie 1968, Clements 1991, Parkinson 1996, Kirchner 1996, Smith 2020a)

- Partial height harmony: nonhigh undergoer vowels approach height of high trigger vowel, but do not necessarily reach it
- Nzebi (Bantu; Gabon) *yotization* (Guthrie 1968): in several verb tenses, roots followed by [i] and each nonhigh root vowel raised
- High-mid /e/ and /o/ raise to [i] and [u] in yotized roots

[betə]	[bit-i]	‘carry’
[bexə]	[bit-i]	‘foretell’
[βoːmə]	[βuːm-i]	‘breathe’
[kolənə]	[kulin-i]	‘go down’

Nzebi Height Harmony

(Guthrie 1968, Clements 1991, Parkinson 1996, Kirchner 1996, Smith 2020a)

- Low-mid /ɛ/ and /ɔ/ raise to [e] and [o] in yotized roots

[sɛbə]	[seb-i]	‘laugh’
[suɛmə]	[suem-i]	‘hide self’
[mɔnə]	[mon-i]	‘see’
[tɔːdə]	[toːd-i]	‘arrive’

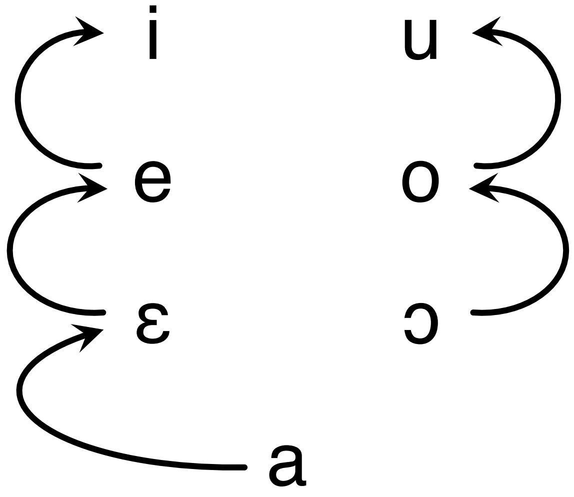
- Low /a/ raises to [e] in yotized roots

[salə]	[sɛl-i]	‘work’
[laxə]	[lɛx-i]	‘show’

Nzebi Height Harmony

(Guthrie 1968, Clements 1991, Parkinson 1996, Kirchner 1996, Smith 2020a)

Nzebi raising harmony: In presence of trigger [i], each nonhigh root vowel raises one ‘step’ along a height scale



Simple Root	Yotized Root	Gloss
[betə]	[bit-i]	‘carry’
[βoːmə]	[βuːm-i]	'breathe’
[sɛbə]	[seb-i]	‘laugh’
[mɔnə]	[mon-i]	‘see’
[salə]	[sɛl-i]	‘work’

Chain Shifts as Derivational Opacity

- Underapplication opacity (McCarthy 1999; Baković 2007, 2011): phonological process appears not to have applied despite its structural description being met in a surface form

- Chain shifts are a type of underapplication opacity:

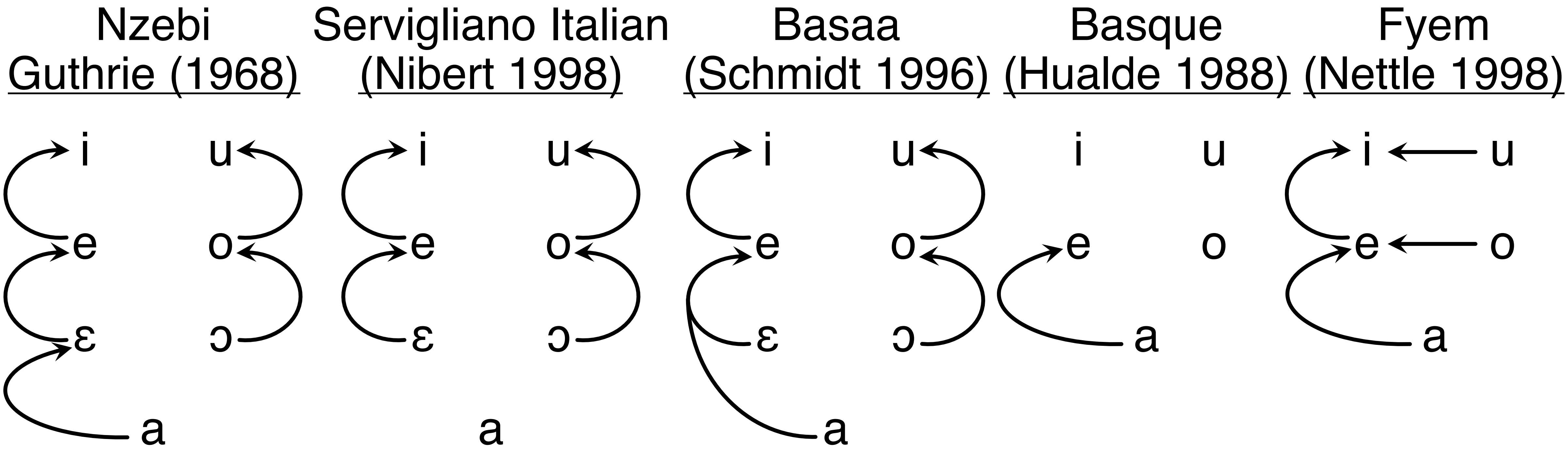
/X/ → [Y]   /Y/ → [Z]

- Challenging for parallel-evaluating, output-driven Optimality Theory (OT; Prince & Smolensky 1993/2004) and Harmonic Grammar (HG; Legendre et al 1990; Smolensky & Legendre 2006):

If /Y/ → [Z], why not /X/ → Y → [Z]?

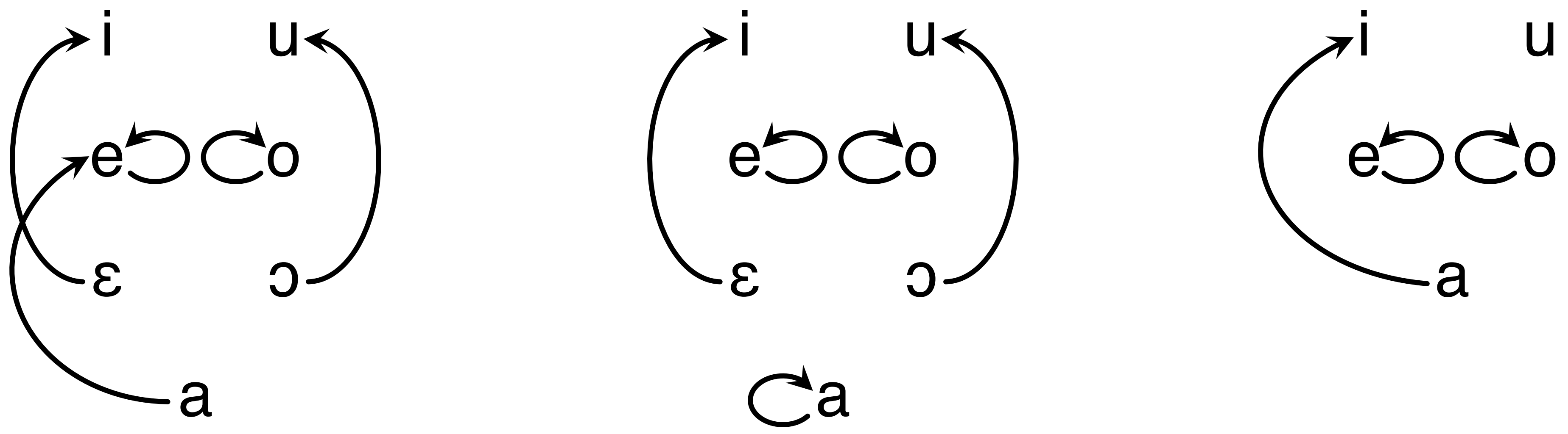
Chain-Shifting Height Harmony

Chain-shifting vowel raising patterns in which vowels raise single step along height scale are well attested:



Unattested Saltatory Height Harmony

Two-step vowel raising patterns that 'skip over' a step in the height scale (i.e., saltation) are unattested (Parkinson 1996):



Saltation as Derivational Opacity

- Saltations are another type of underapplication opacity:

$$/X/ \rightarrow Y \rightarrow [Z] \qquad\qquad /Y/ \rightarrow [Y]$$

- Challenging for Optimality Theory and Harmonic Grammar:

$$\text{If } /X/ \rightarrow Y \rightarrow [Z]\text{, why not } /Y/ \rightarrow [Z]\text{?}$$

- Saltations are rare among phonological processes and apparently unattested in height harmony

The Big Questions

- Chain shifts and saltations cannot be generated in OT or HG using the faithfulness constraints of Correspondence Theory (McCarthy & Prince 1995), e.g. IDENT(F)-IO

Can we formulate a phonological theory that generates derivationally opaque patterns?

- Chain-shifting and saltatory height harmony are both derivationally opaque, but only chain-shifting harmony is well-attested

Can we formulate a phonological theory that predicts robust attestation of chain-shifting harmony and NOT saltatory harmony?

The Gestural Harmony Model
(Smith 2016, 2017ab, 2018, 2020ab)

- Subsegmental units of phonological representation are target-based gestures of Articulatory Phonology (Browman & Goldstein 1986, 1989, et seq.)

- Vowel harmony is result of extension of trigger gesture to overlap gestures of other segments in a word

- Partial height harmony is result of blending between vowel gestures with different target articulatory states (heights)

Proposals: A Gestural Account of Derivationally Opaque Height Harmony

- Partial height harmony via blending in the Gestural Harmony Model generates attested chain-shifting raising and unattested saltatory raising

- Aspects of *learnability* of saltatory height harmony explain its lack of attestation

Learnability and Phonological Typology

- Patterns predicted by phonological framework are determined by setup of grammar, but also by how easy they are to learn (Hayes & Wilson 2008; Pater & Moreton 2012; White 2013; Staubs 2014; Stanton 2016; Hughto 2020; O'Hara 2021)

- For a pattern to be robustly attested, it must be derivable within a phonological framework, but also easily learnable within that framework

The Grammar + Gesture Gradual Learning Algorithm

- Grammar + Gesture Gradual Learning Algorithm (GGGLA): models simultaneous learning of gestural parameter settings and constraint-based phonological grammar

- Modeled the acquisition of grammar and gestural parameter settings that generate chain-shifting and saltatory height harmony

Gestural Harmony Model and GGGLA correctly predict chain-shifting harmony to be more learnable/better attested

Examining the Alternatives: Featural Accounts of Derivationally Opaque Height Harmony

- Assuming non-standard faithfulness constraint definitions, both chain-shifting and saltatory patterns are derivable in OT and HG

- Modeled the acquisition of phonological grammars that derive derivationally opaque patterns in these frameworks

Featural frameworks that derive both chain-shifting and saltatory height harmonies incorrectly predict saltatory harmonies to be more learnable/better attested

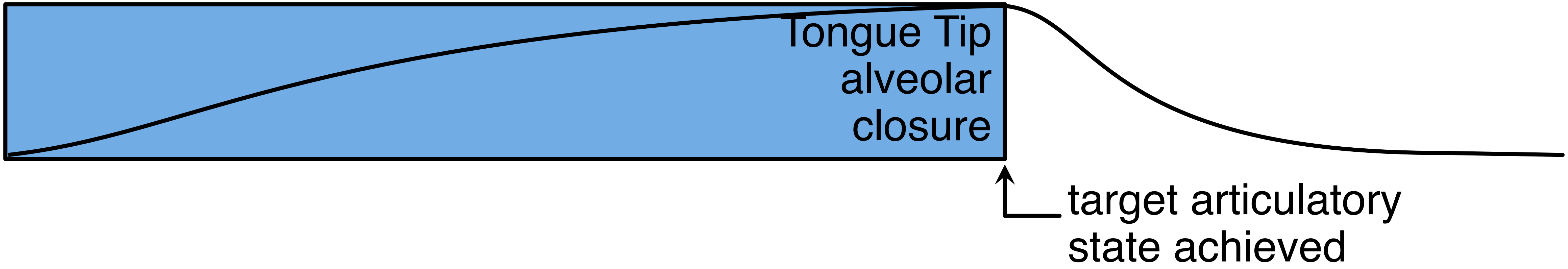
Roadmap

- Gestures as Phonological Units
- Gestural Harmony Model
- Gestural Analysis of Nzebi Chain-Shifting Height Harmony
- Grammar + Gesture Gradual Learning Algorithm
- Generating and Learning Chain-Shifting and Saltatory Height Harmony in Featural Frameworks

Gestures as Phonological Units

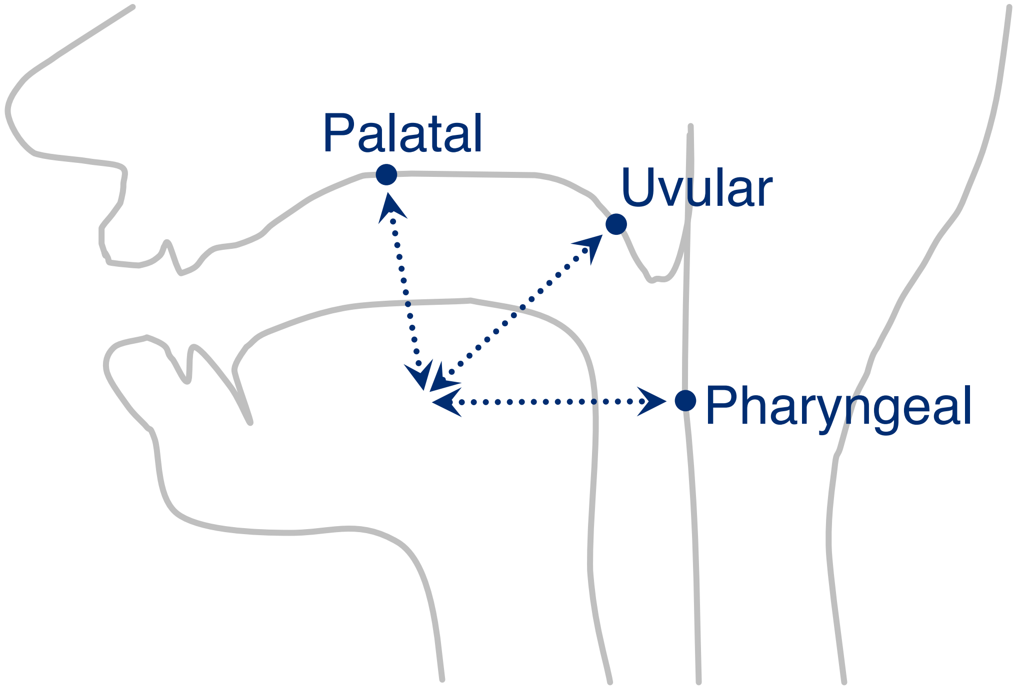
Gestures in Articulatory Phonology
(Browman & Goldstein 1986, 1989 et seq.)

- Gestures: dynamically-defined, goal-based units of phonological representation (Browman & Goldstein 1986, 1989)



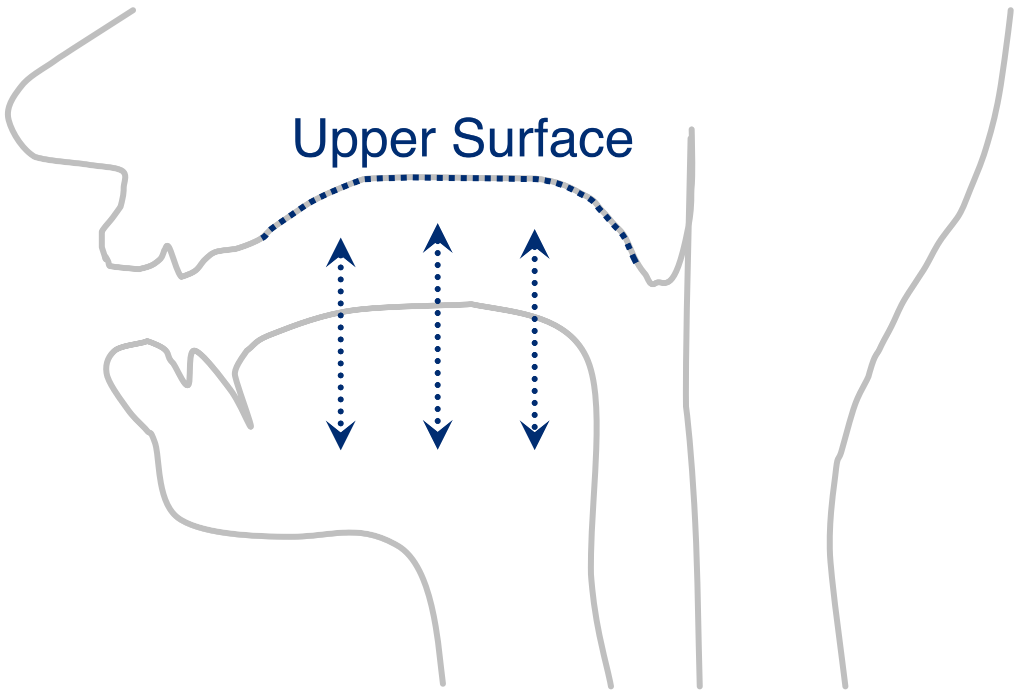
- Target articulatory state:
  - Constriction location
  - Constriction degree
- Blending strength (α): ability to command vocal tract articulators
- Ability to self-activate and self-deactivate (Smith 2016, 2017ab, 2018)

Constriction Location and Degree for Consonantal Gestures



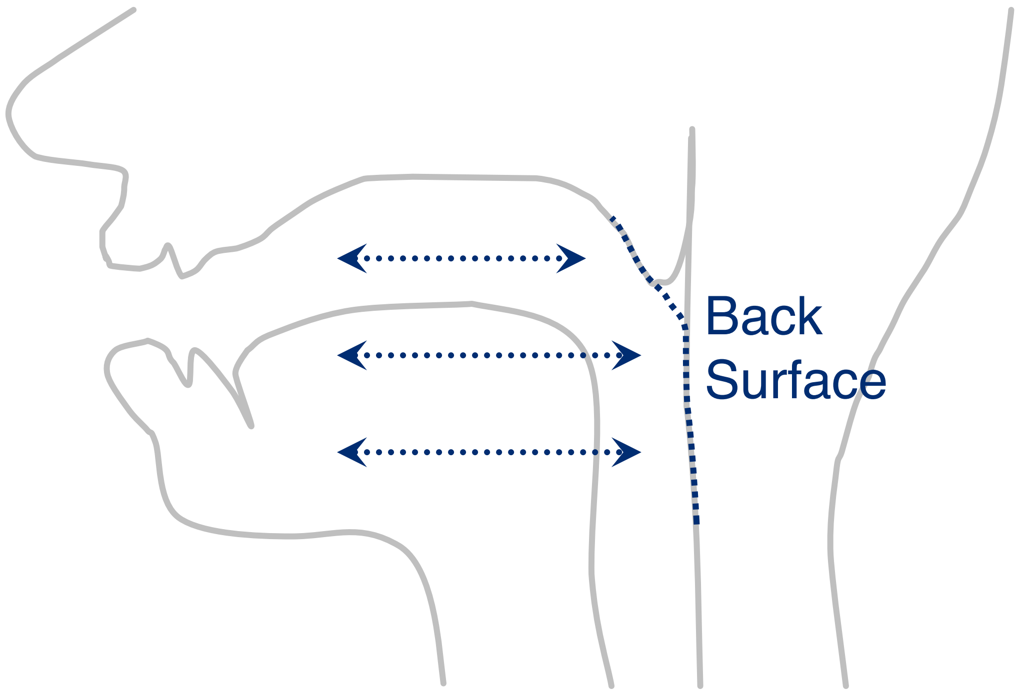
- Constriction location of gesture specifies target point along vocal tract surface
- Constriction degree of gesture specifies distance between active articulator and constriction location point

Constriction Location and Degree for Vowel Gestures
(Smith 2020b)



- Each vowel includes two tongue body gestures:
  - Constriction location 'upper surface'
  - Constriction location 'back surface'
- Constriction degree of upper surface gesture determines vowel height
- Constriction degree of back surface gesture determines vowel backness

Constriction Location and Degree for Vowel Gestures
(Smith 2020b)



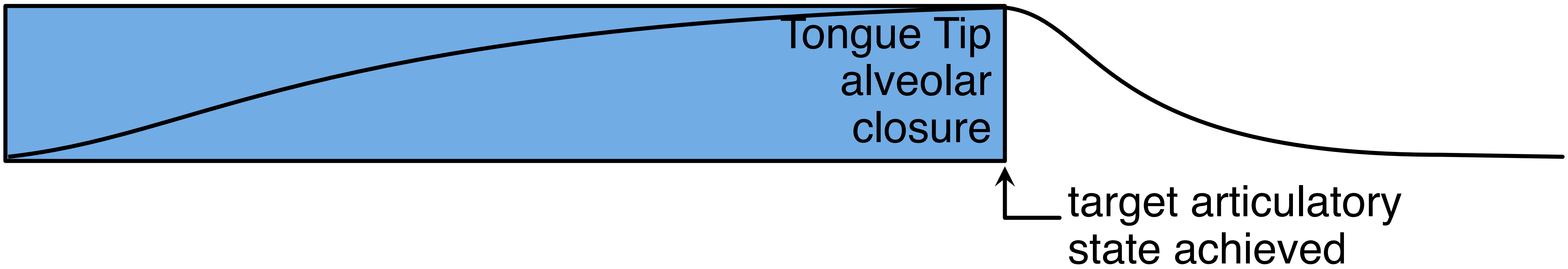
- Each vowel includes two tongue body gestures:
  - Constriction location 'upper surface'
  - Constriction location 'back surface'

- Constriction degree of upper surface gesture determines vowel height

- Constriction degree of back surface gesture determines vowel backness

Gestures in Articulatory Phonology
(Browman & Goldstein 1986, 1989 et seq.)

- Gestures: dynamically-defined, goal-based units of phonological representation in Articulatory Phonology

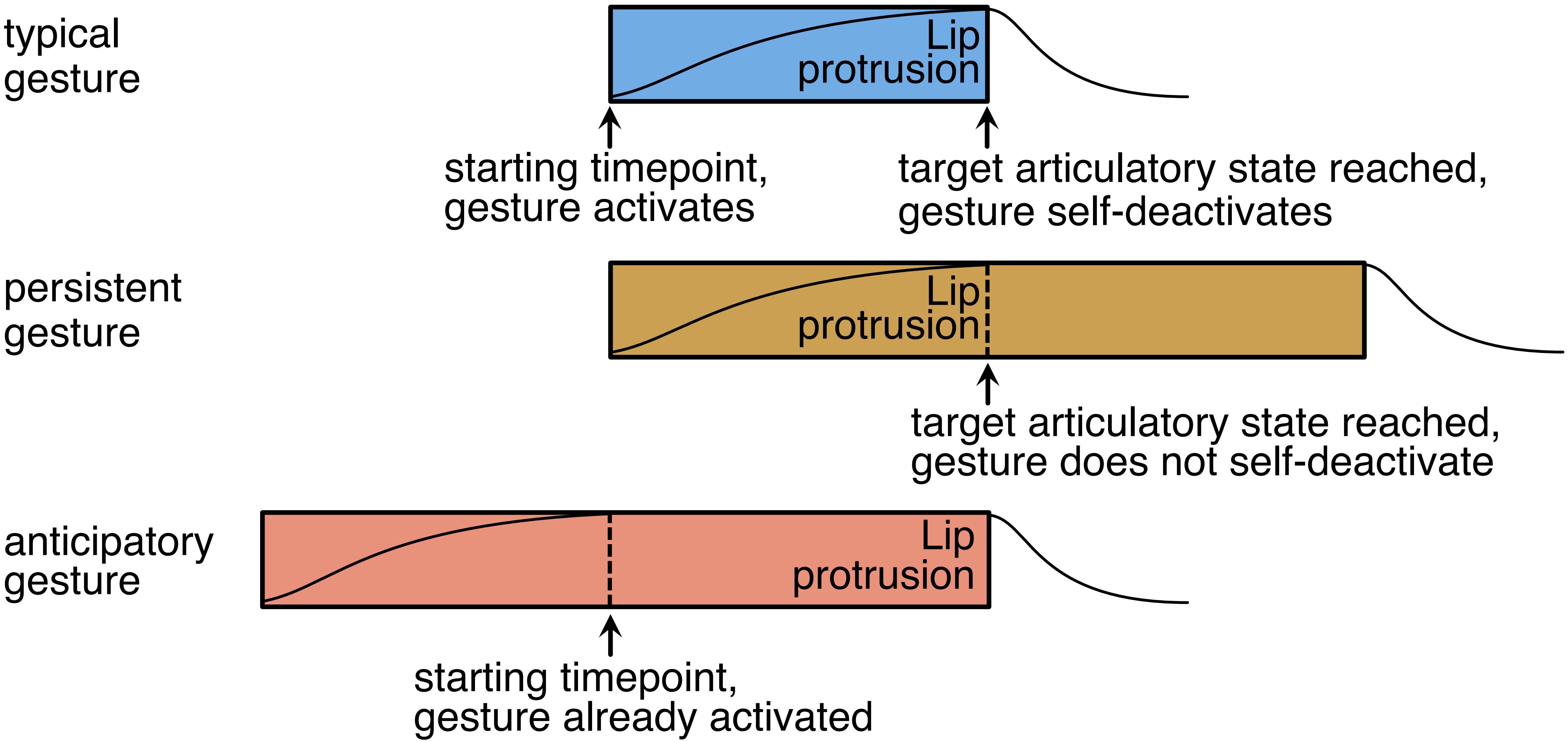


- Target articulatory state:
  - Constriction location
  - Constriction degree
- Blending strength (α): ability to command vocal tract articulators
- Ability to self-activate and self-deactivate (Smith 2016, 2017ab, 2018)

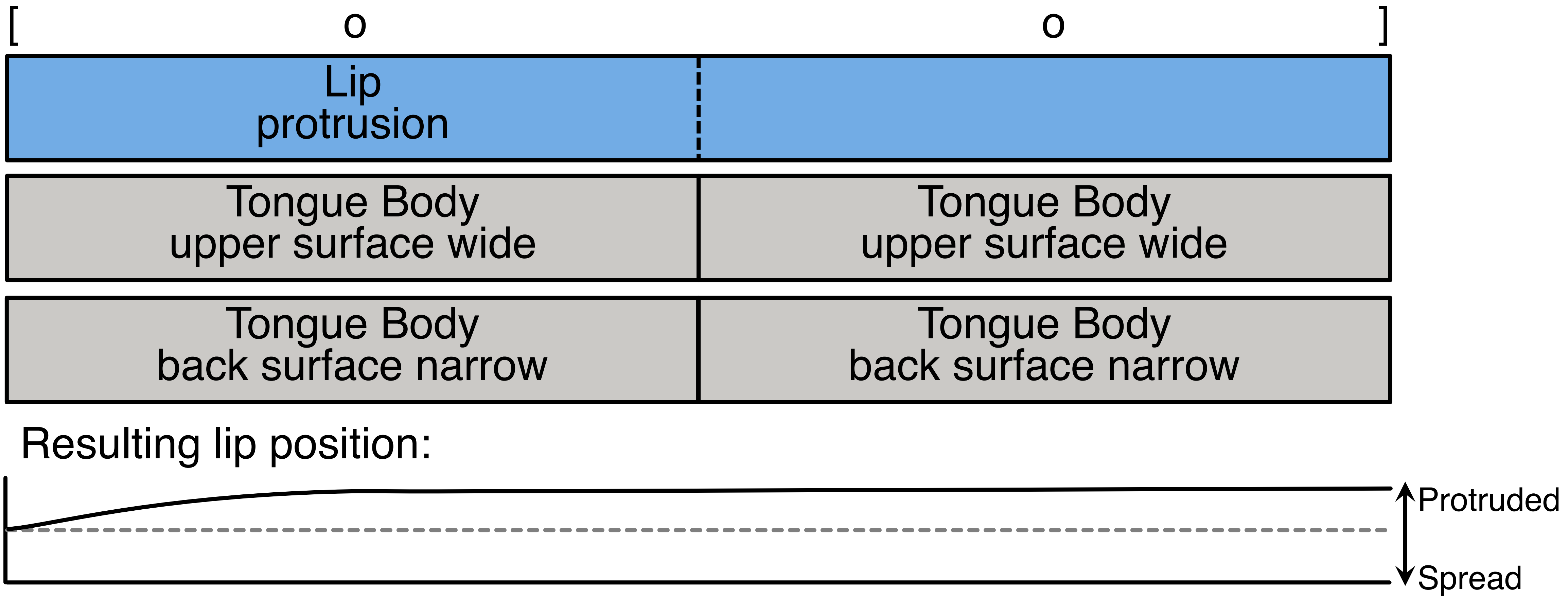
The Gestural Harmony Model

Gestural Activation and Deactivation

(Smith 2016, 2017ab, 2018)

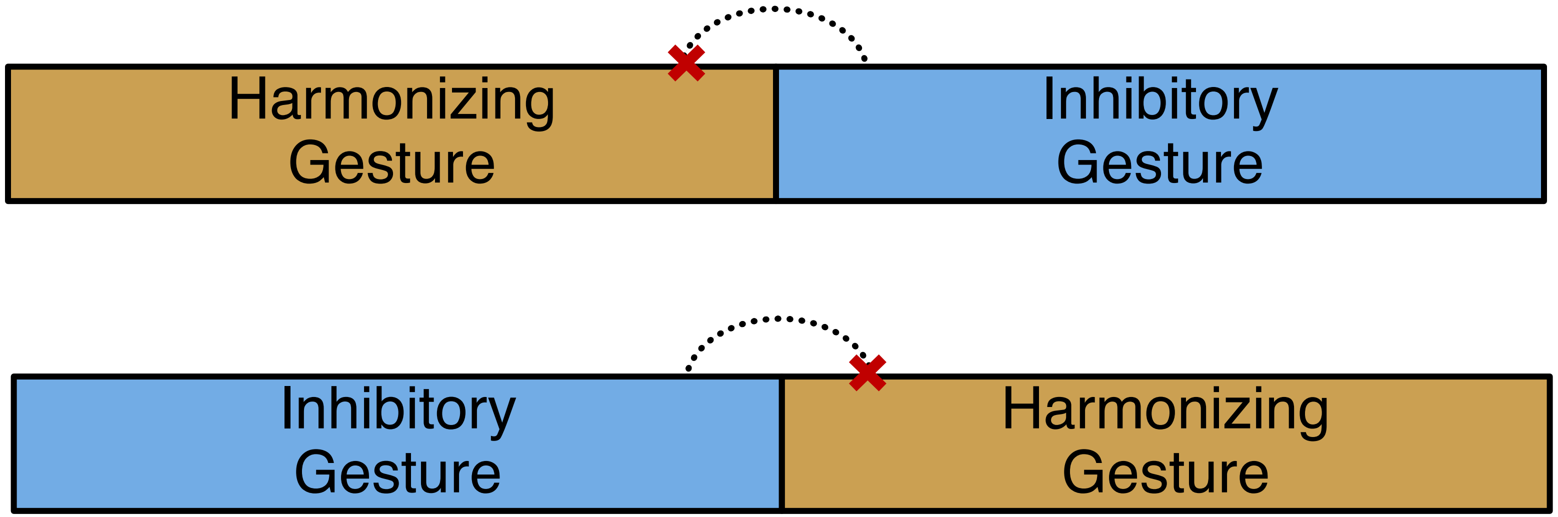


Example: Rounding Harmony



Blocking as Gestural Inhibition

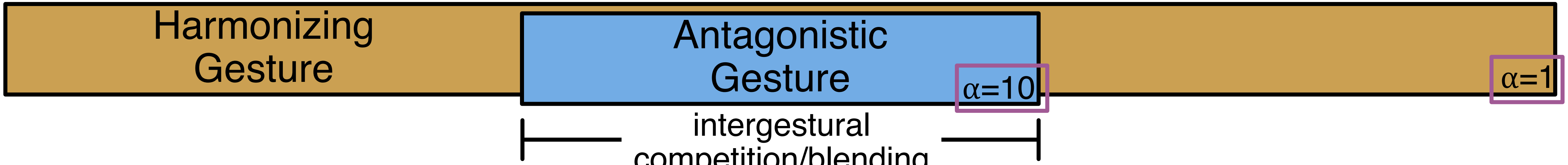
- Blocking: inhibition (deactivation) of harmonizing gesture by incompatible gesture, preventing gestural overlap (Smith 2016, 2018)



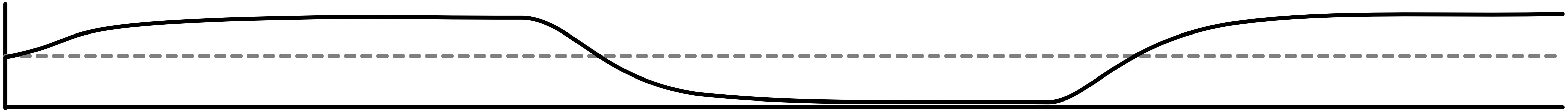
- Inhibited gesture cannot reactivate itself, so harmony ceases

Transparency as Gestural Blending

- Transparency: competition between two concurrently active *antagonistic* gestures (Smith 2016, 2018)
- Gestural antagonism: two concurrently active gestures with opposing target articulatory states



Resulting state of vocal tract for some variable:



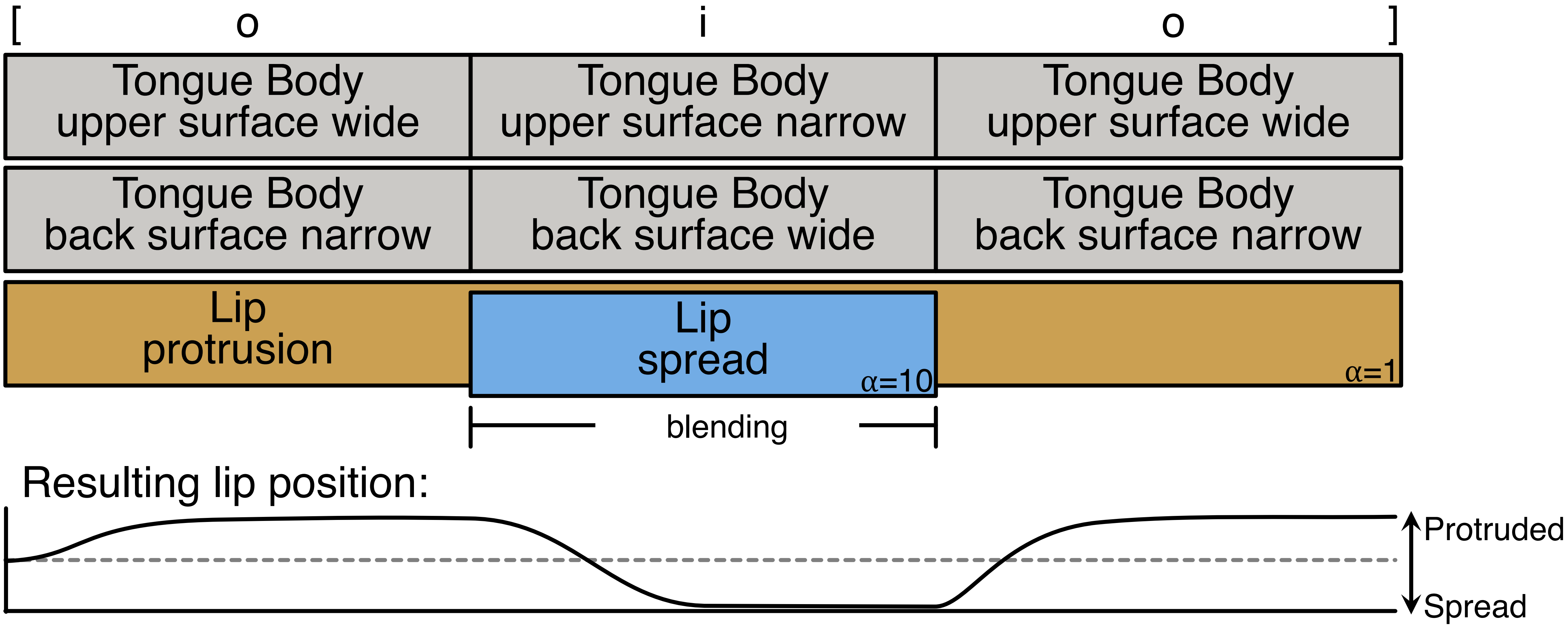
Gestural Strength and Blending

- Antagonistic gestures: gestures with conflicting target articulatory states

- Antagonism resolved by blending goal articulatory states of concurrently active gestures according to Task Dynamic Model of speech production (Saltzman & Munhall 1989, Fowler & Saltzman 1993)

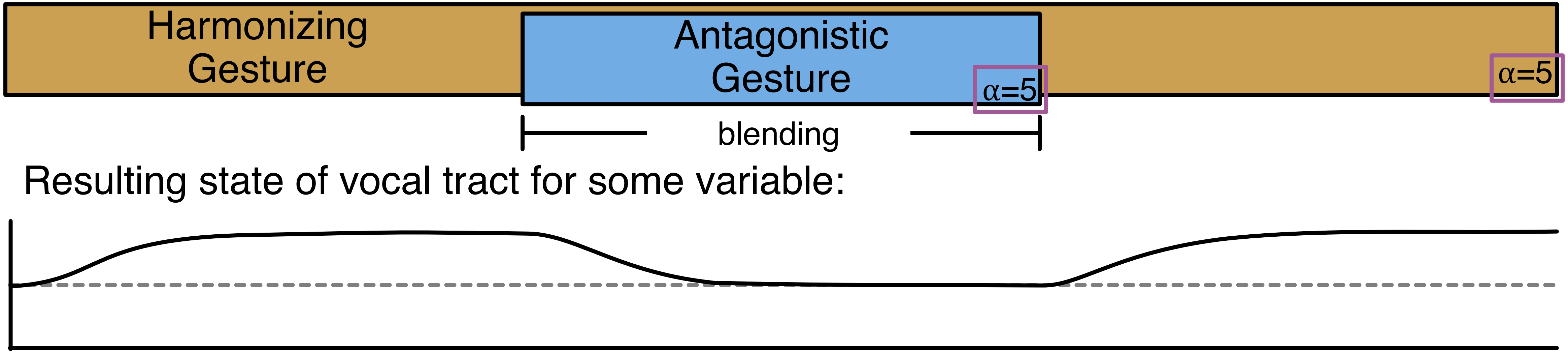
$$\frac{\text{Target}_1 * \alpha_1 + \text{Target}_2 * \alpha_2}{\alpha_1 + \alpha_2} = \text{Blended Target}$$

Example: Transparency in Rounding Harmony



Prediction: Partial Transparency via Gestural Blending

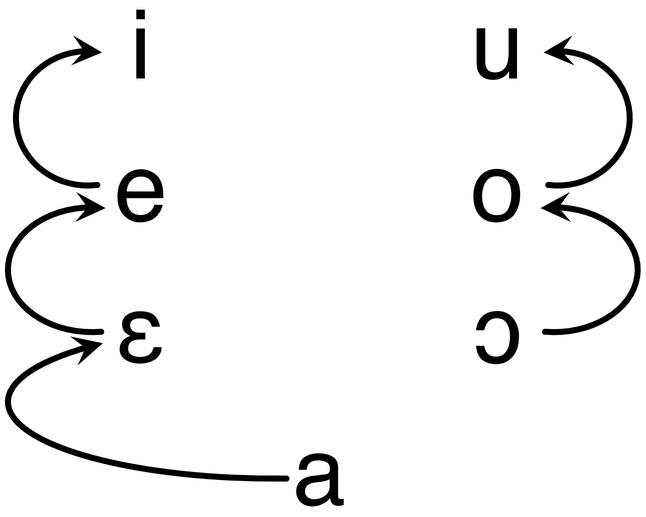
- Full transparency: overlapped gesture of transparent segment is much stronger than harmonizing gesture (e.g. 10-to-1)
- Identical or similar blending strengths of harmonizing gesture and overlapped gesture predicts partial transparency/partial undergoing of harmony



Chain-Shifting Height Harmony in Nzebi

Nzebi Chain-Shifting Height Harmony

(Guthrie 1968, Clements 1991, Parkinson 1996, Kirchner 1996; Smith 2020a)

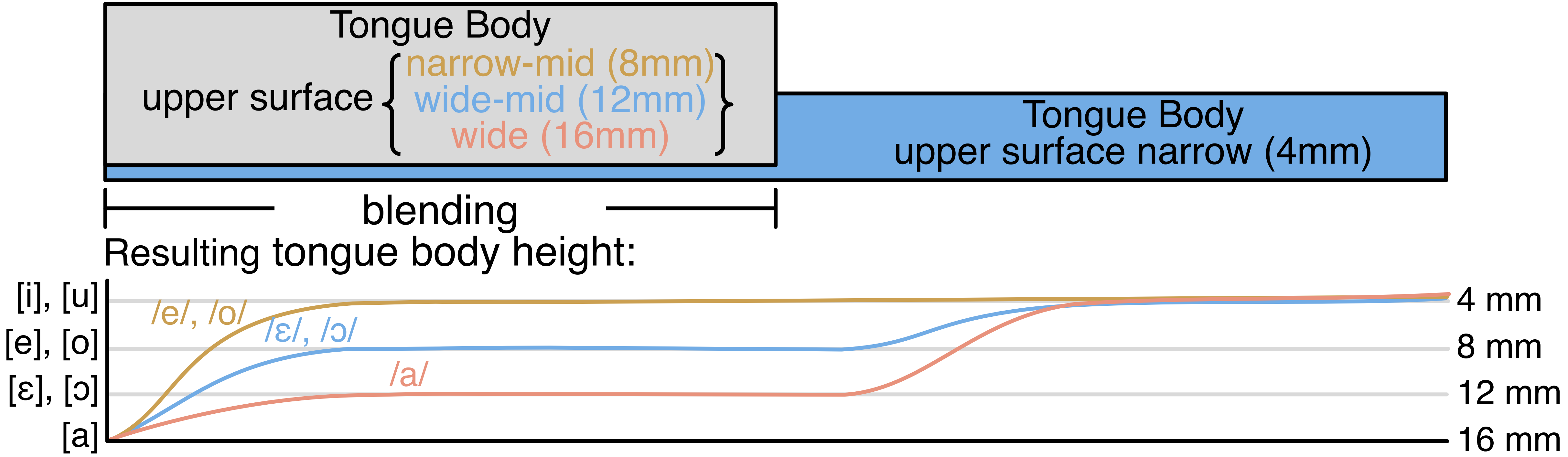


Simple Root	Yotized Root	Gloss
[betə]	[bit-i]	‘carry’
[βoːmə]	[βuːm-i]	'breathe’
[sɛbə]	[seb-i]	‘laugh’
[mɔnə]	[mon-i]	‘see’
[salə]	[sɛl-i]	‘work’

A Gestural Analysis of Nzebi Height Harmony

(Smith 2020a)

- Vowel raising harmony due to overlap by anticipatory upper surface narrowing gesture of suffix high vowel /i/
- Vowels of different heights have antagonistic target states for upper surface constriction degree, resulting in gestural blending



Gestural Overpowering
(Smith & O'Hara 2021)

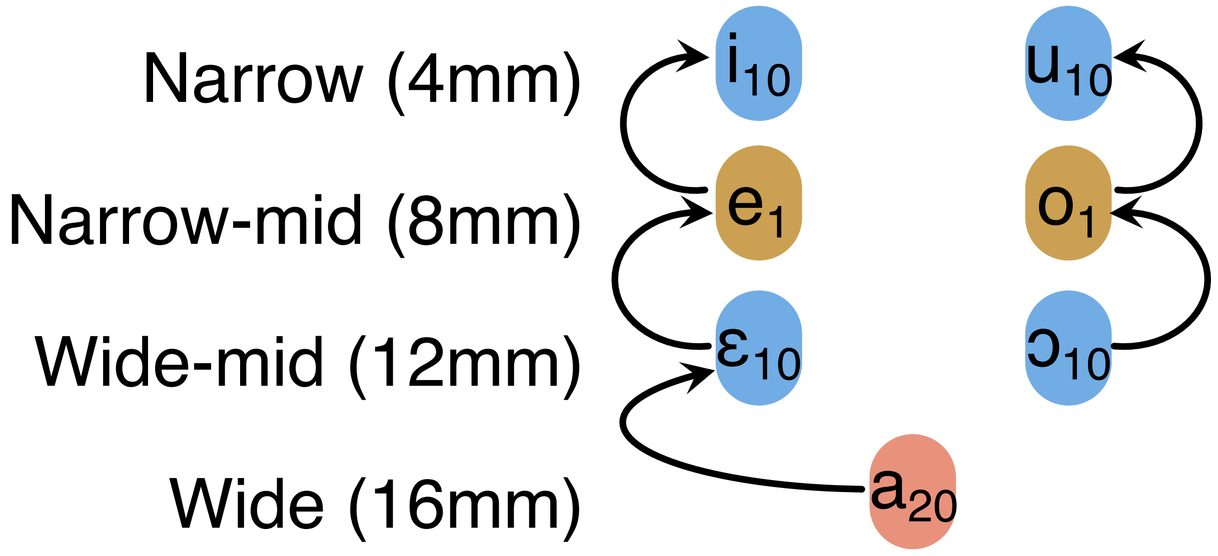
Triggering full assimilation and resisting full assimilation depend on *overpowering* relationships between blended gestures:

- For assimilation of X to Y, Y's gestural strength must be order of magnitude higher than that of X

- For Z to resist assimilation to Y, Z's gestural strength must be order of magnitude higher than that of Y

$$Z_{100} \mapsto Y_{10} \mapsto X_1$$

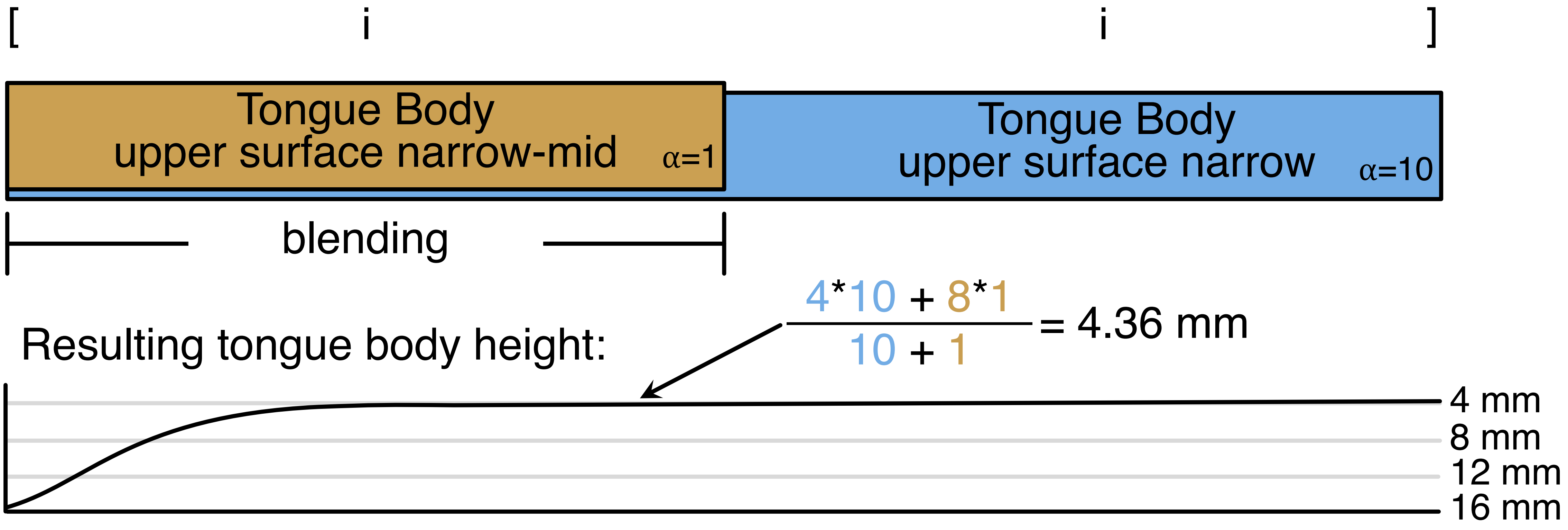
Nzebi Gestural Parameters



- Weak narrow-mid vowels /e/ and /o/ do not resist raising and surface as narrow
- Wide-mid vowels /ɛ/ and /ɔ/ surface as narrow-mid, partially resisting raising to narrow due to strength equal to trigger gesture
- Strong vowel /a/ surfaces as wide-mid, mostly resisting raising due to strength greater than trigger gesture

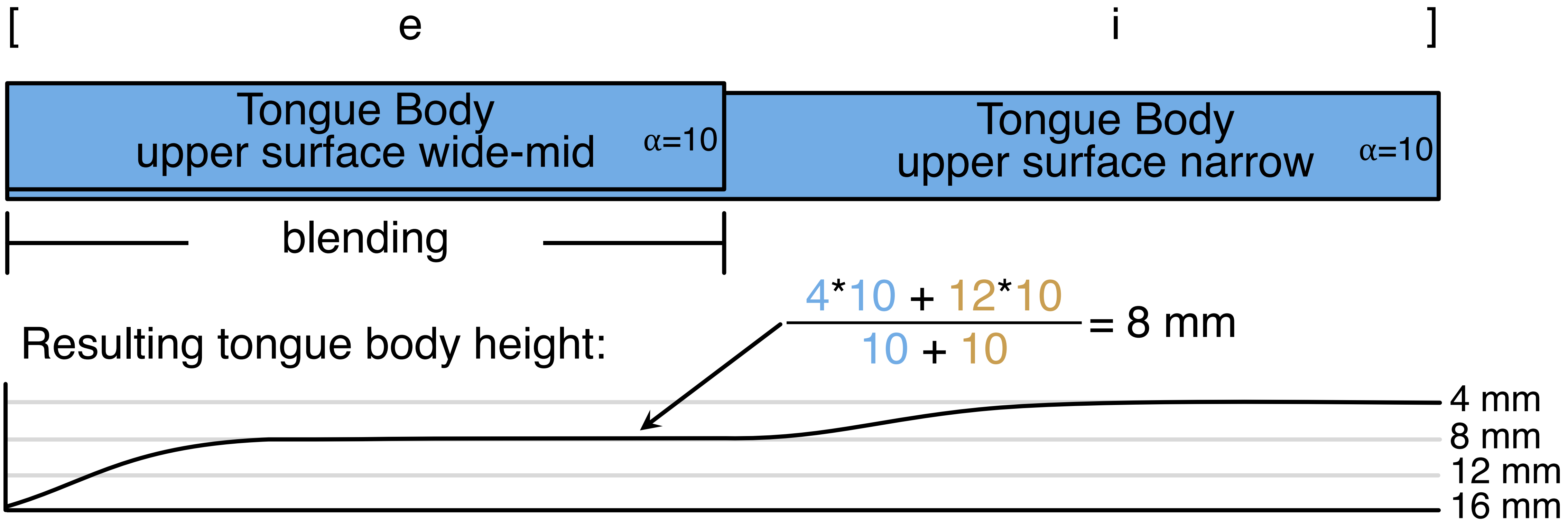
Nzebi Analysis: Narrow-Mid to High Raising

- Narrow-mid vowels /e/ and /o/ fully undergo harmony
- Relative gestural blending strengths favor target constriction degree (narrow upper surface constriction) of high vowels



Nzebi Analysis: Wide-Mid to Narrow-Mid Raising

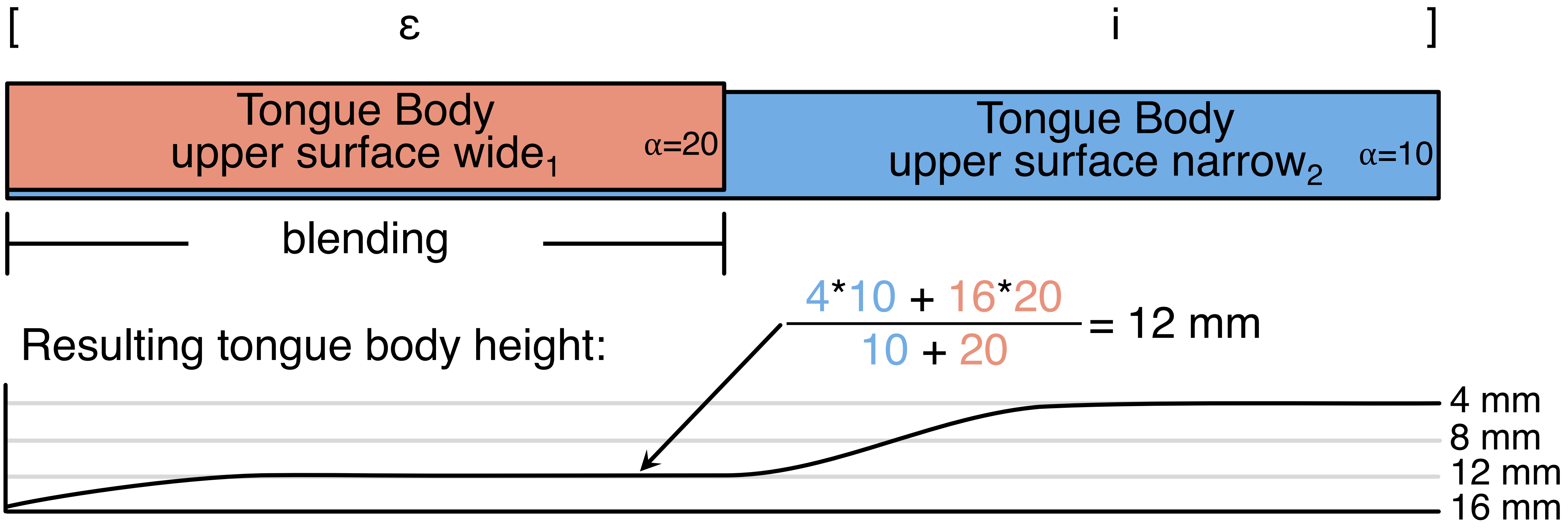
- Overlap between gestures of wide-mid vowels /ɛ/ and /ɔ/ and narrow /i/ produces narrow-mid [e] and [o]
- Intermediate blended articulatory state due to equal gestural strengths



$$\frac{4*10 + 12*10}{10 + 10} = 8 \text{ mm}$$

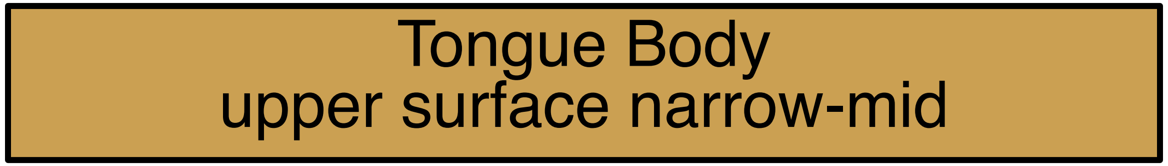
Nzebi Analysis: Wide to Wide-Mid Raising

- Overlap between gestures of wide vowel /a/ and narrow /i/ produces wide-mid vowel [ɛ]
- Blending strengths slightly favor target constriction degree of wide vowel

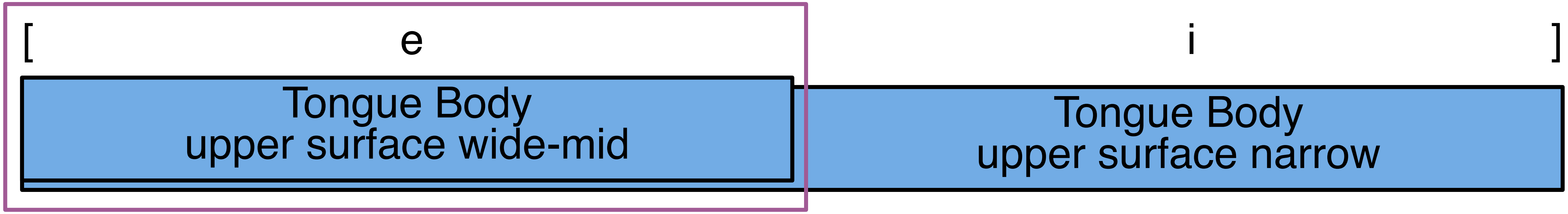


Modeling a Chain Shifting: Underlying and Derived Vowels

- Underlying mid-high vowel /e/:



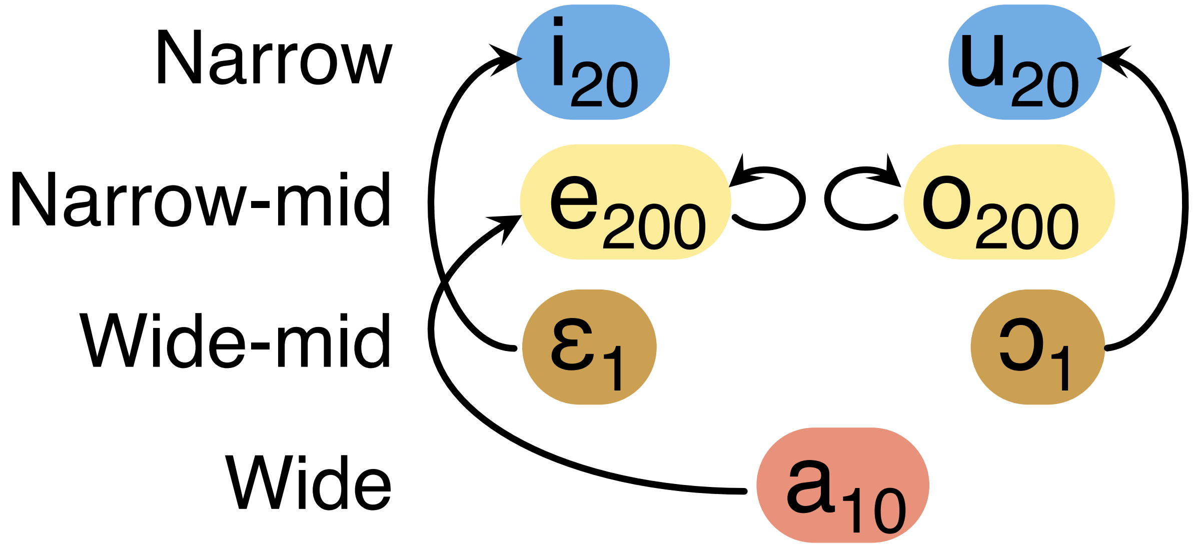
- Mid-high vowel [e] derived by blending /ɛ/ and /i/:



Deriving Saltatory Height Harmony

Gestural Blending for Saltatory Harmony?

With extreme enough strength values, saltatory height harmony can also be generated by the Gestural Harmony Model



Saltatory Height Harmony: Why Such Extreme Strengths?

- Chain-shifting height harmony requires only one overpowering relation between vowels: high vowels overpower high-mid vowels to trigger full assimilation

/i/, /u/ ↦ /e/, /o/

- Saltatory height harmony requires two overpowering relations between vowels:
  - High-mid vowels overpower high vowels to fully resist raising
  - High vowels overpower low-mid vowels to trigger full assimilation

/e/, /o/ ↦ /i/, /u/ ↦ /ɛ/, /ɔ/

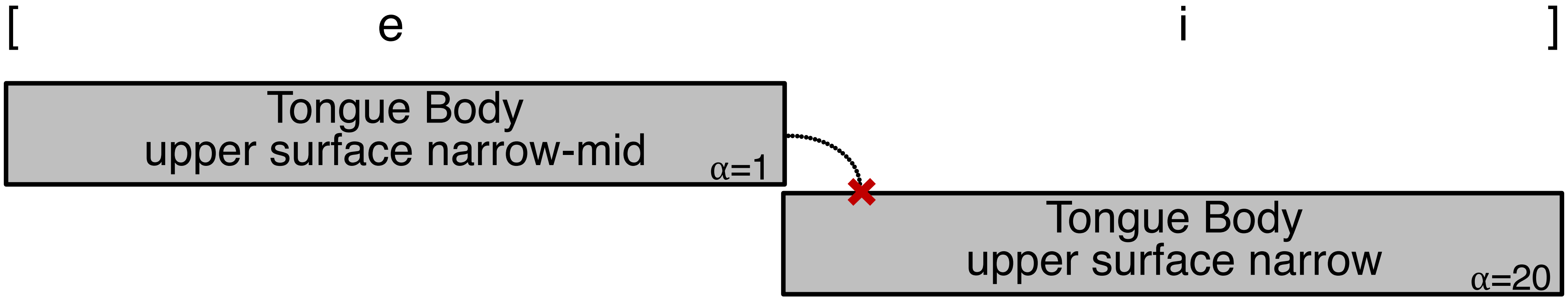
Extreme Strengths Affect Learning

- Smith & O’Hara (2021) simulated learners of chain-shifting and saltatory height harmony via gestural blending, given harmony-producing grammar already known to learner

- Results: extreme strengths necessary to generate saltatory height harmony are substantially slower/harder to learn

An Alternative Representation of Saltatory Height Harmony

- Also possible to generate saltation via blocking of harmony by vowels at non-raising height:



- With saltation via blocking, blocking vowels need not take on extreme strengths in order to prevent raising triggered by high vowel
- True picture of learnability in Gestural Harmony Model must account for overlapping and blocking structures and phonological grammars that derive them

Simulating Learning with the Grammar + Gesture Gradual Learning Algorithm

Learnability Affects Phonological Typology

- Learnability gradiently shapes typological frequency of phonological patterns (Hayes & Wilson 2008; Pater & Moreton 2012; White 2013; Staubs 2014; Stanton 2016; Hughto 2020; O'Hara 2021)
- A pattern that is more difficult to learn is more likely to change across generations, becoming typologically underrepresented

- Saltatory harmony requires much more data to be correctly learned than chain-shifting harmony
- Saltatory harmony is more likely to change across generations and disappear

Grammar + Gesture Gradual Learning Algorithm (GGGLA)

- Learns phonological grammar necessary to correctly cast vowels as triggers, undergoers, and blockers of harmony
- Learns gestural parameters necessary to correctly produce target partial height harmony pattern (chain shift or saltation) when trigger/undergoer overlap occurs
- During training, compares articulatory productions of gestural scores from teacher and learner agents

Constraint Weight Learning in Maximum Entropy Harmonic Grammar

- Constraint-based grammars implemented in Maximum Entropy Harmonic Grammar (MaxEnt; Goldwater & Johnson 2003; Jäger 2007)

- Learning of constraint weights based on Gradual Learning Algorithm for Harmonic Grammar (GLA; Boersma & Pater 2016), based on earlier Perceptron Algorithm (Rosenblatt 1958)

The Constraint Set

(Smith 2018)

- PERSIST(height)
- *GEST(TB high, self-deactivating)
- SELFDEACTIVATE
- *INHIBIT
- INHIBIT(TB upper surface, TB upper surface)
- INHIBIT(TB high, TB mid)
- INHIBIT(TB high, TB low)
- INHIBIT(TB mid, TB low)

PERSIST(height), *GEST(TB high, self-deactivating): drive harmony

SELFDEACTIVATE: penalizes harmony

*INHIBIT: penalizes blocking

INHIBIT constraints: require blocking between various vowel heights

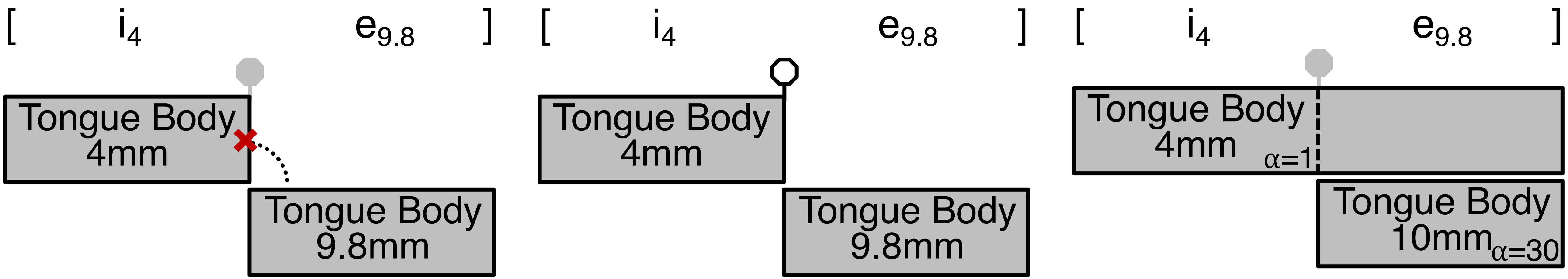
Error-Driven Online Learning

- During a learning trial, randomly generated two-syllable input is evaluated by teacher’s and learner’s individual phonological grammars and winning output candidate gestural score is produced according to their individual gestural parameter settings
- Check learner and teacher productions for a match (i.e. both learner vowels produced within specified window around both teacher vowels)
- In event of error, perform necessary updates to learner’s constraint weights and/or gestural parameter settings (constriction degree, blending strength)

Articulatory Productions and Hidden Structure

- To perform constraint weight updates, learner must compare violations incurred by its and teacher's chosen winner candidates

- Teacher's output candidate gestural score is often unknown to learner based solely on teacher's production:



Hidden Structure Learning

- Robust Interpretive Parsing (RIP; Tesar & Smolensky 1998, Boersma 2003): determine teacher candidate's hidden structure (e.g. metrical structure) from its surface form

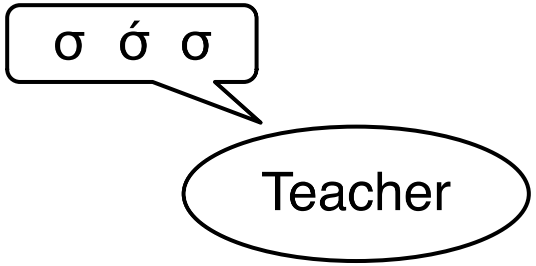
- Robust Interpretive Production Parsing (RIPP): determining teacher candidate's hidden phonological surface form (i.e. gestural score) from its articulatory production

Robust Interpretive Production Parsing (RIPP)

- Standard RIP: sample among learner candidates with surface forms matching teacher's and assume sampled hidden structure as teacher's output parse

Learner
a. (σ́ σ) σ
b. (σ σ́) σ
c. σ (σ́ σ)

sampled to choose assumed teacher parse (b., c.)

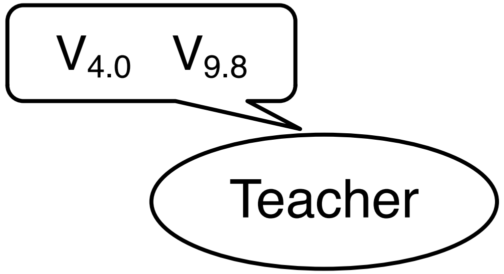


- The problem: Gestural Harmony Model produces numerically-valued productions of each vowel, with no guarantee of any such match in productions between learner and teacher

Robust Interpretive Production Parsing (RIPP)

Learner	P
a. 4.0 12.5	0.25
b. 4.0 15.5	<0.01
c. 4.0 15.5	0.01
d. 4.0 15.5	0.74

← choose closest match to teacher production

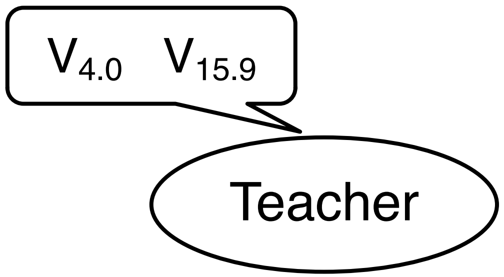


- First, assume teacher's hidden gestural score matches that of whichever learner candidate production most closely matches teacher's production
- If single closest match exists, choose its gestural score as teacher's parse

Robust Interpretive Production Parsing (RIPP)

	Learner	P
a.	4.0 12.5	0.25
b.	4.0 15.5	<0.01
c.	4.0 15.5	0.01
d.	4.0 15.5	0.74

(b.–d.) equally close to teacher production, sampled to choose teacher parse

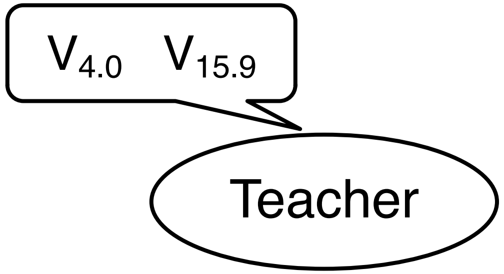


- If two or more learner candidates have equally close productions to teacher's production, sample from among them according to probabilities generated by learner's current MaxEnt grammar
- Choose sampled candidate's gestural score as teacher's parse

Robust Interpretive Production Parsing (RIPP)

Learner	P
a. 4.0 12.5	0.25
b. 4.0 15.5	<0.01
c. 4.0 15.5	0.01
d. 4.0 15.5	0.74

(b.–d.:) equally close to teacher production, sampled to choose teacher parse

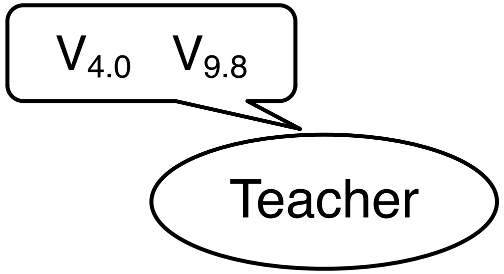


Once teacher parse is chosen, if learner's candidate gestural score and teacher's parsed gestural score do not match, perform constraint weight updates according to Gradual Learning Algorithm

Robust Interpretive Production Parsing (RIPP)

Learner	P
a. 4.0 12.5	0.25
b. 4.0 15.5	<0.01
c. 4.0 15.5	0.01
d. 4.0 15.5	0.74

← teacher parse



- After constraint weight updates, learner compares its production of teacher’s parsed gestural score to teacher’s production
- If learner’s production of teacher’s parsed gestural score does not match teacher’s production, perform gestural parameter updates

Gestural Parameter Updates

- Learning of gestural parameter settings based on Gestural Gradual Learning Algorithm (GGLA; Smith & O'Hara 2021)
- For each incorrect vowel production, perform parameter updates to any vowel gestures involved in that production
- Constriction degree update for blended and unblended vowel productions:

$$\text{sgn(teacher prod.} - \text{learner prod.)} \times \text{LR}$$

- Blending strength updates for blended vowel gestures V_1 and V_2:

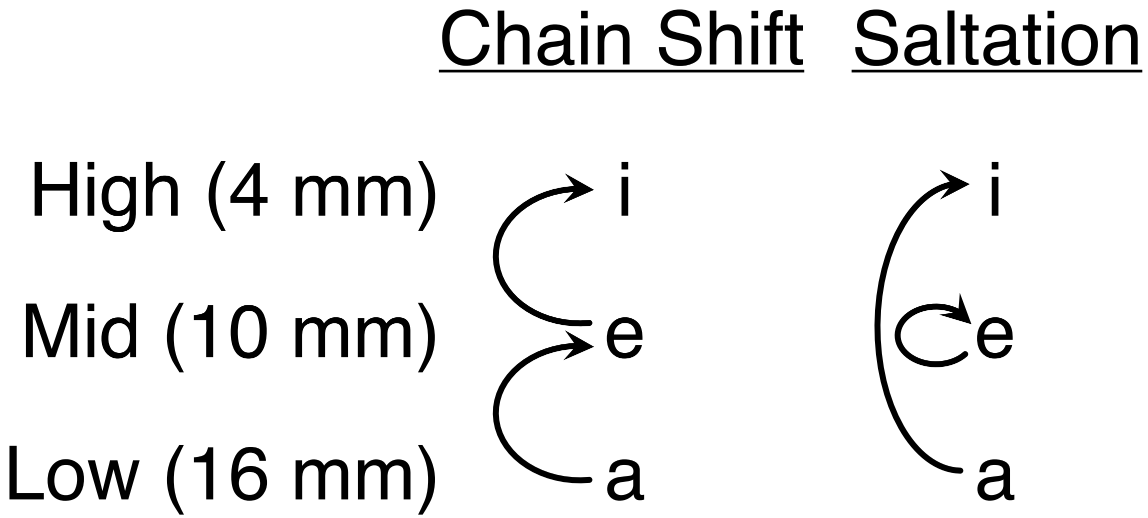
V_1: sgn(learner V_1 CD – learner V_2 CD) × sgn(teacher prod. – learner prod.) × LR
V_2: sgn(learner V_2 CD – learner V_1 CD) × sgn(teacher prod. – learner prod.) × LR

CD = constriction degree, LR = learning rate

Our Models

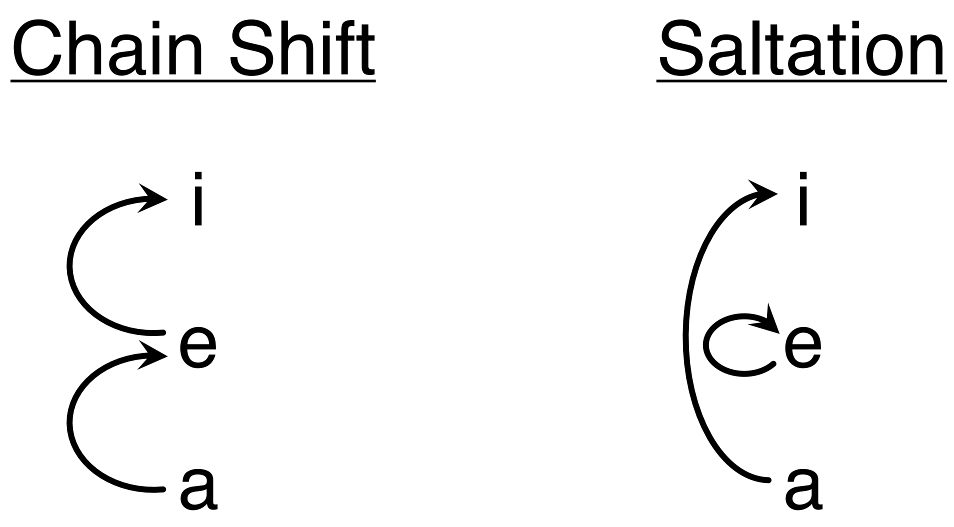
Two progressive (rightward) harmony patterns tested:

- Three-height inventory, chain-shifting raising after high vowel trigger (Nzebi-like)
- Three-height inventory, saltatory (two-step) raising after high vowel trigger (unattested)



Our Models

- Ran 100 models of each type until convergence
- Convergence: for each input, assigning cumulative ≥ 99% probability to all candidates whose productions match teacher's production



Results: Learned Gestural Parameters

	Chain Shift		Saltation	
	Constriction Degree	Blending Strength	Constriction Degree	Blending Strength
/i/	3.90mm	17.63	3.90mm	28.40
/e/	9.82mm	1.04	10.17mm	10.59
/a/	16.19mm	16.41	15.81mm	1.01

Learned gestural parameter settings for chain-shifting learners are consistent with overpowering of /e/ by /i/, and intermediate blending between /i/ and /a/

Results: Learned Gestural Parameters

	Chain Shift		Saltation	
	Constriction Degree	Blending Strength	Constriction Degree	Blending Strength
/i/	3.90mm	17.63	3.90mm	28.40
/e/	9.82mm	1.04	10.17mm	10.59
/a/	16.19mm	16.41	15.81mm	1.01

- Learned gestural parameter settings for saltatory learners are consistent with over-powering of /a/ by /i/
- Grammar ensures no overlap between /e/ and /i/, so /e/ does not need strength high enough to overpower it

Results: Learned Grammars

- Learners of chain-shifting height harmony converged on grammars with high vowel triggering and no blocking of high vowel overlap:

w(SELFDEACTIVATE) > w(PERSIST)

w(PERSIST(ht.)) + w(*GEST(TB high, self-deact.)) > w(SELFDEACTIVATE)

w(*INHIBIT) > w(INHIBIT(TB upper, TB upper) + {w(INHIBIT(TB high, TBmid)) / w(INHIBIT(TB high, TB low))}

- Learners of saltatory height harmony converged on grammars with high vowel triggering and blocking of high vowel overlap by mid vowels:

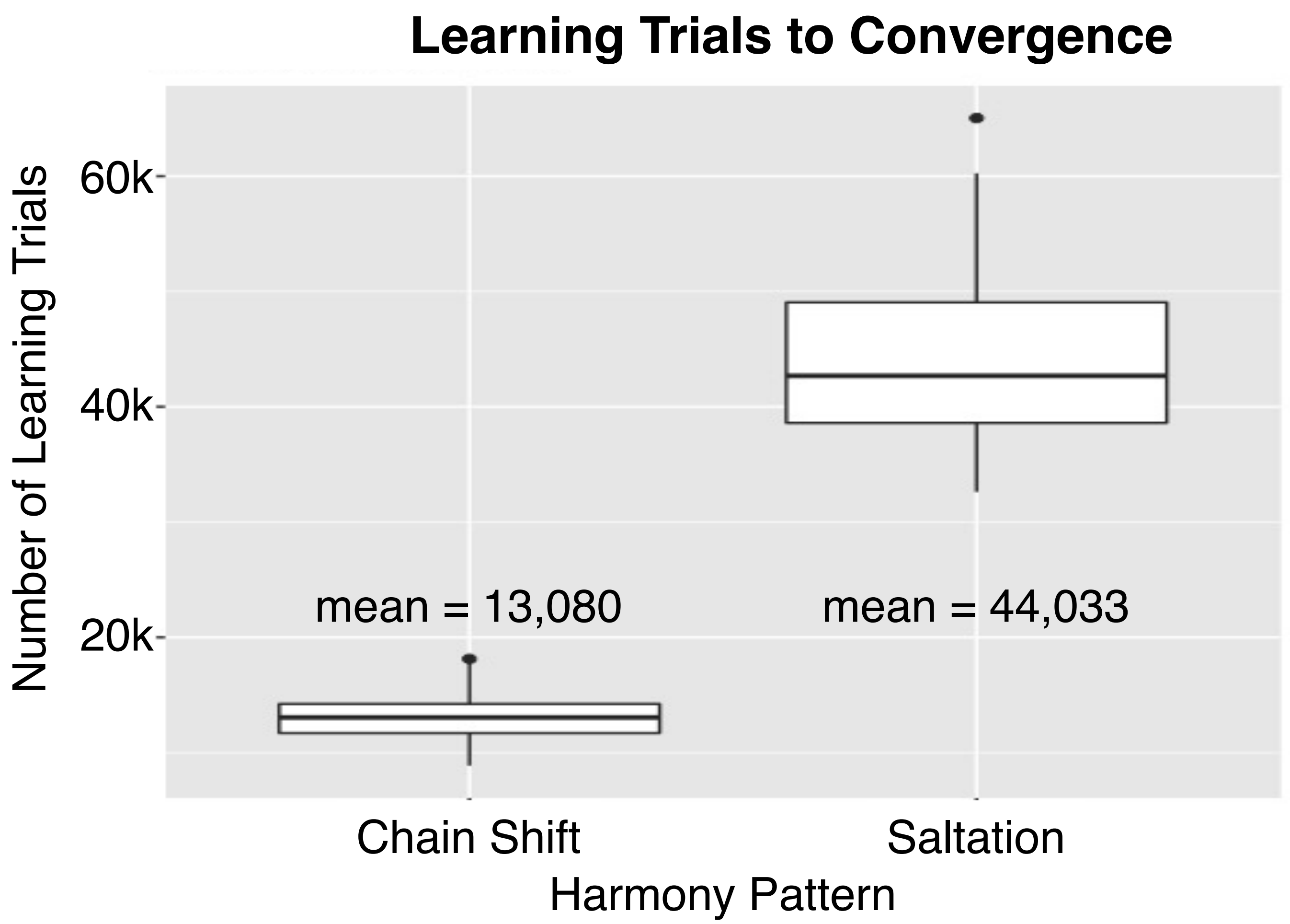
w(SELFDEACTIVATE) > w(PERSIST(ht.))

w(PERSIST(ht.)) + w(*GEST(TB high, self-deact.)) > w(SELFDEACTIVATE)

w(INHIBIT(TB upper, TB upper)) + w(INHIBIT(TB high, TB mid) > w(*INHIBIT)

Results: Time to Model Convergence

- Chain-shifting height harmony models converged substantially faster than saltatory harmony models
- Saltation takes ~3.4 times as many learning trials to learn
- Saltation is harder to learn, making it more likely to be mis-learned across generations and become less frequent typologically



Generational Stability Model
(O'Hara 2021)

- Iterated learning model (Kirby & Hurford 2002; Staubs 2014; Hughto 2020):
    - Learner of a MaxEnt grammar trained by comparing its productions to its teacher's
    - Learner matures and becomes teacher for new learner of next generation
    - Imperfect learning at each generation leads to pattern changes across generations

- Models transmissibility of phonological patterns: greater cross-generational stability leads to more robust attestation

Work In Progress: Generational Stability Model Utilizing GGGLA

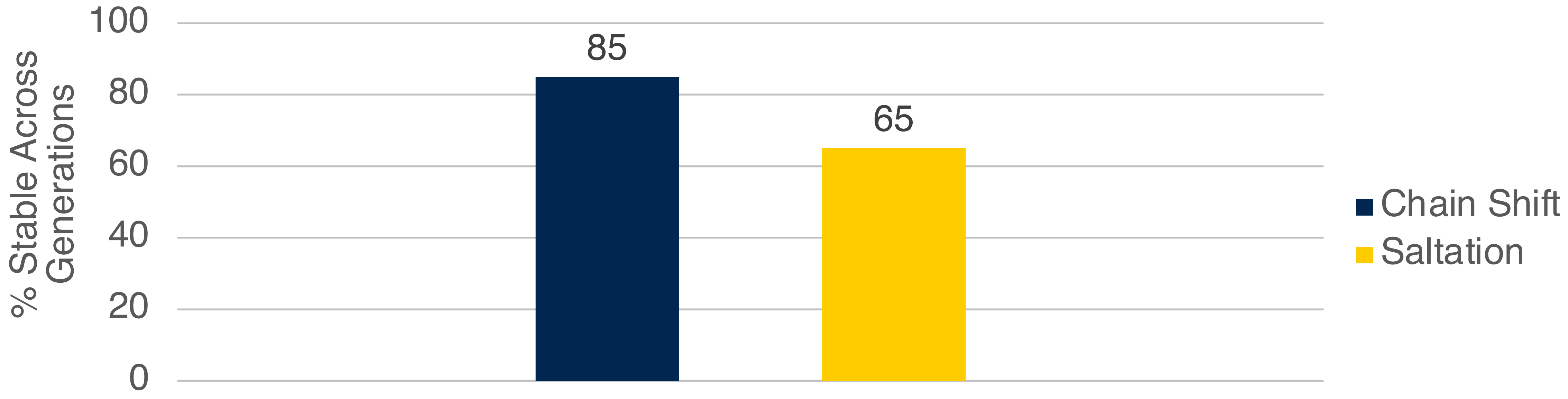
- Preliminary results based on 20 generational simulations per pattern type (chain-shifting versus saltatory)

- In progress: testing different numbers of learning trials per generation and numbers of generations for effects on effect size

- Current preliminary results: 50,000 learning trials per generation over 20 generations

Preliminary Results: Gestural Harmony Model + GGGLA

20 learning simulations with 20 generations and 50,000 learning trials per generation:



Summary

- Gestural Harmony Model generates both chain-shifting and saltatory height harmony
- Learning models based on GGGLA show that chain-shifting harmony is easier/faster to learn than saltatory harmony
- Learnability affects typology: patterns that are easier to learn (e.g. chain-shifting height harmony) are predicted to be more robustly attested crosslinguistically

Generating & Learning Chain-Shifting and Saltatory Height Harmony in Featural Frameworks

Chain Shifts and Saltations in Harmonic Grammar

- Chain shifts and saltations cannot be generated in Harmonic Grammar using the faithfulness constraints of Correspondence Theory (e.g., IDENT(F)-IO)

- Cumulative constraint interaction ('ganging') of faithfulness in Harmonic Grammar does not rule out multistep raising and cannot generate chain shifts (Albright et al. 2008; Farris-Trimble 2008)

- Ganging of markedness and faithfulness in Harmonic Grammar does not favor multistep raising and cannot generate saltations (White 2013; Hayes & White 2015; J. Smith in press)

Rethinking Faithfulness Constraints

(Tesar 2013; Magri 2018ab)

- Ability to generate underapplication opacity is characteristic of violation profiles of individual faithfulness constraints, not constraint interaction (Tesar 2013; Magri 2018ab)

- Violation profile of IDENT(F)-IO:

$$\text{IDENT}(/X/\rightarrow Y\rightarrow[Z]) = \text{IDENT}(/X/\rightarrow[Y]) + \text{IDENT}(/Y/\rightarrow[Z])$$

- IDENT violations incurred by less-faithful mapping are exactly those incurred by more-faithful component mappings

Rethinking Faithfulness Constraints

(Tesar 2013; Magri 2018ab)

- Chain shift requires constraint C that penalizes extra-unfaithful mapping *more* than its component more-faithful mappings:

C(/X/→Y→[Z]) > C(/X/→[Y]) + C(/Y/→[Z])

- Saltation requires constraint S that penalizes extra-unfaithful mapping *less* than its component more-faithful mappings:

S(/X/→Y→[Z]) < S(/X/→[Y]) + S(/Y/→[Z])

Rethinking Faithfulness Constraints

Alternative formulations of faithfulness fit violation profiles necessary to generate derivationally opaque chain shifts and saltations:

- Scalar and categorical faithfulness to scalar feature values (Gnanadesikan 1997)
- Distinct faithfulness constraints (*MAP(X,Y)) for all input-output mappings (Zuraw 2007; White 2013; Hayes & White 2015)

Testing Learnability of Opaque Harmony Processes in Featural Frameworks

- Error-driven learner of constraint weighting in MaxEnt grammars trained via Gradual Learning Algorithm to convergence (for each input, assigning ≥90% probability to intended winning candidate in target grammar)

- Generational stability of height harmony patterns tested via Generational Stability Model (O’Hara 2021)

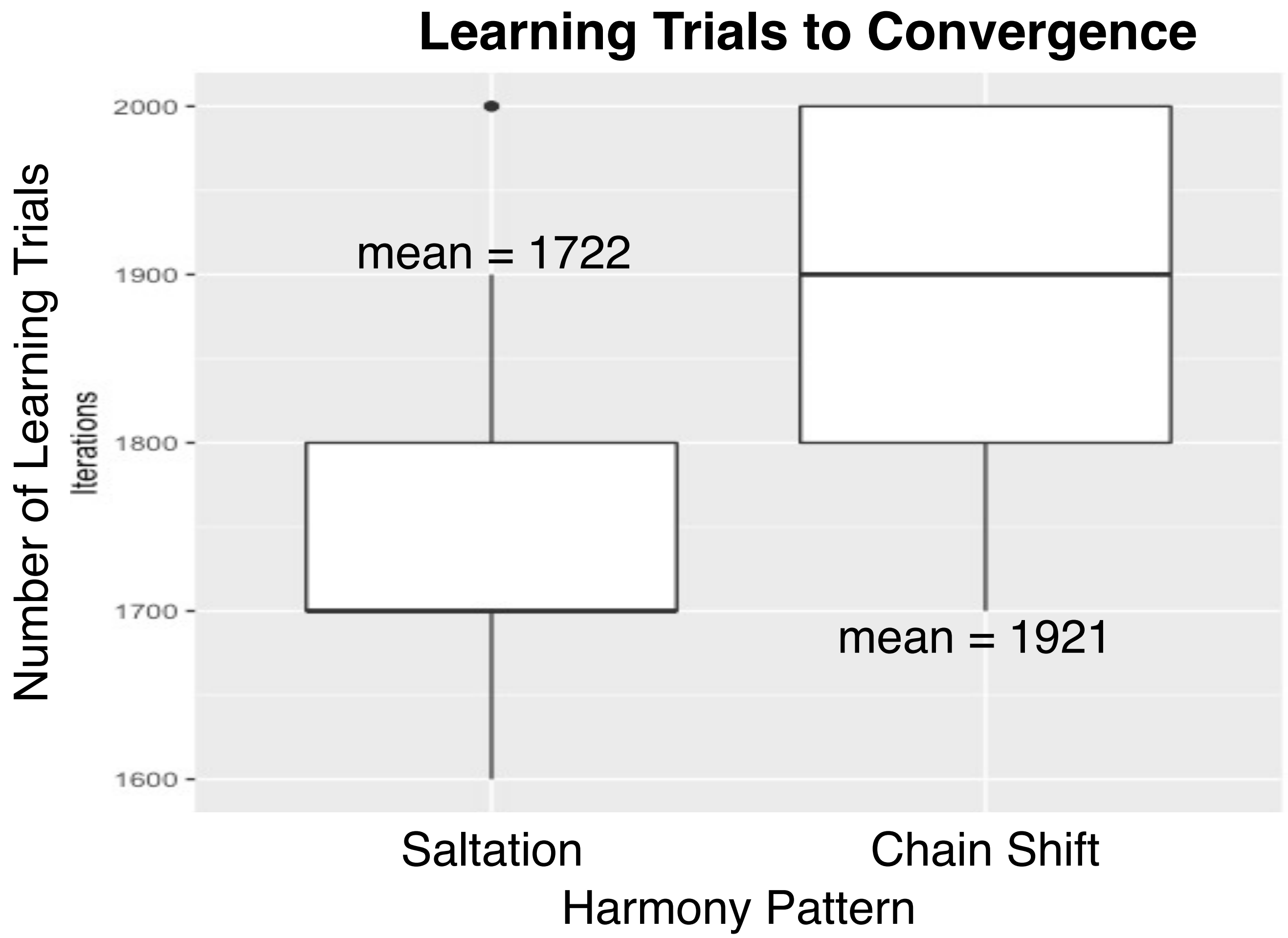
Learning Simulation Setup

- Two sets of simulations based on constraint sets capable of generating chain shifts and saltations:
  - Scalar/categorical markedness and faithfulness constraints from Feature Scales theory
  - Harmony-driving markedness constraints and distinct *MAP faithfulness constraints
- Greater initial weightings for *MAP constraints penalizing more steps of raising
- Three initial constraint weighting conditions for distinct faithfulness constraint simulations
  - M > F: Harmony-driving markedness over *MAP faithfulness
  - M ≈ F: Harmony-driving markedness equal to lowest-weighted *MAP faithfulness
  - M < F: Harmony-driving markedness below *MAP faithfulness

Scalar/Categorical Faithfulness Results: Time to Model Convergence

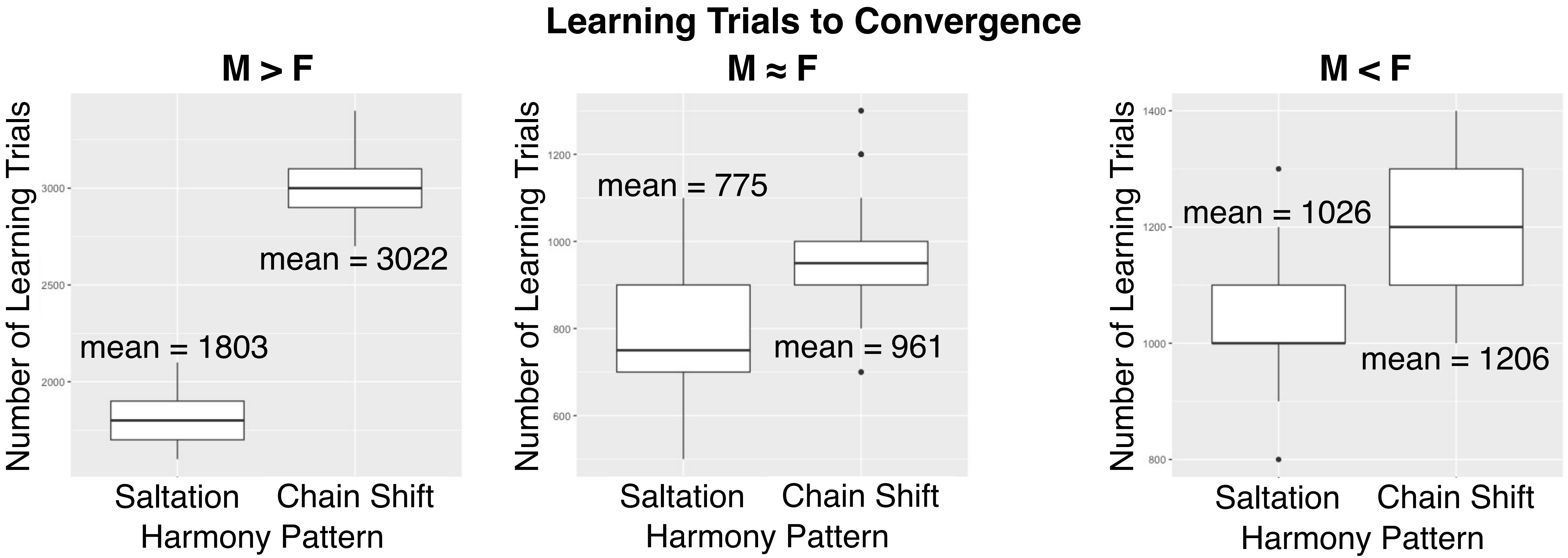
- Chain-shifting height harmony models converged more slowly than saltatory harmony models

- Incorrectly predicts that chain-shifting height harmony should be harder to learn



Distinct Faithfulness Results:
Time to Model Convergence

In all initial weighting conditions, chain-shifting height harmony models converged more slowly than saltatory harmony models

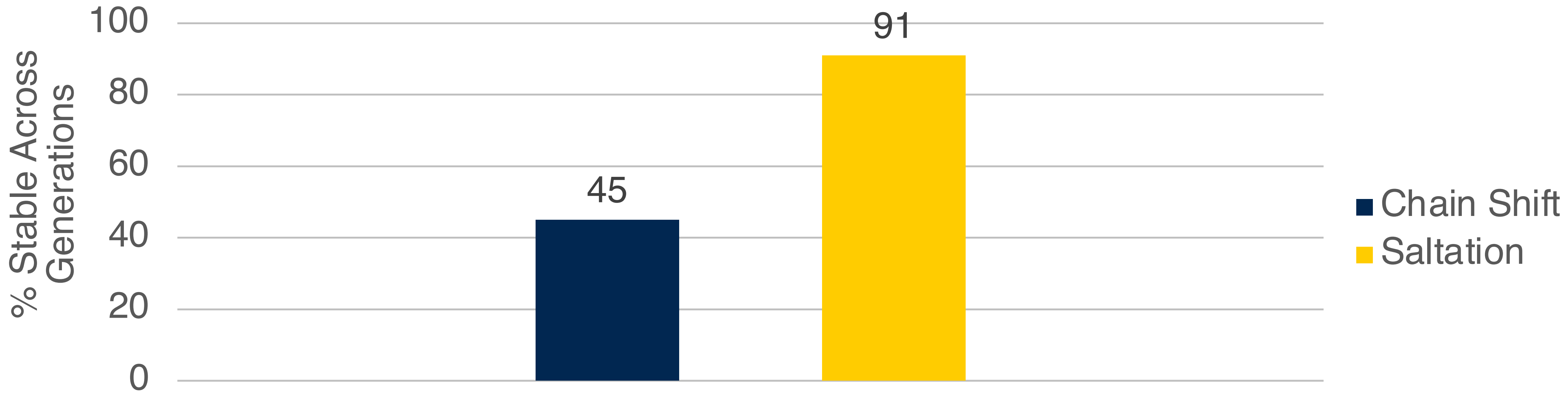


Generational Stability Model: Learning Simulation Setup

- 100 simulations per pattern type (chain-shifting versus saltatory) per constraint set (scalar/categorical versus distinct) per initial constraint weighting condition
- 20 generations per simulation
- Various numbers of learning trials per generation, depending on overall pattern learnability for each model type/initial weighting condition
  - 2,500 learning trails per generation for scalar/categorical faithfulness
  - 3,600 learning trials per generation for distinct faithfulness models with M > F initial weighting condition
  - 2,000 learning trails per generation for distinct faithfulness models with other initial weighting conditions

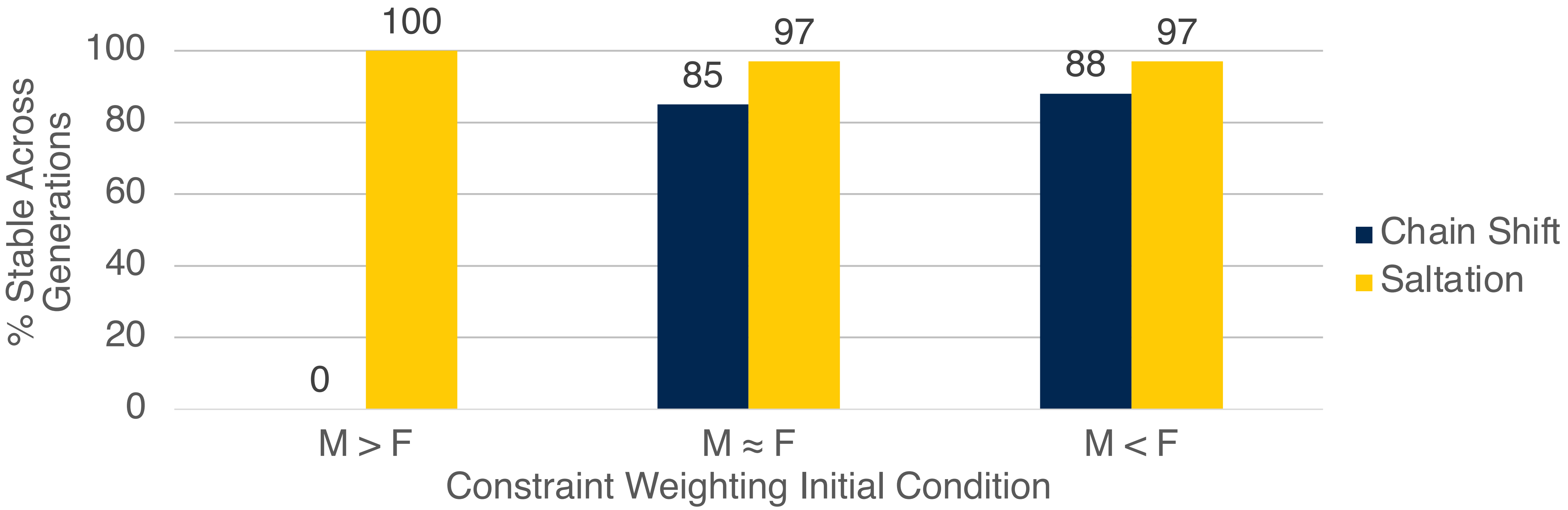
Scalar/Categorical Faithfulness Results: Generational Stability

Generational stability of patterns using scalar and categorical faithfulness constraints over feature scales:



Distinct Faithfulness Results: Generational Stability

Generational stability of patterns using distinct faithfulness constraints:



Summary

- Two feature-based approaches to generating underapplication opacity with featural representations in Harmonic Grammar

- Both scalar/categorical faithfulness and distinct faithfulness incorrectly predict:
  - Saltatory height harmony is easier/faster to learn and more stably transmitted across generations
  - Saltatory height harmony should be more widely attested crosslinguistically than chain-shifting height harmony

Conclusion

Conclusion

- Gestural Harmony Model is sufficiently powerful to generate apparently derivationally opaque chain-shifting and saltatory height harmony patterns

- Featural frameworks that eschew Correspondence Theory-based faithfulness constraints also powerful enough to generate derivationally opaque chain shifts and saltations

Conclusion

- Results of learning simulations using Grammar + Gesture Gradual Learning Algorithm correctly indicate a typological bias favoring attested chain-shifting harmony and against saltatory harmony

- Results of learning simulations in featural frameworks incorrectly indicate a bias favoring saltatory harmony and against chain-shifting harmony

References

Albright, Adam, Magri, Giorgio, & Michaels, Jennifer (2008) Modeling Doubly Marked Lags with a Split Additive Model. In H. Chan, H. Jacob, & E. Kapia (Eds.), *Proceedings of the 32nd Annual Boston University Conference on Language Development* (pp. 36–47). Cascadilla Press.

Baković, Eric (2007) A revised typology of opaque generalisations. *Phonology*, 24, 217–259.

Baković, Eric (2011) Opacity and Ordering. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (Second Edition, pp. 40–67). Blackwell Publishing.

Boersma, Paul (2003) Review: Bruce Tesar and Paul Smolensky (2000) Learnability in Optimality Theory. *Phonology*, 20(3), 436–446.

Boersma, Paul, & Pater, Joe (2016) Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar. In J. J. McCarthy & J. Pater (Eds.), *Harmonic Grammar and Harmonic Serialism*. Equinox.

Browman, Catherine P., & Goldstein, Louis (1986) Towards an Articulatory Phonology. *Phonology Yearbook*, 3, 219–252.

Browman, Catherine P., & Goldstein, Louis (1989) Articulatory gestures as phonological units. *Phonology*, 6(2), 201–251.

References

Clements, G. N. (1991) Vowel Height Assimilation in Bantu Languages. In K. Hubbard (Ed.), *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society: Special Session on African Language Structures* (pp. 25–64). Berkeley Linguistics Society.

Farris-Trimble, Ashley (2008) *Cumulative Faithfulness Effects in Phonology*. Doctoral Dissertation, Indiana University.

Fowler, Carol A., & Saltzman, Elliot (1993) Coordination and Coarticulation in Speech Production. *Language and Speech*, 36, 171–195.

Gnanadesikan, Amalia (1997) *Phonology with Ternary Scales*. Doctoral Dissertation, University of Massachusetts Amherst.

Goldwater, Sharon, & Johnson, Mark (2003) Learning OT Constraint Rankings Using a Maximum Entropy Model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory* (pp. 111–120).

Guthrie, Malcolm (1968) Notes on Nzεbi (Gabon). *Journal of African Languages*, 7(2), 101–129.

Hayes, Bruce, & White, James (2015) Saltation and the P-map. *Phonology*, 32, 267–302.

Hayes, Bruce, & Wilson, Colin (2008) A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry*, 39(3), 379–440.

References

Hualde, José (1988) *A Lexical Phonology of Basque*. Doctoral Dissertation, University of Southern California.

Hughto, Coral (2020) *Emergent Typological Effects of Agent-Based Learning Models in Maximum Entropy Grammar*. Doctoral Dissertation, University of Massachusetts Amherst.

Jäger, Gerhard (2007) Maximum Entropy Models and Stochastic Optimality Theory. In J. Grimshaw, J. Maling, C. Manning, J. Simpson, & A. Zaenen (Eds.), *Architectures, Rules, and Rreferences: A Festschrift for Joan Bresnan* (pp. 467–479). CSLI Publications.

Kirby, Simon, & Hurford, James R. (2002) The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model. In A. Cangelosi & D. Parisi (Eds.), S*imulating the Evolution of Language* (pp. 121–147). Springer.

Kirchner, Robert (1996) Synchronic Chain Shifts in Optimality Theory. *Linguistic Inquiry*, 27(2), 341–350.

Legendre, Géraldine, Miyata, Yoshiro, & Smolensky, Paul (1990) Can Connectionism Contribute to Syntax? Harmonic Grammar, with an Application. In K. Deaton, M. Noske, & M. Ziolkowski (Eds.), *Proceedings of the 26th Annual Meeting of the Chicago Linguistic Society* (pp. 237–252). Chicago Linguistic Society.

References

Magri, Giorgio (2018a) Idempotency, Output-Drivenness and the Faithfulness Triangle Inequality: Some Consequences of McCarthy's (2003) Categoricity Generalization. *Journal of Logic, Language and Information*, 27, 1–60.

Magri, Giorgio (2018b) Idempotency in Optimality Theory. *Journal of Linguistics*, 54(1), 139–187.

McCarthy, John J. (1999) Sympathy and phonological opacity. *Phonology*, 16, 331–399.

McCarthy, John J., & Prince, Alan S. (1995) Faithfulness and Reduplicative Identity. In J. Beckman, L. Walsh Dickey, S. Urbanczyk (Eds.), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory* (pp. 249–384). University of Massachusetts Graduate Linguistic Student Association.

Nettle, Daniel (1998) *The Fyem Language of Northern Nigeria*. Lincom Europa.

Nibert, Holly J. (1998) Processes of vowel harmony in the Servigliano dialect of Italian: A comparison of two non-linear proposals for the representation of vowel height. *Probus*, 10, 67–102.

O'Hara, Charles (2021) *Soft Biases in Phonology: Learnability Meets Grammar*. Doctoral Dissertation, University of Southern California.

Parkinson, Frederick Brooke (1996) *The Representation of Vowel Height in Phonology*. Doctoral Dissertation, The Ohio State University.

References

Pater, Joe, & Moreton, Elliott (2012) Structurally biased phonology: Complexity in learning and typology. *The EFL Journal*, 3(2), 1–44.

Prince, Alan, & Smolensky, Paul (1993/2004) *Optimality Theory: Constraint Interaction in Generative Grammar*. Published 2004 by Blackwell Publishing.

Rosenblatt, F. (1958) The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6), 386–408.

Saltzman, Elliot, & Munhall, Kevin G. (1989) A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology*, 1(4), 333–382.

Schmidt, Deborah Schlindwein (1996) Vowel raising in Basaa: A synchronic chain shift. *Phonology*, 13(2), 239–267.

Smith, Caitlin (2016) A gestural account of neutral segment asymmetries in harmony. In G. Ó. Hansson, A. Farris-Trimble, K. McMullin, & D. Pulleyblank (Eds.), *Proceedings of the 2015 Annual Meeting on Phonology.* Linguistic Society of America.

Smith, Caitlin (2017a) Harmony Triggering as a Segmental Property. In A. Lamont & K. Tetzloff (Eds.), *Proceedings of the 47th Annual Meeting of the North East Linguistic Society*. GLSA.

References

Smith, Caitlin (2017b) Harmony Triggering as a Contrastive Property of Segments.In K. Jesney, C. O'Hara, C. Smith, & R. Walker (Eds.), *Proceedings of the 2016 Annual Meeting on Phonology*. Linguistic Society of America.

Smith, Caitlin (2018) Harmony in Gestural Phonology. Doctoral Dissertation, University of Southern California.

Smith, Caitlin (2020a) Stepwise height harmony as partial transparency. In M. Asatryan, Y. Song, & A. Whitmal (Eds.), *Proceedings of the 50th Annual Meeting of the North East Linguistic Society* (pp. 141–154). GLSA.

Smith, Caitlin (2020b) Partial Height Harmony as Partial Transparency. In H. Baek, C. Takahashi, & A. H.-L. Yeung (Eds.), *Proceedings of the 2019 Annual Meeting on Phonology*. Linguistic Society of America.

Smith, Caitlin, & O'Hara, Charlie (2021) Learnability of derivationally opaque processes in the Gestural Harmony Model. In *Proceedings of the Society for Computation in Linguistics: Vol. 4* (pp. 396–400).

Smith, Jennifer L. (in press) Some Formal Implications of Deletion Saltation. To appear in *Linguistic Inquiry*.

Smolensky, Paul, & Legendre, Géraldine (Eds.) (2006) *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, MA: The MIT Press.

References

Stanton, Juliet (2016) Learnability shapes typology: The case of the midpoint pathology. *Language*, 92(4), 753–791.

Staubs, Robert (2014) *Computational Modeling of Learning Biases in Stress Typology*. Doctoral Dissertation, University of Massachusetts Amherst.

Tesar, Bruce (2013) *Output-Driven Phonology: Theory and Learning*. Cambridge University Press.

Tesar, Bruce, & Smolensky, Paul (1998) Learnability in Optimality Theory. *Linguistic Inquiry*, 29(2), 229–268.

White, James (2013) *Bias in Phonological Learning: Evidence from Saltation*. Doctoral Dissertation, University of California Los Angeles.

Zuraw, Kie (2007) The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from tagalog infixation. *Language*, 83(2), 277–316.

Appendix: Gestural Harmony Model Constraint Definitions

Gestural Harmony Model Constraints
(Smith 2018)

- PERSIST(height): Assign a violation to a TB upper surface gesture that is self-deactivating. (harmony driver)

- *GEST(TB high, self-deactivating): Assign a violation to a gesture specified for TB upper surface high and self-deactivation. (high vowel-specific harmony driver)

- SELFDEACTIVATE: Assign a violation to a persistent gesture. (penalizes harmony)

- *INHIBIT: Assign a violation to an inhibition relation. (penalizes blocking)

Gestural Harmony Model Constraints
(Smith 2018)

- INHIBIT(TB upper, TB upper): Assign a violation to a pair of TB upper surface gestures without an inhibition relation between them. (block harmony between TB upper surface gestures)
- INHIBIT(TB high, TB mid): Assign a violation to a pair of TB upper surface high and mid gestures without an inhibition relation between them. (block harmony between TB high and mid gestures)
- INHIBIT(TB high, TB low): Assign a violation to a pair of TB upper surface high and low gestures without an inhibition relation between them. (block harmony between TB high and low gestures)
- INHIBIT(TB mid, TB low): Assign a violation to a pair of TB upper surface mid and low gestures without an inhibition relation between them. (block harmony between TB mid and low gestures)

Appendix: Generating Chain-Shifting and Saltatory Height Harmony in Featural Frameworks (Expanded Discussion)

Chain Shifts and Saltations in Harmonic Grammar

- Chain shifts and saltations cannot be generated in Harmonic Grammar using the faithfulness constraints of Correspondence Theory (e.g., IDENT(F)-IO)
- Cumulative constraint interaction ('ganging') of faithfulness in Harmonic Grammar does not rule out multistep raising and cannot generate chain shifts (Albright et al. 2008; Farris-Trimble 2008)
- Ganging of markedness and faithfulness in Harmonic Grammar does not favor multistep raising and cannot generate saltations (White 2013; Hayes & White 2015; Smith to appear)

Chain Shifts in Harmonic Grammar with IDENT

(Albright et al. 2008; Farris-Trimble 2008)

Input: /e-i/	HARMONY(Height) w=3	IDENT(high) w=2	IDENT(ATR) w=2	$\mathcal{H}$
☞ a. [i-i]		-1		-2
b. [e-i]	-1			-3

Input: /ɛ-i/	HARMONY(Height) w=3	IDENT(high) w=2	IDENT(ATR) w=2	$\mathcal{H}$
☹ a. [i-i]		-1	-1	-4
Intended winner → b. [e-i]	-1		-1	-5
c. [ɛ-i]	-2			-6

Shared violation, no *asymmetric tradeoff* (Pater 2009)

Saltation in Harmonic Grammar with IDENT

(White 2013; Hayes & White 2015; Smith to appear)

Input: /e-i/	IDENT(high) w=5	HARMONY(Height) w=4	IDENT(ATR) w=2	$\mathcal{H}$
a. [i-i]	-1			-5
☞ b. [e-i]		-1		-4

Input: /ɛ-i/	IDENT(high) w=5	HARMONY(Height) w=4	IDENT(ATR) w=2	$\mathcal{H}$
a. [i-i]	-1		-1	-7
☹ b. [e-i]		-1	-1	-6
c. [ɛ-i]		-2		-8

Intended winner → a. [i-i]

Shared violation, no asymmetric tradeoff

Rethinking Faithfulness Constraints

(Tesar 2013; Magri 2018ab)

- Ability to generate underapplication opacity is characteristic of violation profiles of individual faithfulness constraints, not constraint interaction (Tesar 2013; Magri 2018ab)

- Violation profile of IDENT(F)-IO:

IDENT(/X/→Y→[Z]) = IDENT(/X/→[Y]) + IDENT(/Y/→[Z])

- IDENT violations incurred by less-faithful mapping are exactly those incurred by more-faithful component mappings

Rethinking Faithfulness Constraints

(Tesar 2013; Magri 2018ab)

- Chain shift requires constraint C that penalizes extra-unfaithful mapping *more* than its component more-faithful mappings:

$$C(/X/\rightarrow Y\rightarrow[Z]) > C(/X/\rightarrow[Y]) + C(/Y/\rightarrow[Z])$$

- Saltation requires constraint S that penalizes extra-unfaithful mapping *less* than its component more-faithful mappings:

$$S(/X/\rightarrow Y\rightarrow[Z]) < S(/X/\rightarrow[Y]) + S(/Y/\rightarrow[Z])$$

Rethinking Faithfulness Constraints

Alternative formulations of faithfulness fit violation profiles necessary to generate derivationally opaque chain shifts and saltations:

- Scalar and categorical faithfulness to scalar feature values (Gnanadesikan 1997)
- Distinct faithfulness constraints (*MAP(X,Y)) for all input-output mappings (Zuraw 2007; White 2013; Hayes & White 2015)

Feature Scales Theory
(Gnanadesikan 1997)

- Feature Scales Theory: specific feature values represented by position on feature scale

- Ternary vowel height scale:

  High = 1 Mid = 2 Low = 3

- Quaternary vowel height scale:

  High = 1 High-Mid = 2 Low-Mid = 3 Low = 4

Scalar and Categorical Faithfulness

(Gnanadesikan 1997)

Multiple versions of the featural faithfulness constraint IDENT:

- IDENT(X): Given an input segment A and its correspondent output segment B, then A and B must have values on scale X that are identical.

- IDENT-ADJACENT(X): Given an input segment A and its correspondent output segment B, then A and B must have values on scale X that are identical or adjacent.

- IDENT-PARTIAL(X): Given an input segment A and its correspondent output segment B, then A and B must have values on scale X that are identical, adjacent, or within-two.

Violation Profiles of Scalar and Categorical Faithfulness Constraints

Input: /a/	IDENT(Height)	IDENT-ADJ(Height)	IDENT-PART(Height)
a. [i]	*	*	*
b. [e]	*	*	
c. [ɛ]	*		
d. [a]			

IDENT(X) penalizes slightly unfaithful mappings just as much as very unfaithful mappings

IDENT-ADJ(X) and IDENT-PART(X) characterize mappings as ‘faithful enough’ or ‘too unfaithful’ rather than ‘faithful’ or ‘unfaithful’

Generating a Chain Shift with Scalar Faithfulness

One-step raising does not violate IDENT-ADJ and IDENT-PART, but does better satisfy harmony-driving ASSIM constraints:

Input: /a-i/	IDENT-ADJ(Ht.) w=3	IDENT-PART(Ht.) w=2	ASSIM-PART(Ht.) w=2	ASSIM-ADJ(Ht.) w=2	ASSIM(Ht.) w=2	IDENT(Ht.) w=1	$\mathcal{H}$
a. [i-i]	-1	-1				-1	-6
b. [e-i]	-1				-1	-1	-6
☞ c. [ɛ-i]				-1	-1	-1	-5
d. [a-i]			-1	-1	-1		-6

Input: /ɛ-i/	IDENT-ADJ(Ht.) w=3	IDENT-PART(Ht.) w=2	ASSIM-PART(Ht.) w=2	ASSIM-ADJ(Ht.) w=2	ASSIM(Ht.) w=2	IDENT(Ht.) w=1	$\mathcal{H}$
a. [i-i]	-1					-1	-4
☞ b. [e-i]					-1	-1	-3
c. [ɛ-i]				-1	-1		-4

Generating Saltation with Categorical Faithfulness

IDENT(Height) penalizes all raising equally, motivating two-step raising to satisfy harmony-driving ASSIM constraints:

Input: /ɛ-i/	IDENT(Ht.) w=3	IDENT-PART(Ht.) w=3	ASSIM-ADJ(Ht.) w=3	ASSIM(Ht.) w=2	ASSIM-PART(Ht.) w=2	IDENT-ADJ(Ht.) w=1	$\mathcal{H}$
☞ a.[i-i]	-1					-1	-4
b.[e-i]	-1			-1			-5
c.[ɛ-i]			-1	-1			-5

Input: /e-i/	IDENT(Ht.) w=3	IDENT-PART(Ht.) w=3	ASSIM-ADJ(Ht.) w=3	ASSIM(Ht.) w=2	ASSIM-PART(Ht.) w=2	IDENT-ADJ(Ht.) w=1	$\mathcal{H}$
a.[i-i]	-1						-3
☞ b.[e-i]				-1			-2

Distinct Faithfulness

- *MAP constraints (Zuraw 2007; White 2013; Hayes & White 2015) assign distinct violation profiles to every input-output mapping

- *MAP(X,Y): Assign a violation when a segment that is a member of class X is in correspondence with a segment of class Y.

Generating a Chain Shift with Distinct Faithfulness

*MAP(a,i), *MAP(a,e), and *MAP(ɛ,i) penalize only multi-step raising:

Input: /a-i/	*MAP(a,i) w=6	*MAP(a,e) w=4	*MAP(ɛ,i) w=4	HARMONY(high) w=2	HARMONY(ATR) w=2	HARMONY(low) w=2	$\mathcal{H}$
a. [i-i]	-1						-6
b. [e-i]		-1		-1			-6
☞ c. [ɛ-i]				-1	-1		-4
d. [a-i]				-1	-1	-1	-6

Input: /ɛ-i/	*MAP(a,i) w=6	*MAP(a,e) w=4	*MAP(ɛ,i) w=4	HARMONY(high) w=2	HARMONY(ATR) w=2	HARMONY(low) w=2	$\mathcal{H}$
e. [i-i]			-1				-4
☞ f. [e-i]				-1			-2
g. [ɛ-i]				-1	-1		-4

Generating Saltation with Distinct Faithfulness

*MAP(e,i) penalizes only one-step raising from high-mid to high:

Input: /ɛ-i/	*MAP(e,i) w=2	HARMONY(high) w=1	HARMONY(ATR) w=1	*MAP(ɛ,e) w=1	*MAP(ɛ,i) w=1	$\mathcal{H}$
☞ a. [i-i]					-1	-1
b. [e-i]		-1		-1		-2
c. [ɛ-i]		-1	-1			-2

Input: /e-i/	*MAP(e,i) w=2	HARMONY(high) w=1	HARMONY(ATR) w=1	*MAP(ɛ,e) w=1	*MAP(ɛ,i) w=1	$\mathcal{H}$
c. [i-i]	-1					-2
☞ b. [e-i]		-1				-1