Learning Derivationally Opaque Patterns in the Gestural Harmony Model

Caitlin Smith and Charlie O'Hara
(in revision for *Phonology* 10/29/21*)*

In this paper, we examine the learnability of two apparently derivationally opaque vowel harmony patterns: attested chain-shifting height harmony and unattested saltatory height harmony. We analyze these patterns within the Gestural Harmony Model (Smith 2018) and introduce a learning algorithm for setting the gestural parameters that generate these harmony patterns. Results of the learning model indicate a learning bias in favor of the attested chain-shifting pattern and against the unattested saltation pattern, providing a potential explanation for the differences in attestation between these two derivationally opaque patterns. Furthermore, we show that feature-based learning models of these patterns show no such learning bias and provide no account of the typological asymmetry between chain-shifting and saltatory height harmony.

1. Introduction

In partial vowel height harmony, nonhigh vowels are raised, approaching the height of a high vowel trigger without necessarily reaching it. In a language with multiple vowel heights, there are multiple possible patterns that partial raising may follow. One possibility is that each vowel raises one 'step' along what can be considered a scale of vowel height; such patterns are sometimes referred to as stepwise height harmonies.

One example of stepwise height harmony comes from Nzebi, a Bantu language spoken in Gabon (Guthrie 1968; Clements 1991; Kirchner 1996; Parkinson 1996). In Nzebi, the suffix /-i-/ occurs immediately after verb stems in some tenses. This high vowel triggers one-step raising of preceding non-reduced stem vowels, a process Guthrie (1968) refers to as *yotization*. As part of this raising process, the high-mid vowels /e/ and /o/ surface faithfully when no high suffix vowel follows, as in (1a-c), but surface as [i] and [u], respectively, before suffix [i], as in (1d-f).

(1)	a. [be̲t]	d. [bi̲t-i]	'carry'
	b. [βo̲ːm]	e. [βu̲ːm-i]	'breathe'
	c. [ko̲lən]	f. [ku̲lin-i]	'go down'

The low-mid and low vowels also undergo raising when followed by suffix /-i-/, but do not surface as high vowels themselves. Underlying low-mid /ɛ/ and /ɔ/ surface faithfully in (2a-d), and as high-mid [e] and [o], respectively, before suffix [i] in (2e-h).

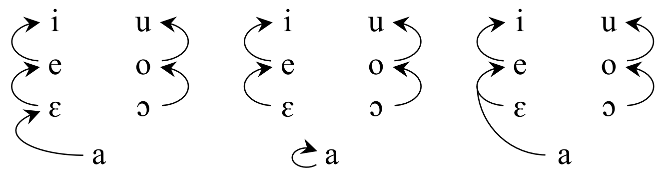
(2)	a. [sɛ̲b]	e. [se̲b-i]	'laugh'
	b. [suɛ̲m]	f. [sue̲m-i]	'hide self'
	c. [mɔ̲n]	g. [mo̲n-i]	'see'
	d. [tɔ̲ːd]	h. [to̲ːd-i]	'arrive'

Finally, underlying /a/ also surfaces faithfully in (3a-b), but undergoes single-step raising to surface as low-mid [ɛ] before suffix [i], as in (3c-d).

(3) a. [sa̲l] c. [sɛ̲l-i] 'work'
b. [tsia̲t] d. [tsiɛ̲t-i] 'trample'

This sort of stepwise height harmony process resembles a synchronic chain shift. In a chain-shifting phonological process, an underlying segment /X/ maps to surface [Y], while underlying /Y/ maps to surface [Z]. In Nzebi, height harmony creates a chain shift whereby /a/ → [ɛ], /ɛ/ → [e], and /e/ → [i]; while /ɔ/ → [o] and /o/ → [u].

Similar patterns of chain-shifting vowel raising are well-attested beyond this example from Nzebi. In Servigliano Italian, a similar pattern of one-step raising occurs among mid vowels (Camilli 1929; Kaze 1989; Nibert 1998; Mascaró 2011; Walker 2011). In Basaa, morphologically-conditioned stem vowel mutation results in chain-shifting raising as well (Schmidt 1996). These patterns of vowel raising are illustrated in Figure (1).

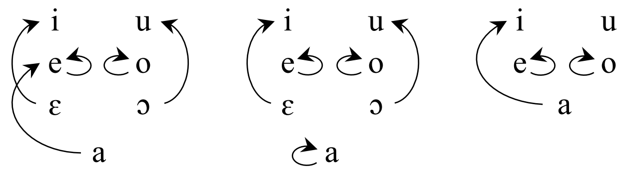


(a) *(b)* *(c)*

Figure (1): Patterns of vowel raising observed in (a) Nzebi height harmony; (b) Servigliano Italian metaphony; (c) Basaa morphological raising

All of these chain-shifting vowel raising patterns are examples of what McCarthy (1999) and Baković (2007, 2011) refer to as *underapplication opacity*. Underapplication is broadly defined as a situation in which a phonological process appears not to have applied when it should have, i.e. despite its structural description having been met. For example, we can envision the pattern of vowel raising harmony in Nzebi as the result of a set of several vowel raising processes by which high-mid vowels raise to high, low-mid to high-mid, and low to low-mid. High-mid to high raising applies to underlying /e/ and /o/, deriving [i] and [u], respectively. However, the same process does not apply to the [e] and [o] that are derived from underlying /ɛ/ and /ɔ/ by low-mid to high-mid raising. Likewise, low-mid to high-mid raising applies to underlying /ɛ/ and /ɔ/, but underapplies to the [ɛ] derived from /a/ by low to low-mid raising.

While stepwise, chain-shifting vowel raising patterns are well-attested, Parkinson (1996) observes that similar patterns involving exclusively two-step raising are apparently unattested. Several logically possible but unattested two-step vowel raising patterns are provided in Figure (2). In all of these hypothetical patterns, vowels raise either two steps in height or not at all, resulting in instances of *saltation*, a type of phonologically-derived environment effect. In a saltatory phonological process, an underlying segment /X/ maps to surface [Z] rather than [Y], despite X being more phonologically similar to Y than to Z. Meanwhile, underlying /Y/ maps faithfully to surface [Y]. For example, in Figure (2a), /ɛ/ raises to [i], 'skipping over' more similar [e], while underlying /e/ surfaces faithfully.



(a) *(b)* *(c)*

Figure (2): Unattested two-step vowel raising patterns

Like chain shifts, saltations also represent examples of underapplication opacity. Again, we can envision the pattern of vowel raising in Figure (2a) as the result of a set of single-step vowel raising processes. For /ɛ/ to raise to [i], it must undergo two vowel raising processes: one raising low-mid vowels to high-mid, and one raising high-mid vowels to high. However, in order to derive a saltation the second of these processes must crucially apply only to high-mid vowels derived by low-mid to high-mid raising. The process can then be described as underapplying to underlying high-mid vowels.

In both chain shifts and saltations, then, whether a process applies depends on the underlying versus intermediate status of a segment. Such a scenario presents difficulty for output-oriented phonological frameworks such as Optimality Theory (Prince & Smolensky 1993/2004) and Harmonic Grammar (Legendre et al. 1990; Smolensky & Legendre 2006), in which there are no intermediate forms and a segment's status as either underlying or derived therefore does not factor into the evaluation of output candidates. The inability of Optimality Theory to generate many derivationally opaque patterns, including chain shifts and saltations, was uncovered quickly, as discussed in many of the papers compiled by Roca (1997), as well as by Kirchner (1996), McCarthy (1999), Łubowicz (2002), Moreton (2004), and Baković (2007), among others. Harmonic Grammar, in which constraints are weighted rather than strictly ranked, is similarly unable to generate chain shifts (Albright et al. 2008; Farris-Trimble 2008) and saltations (White 2013; Hayes & White 2015; Smith to appear).

In this paper, we propose an analysis of the chain-shifting partial height harmony of Nzebi that is situated within the Gestural Harmony Model (Smith 2018). In this model, the subsegmental units of phonological representation are gestures, the goal-based, dynamically-defined units of the framework of Articulatory Phonology (Browman and Goldstein 1986, 1989, et seq.). Vowel harmony is the result of a gesture extending to overlap the gestures of other segments in a word. We propose that cases in which vowels seem to partially undergo harmony, as in Nzebi height harmony, result from competition between gestures with conflicting articulatory targets. In doing so, we cast Nzebi height harmony as a derivationally transparent process. The same approach can be applied to the types of saltatory height harmonies depicted in Figure (2). However, we show that within the Gestural Harmony Model, predictions regarding the relative *learnability* of chain-shifting and saltatory harmony are consistent with the attestedness of chain-shifting harmony and the unattestedness of saltatory harmony. We introduce the Gestural Gradual Learning Algorithm and present the results of learning models that utilize it.

We also compare our learning results to those of feature-based computational learning models of chain-shifting and saltatory height harmony. While neither type of pattern is generable in either Optimality Theory or Harmonic Grammar utilizing the typical faithfulness constraints of Correspondence Theory (McCarthy & Prince 1995), several approaches have been developed to capture such phenomena. These include the adoption of constraints on featural scales (Gnanadesikan 1997) and of faithfulness constraints on specific input-output mappings rather than individual feature changes (Zuraw 2007; White 2013). Like the Gestural Harmony Model, these theories of faithfulness can generate both chain shifts and saltations. However, we show that while an analysis situated within the Gestural Harmony Model can appeal to learnability to explain the unattestedness of saltatory harmony, there is no such learning bias against saltatory patterns under these alternative featural approaches.

The paper is organized as follows. Section 2 provides an introduction to the phonological unit of representation assumed within Articulatory Phonology, the gesture, and describes the Gestural Harmony Model. Section 3 provides an analysis of the chain-shifting height harmony of Nzebi in that framework and shows how that framework can also be used to generate saltatory height harmony. Section 4 introduces the Gestural Gradual Learning Algorithm and presents the results of computationally modeling the learning of both chain-shifting and saltatory height harmony in the Gestural Harmony Model. Section 5 compares those results to those of learning models based on featural frameworks that are powerful enough to generate these derivationally opaque patterns. Section 6 concludes.

2. Representing Harmony with Gestures

2.1. Gestures as Phonological Units

In our analysis of Nzebi height harmony, we assume that the units of subsegmental representation are gestures, as in Articulatory Phonology (Browman & Goldstein 1986, 1989, et seq.). In this framework, each gesture is specified for a target articulatory state, the achievement of which unfolds over time according to a dynamically-defined equation of motion. The length of time over which a gesture commands one or more articulators in the vocal tract to achieve its target state is its period of activation. When enough time has passed for its target state to be achieved, a gesture deactivates, allowing its articulators to return to their specified neutral positions until they are recruited by subsequent gestures.

A gesture's target articulatory state is specified in terms of a primary articulator, a constriction location, and a constriction degree. The constriction location of a consonantal gesture is specified as some point along the static surface of the vocal tract. Constriction degree refers to the aperture of the constriction between the primary articulator and the constriction location. For instance, a dorsal stop may be characterized in part by a gesture with a target articulatory state in which a closure (the constriction degree) is formed between the tongue body (the primary articulator) and the velum (the constriction location).

Like dorsal consonants, vowels are usually assumed to include a tongue body gesture with a target constriction location specifying a point along the static surface of the vocal tract (e.g., palate, velum, uvula). However, this assumption does not admit the direct encoding of vowel height and backness, dimensions along which vowel are typically phonologically active. Therefore, we diverge from the standard gestural representation of vowels. According to the new representational scheme that we propose, vocalic gestures are still specified for constriction location and degree, just as consonantal gestures are. However, rather than specifying constriction location as a specific point along the vocal tract, we propose that each vowel is composed of two tongue body gestures whose constriction locations are specified for wider regions of the vocal tract. One of these gestures has a target constriction location comprising much of the upper surface of the vocal tract, including the palate and velum. The target articulatory state of this gesture is to make a constriction anywhere in this region. Its constriction degree determines a vowel's height; a narrow constriction at the upper surface produces a high vowel, while a wide constriction produces a low vowel. The second of these vowel gestures has a target constriction location comprising much of the back surface of the vocal tract, including the uvula and pharyngeal wall. The constriction degree of a tongue body back surface constriction gesture determines a vowel's backness. A narrow constriction in this region produces a back vowel, while a wide constriction produces a front vowel. The target constriction locations for these vowel gestures are illustrated in Figure (3).

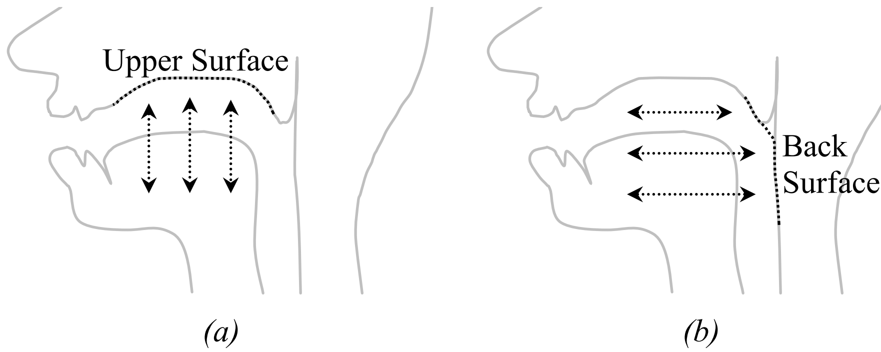


Figure (3): (a) Constriction location for tongue body upper surface gesture; (b) constriction location for tongue body back surface gesture

Adopting these constriction location regions for vocalic gestures allows for vowel height and backness to be encoded directly by gestural representations. The ability to gesturally represent vowel height is particularly important for formulating an analysis of Nzebi height harmony within the Gestural Harmony Model.

2.2. The Gestural Harmony Model

In the Gestural Harmony Model (Smith 2016, 2018), vowel and vowel-consonant harmony is the result of a gesture extending its activation period to overlap the gestures of other segments in a domain. To achieve this extended activation, Smith proposes that each gesture's representation includes parameters determining whether it self-activates at its specified starting point as well as whether it self-deactivates once it reaches its target articulatory state. These parameters determine whether a gesture is a trigger of harmony.

The effects of these gestural parameter specifications are illustrated in Figure (4). The top gesture is a typical self-activating and self-deactivating lip protrusion gesture, which is used to represent lip rounding. When this gesture reaches its target articulatory state (protruded lips), it self-deactivates. The middle gesture is a *persistent* gesture, which does not self-deactivate once it has reached its target articulatory state. Instead, it remains active, extending to overlap the gestures of following segments, thus triggering progressive (rightward) rounding harmony. To illustrate this extension, the dashed line indicates the point at which the gesture reaches its target articulatory state but does not deactivate. The bottom gesture is an anticipatory, or early-activating, gesture. The dashed line indicates the point at which the gesture is scheduled to start according to its position in a word. However, since it is an anticipatory gesture it has activated before that point, extending to overlap the gestures of preceding segments, thus triggering regressive (leftward) rounding harmony.

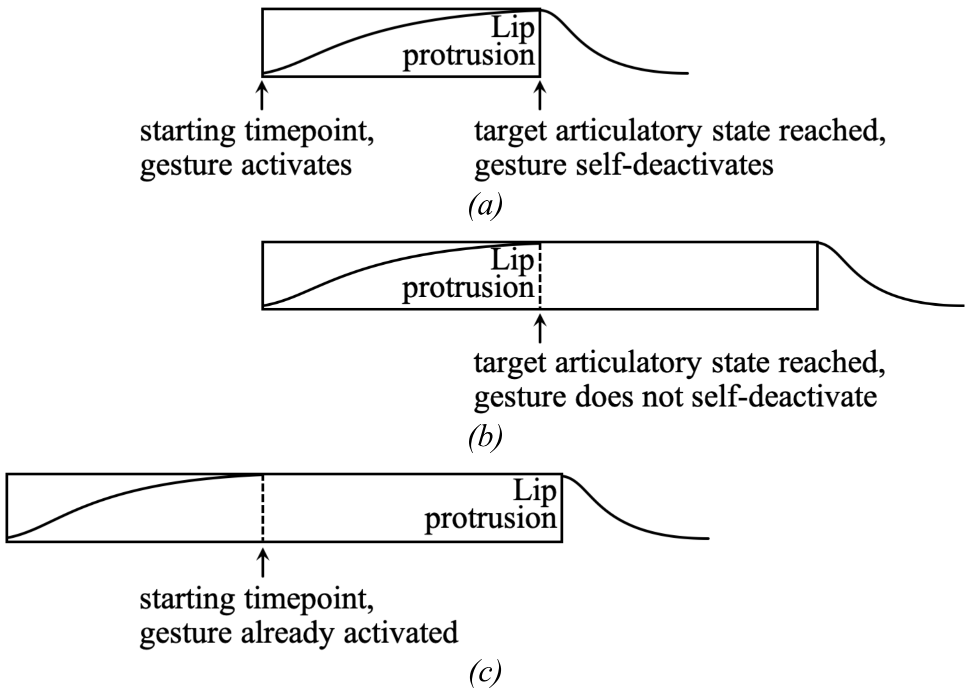


Figure (4): (a) Typical, (b) persistent, and (c) anticipatory lip protrusion gestures

In the Gestural Harmony Model, harmony arises when a segment includes a gesture that is either persistent, anticipatory, or both; that segment is the trigger of harmony. Other segments undergo harmony when their composite gestures are overlapped by a harmonizing gesture.

Figure (5) illustrates the workings of the model with a *gestural score* for an [o-o] sequence, derived from underlying /o-ɤ/ by rounding harmony. The segmental transcription is provided along the top, and the subscript for each segment matches the subscripts of its composite gestures. In this gestural score, both vowels include a tongue body upper surface gesture with a wide constriction degree, indicating that they are nonhigh. In addition, they each include a tongue body back surface gesture with a narrow constriction degree, indicating that they are back vowels. The first [o] also includes a persistent lip protrusion gesture that overlaps the gestures of the following vowel, which surfaces as rounded as a result. The time course of lip protrusion below the gestural score indicates that the lips reach their target protruded state and remain there throughout the word. This is the basic representation of harmony within this model: overlap by a single, uninterrupted harmonizing gesture with an extended period of activation.

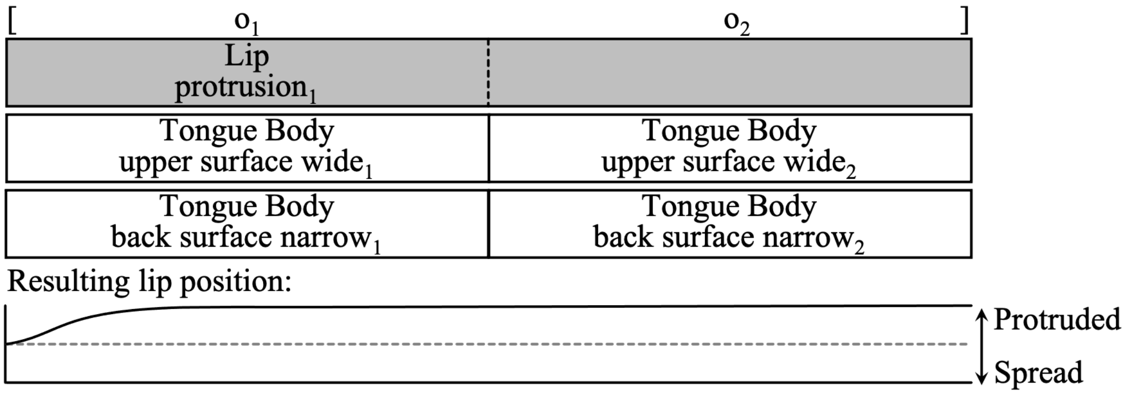


Figure (5): Rounding harmony due to overlap by a persistent lip protrusion gesture

In some languages, certain segments may be *transparent* to a harmony process, appearing to have been skipped by the spread of a harmonizing property. For instance, if the vowel /i/ is transparent to a rounding harmony process, then an underlying vowel sequence such as /o-i-ɤ/ would surface as [o-i-o], with the third vowel surfacing as rounded while the medial vowel is seemingly unaffected. To account for such transparency, the Gestural Harmony Model takes advantage of the fact that gestures are goal-based units; while they are specified for a target articulatory state, they might not necessarily achieve it successfully. In this model, transparent segments are considered a special type of undergoer; rather than being skipped over by a harmony process, transparent segments are overlapped by a harmonizing gesture just as typical undergoer segments are. Transparency arises when gestural overlap results in *antagonism*, a situation in which two concurrently active gestures have opposing target articulatory states. Examples of antagonistic pairs of gestures include concurrently active gestures for lip protrusion and lip spreading, velum opening and velum closure, and narrow and wide constrictions at the upper surface of the vocal tract.

Because the concurrent achievement of two conflicting target articulatory states is not possible, antagonistic gestures are essentially in competition with one another for control of the vocal tract. The outcome of the competition between antagonistic gestures is determined by the relative *blending strengths* that are specified for each gesture. According to the Task Dynamic Model of speech production (Saltzman and Munhall 1989; Fowler and Saltzman 1993), when gestural antagonism occurs, it is resolved by blending the competing target articulatory states of these gestures to create an intermediate target state that holds during the period of their concurrent activation. This blended target state is the weighted average of two gestures' individual target articulatory states, and the weighting in this averaging function is contributed by the gestures' strength parameters, denoted α. The gestural blending function is provided in (4).

(4) Gestural blending function in the Task Dynamic Model of speech production

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

Figure (6) contains the gestural score for an [o-i-o] sequence, illustrating transparency to rounding harmony via gestural blending. The persistent lip protrusion gesture of the first [o] overlaps the gestures of all other vowels. The third vowel in the sequence surfaces as rounded [o] due to this overlap. However, the medial high front /i/ surfaces as unrounded [i] rather than rounded

[y], despite also being overlapped by the lip protrusion gesture. This is because the representation of /i/ includes a lip spreading gesture in addition to its tongue body gestures. This lip spreading gesture is antagonistic to the harmony-triggering lip protrusion gesture and is specified for a relatively high gestural strength. In this example, its strength is ten times the strength of the harmonizing gesture. This allows the lip spreading gesture to counteract the effect of the harmonizing lip protrusion gesture upon the state of the vocal tract during the period of their concurrent activation.

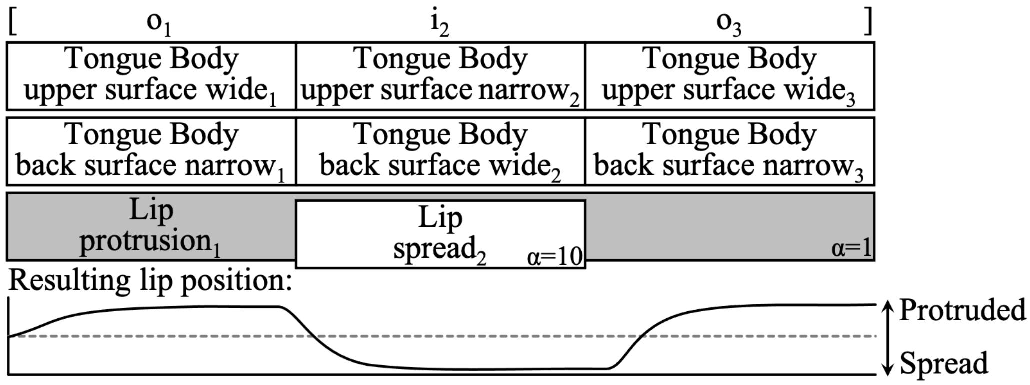


Figure (6): Transparency due to relatively strong antagonistic lip spreading gesture active during production of relatively weak harmonizing lip protrusion gesture

There are several advantages to the representation of transparency to harmony as the result of the concurrent activation of antagonistic gestures. First, it correctly predicts that only those segments that include a gesture that is antagonistic to a harmonizing gesture may be transparent within a given type of harmony. Smith (2016, 2018) claims that this prediction is important for typological reasons: in rounding harmony and in nasal harmony, the sets of cross-linguistically attested transparent segments are limited to the classes of segments that include gestures that are antagonistic to a harmonizing gesture, as suggested by instrumental study. In addition, this model provides a representation of harmony in which spreading is completely local; harmony does not skip segments, even those that surface as transparent.

The full transparency of high front [i] to rounding harmony in Figure (6) is dependent on the strength of its antagonistic lip spreading gesture being much greater than that of the harmonizing lip protrusion gesture. However, their strengths are not categorically ‘strong’ and ‘weak’ but rather are defined numerically. In this example, the strength ratio between the two gestures is 10-to-1. Because of this numerical definition of gestural strength, there is no restriction within the Gestural Harmony Model dictating that for a given pair of overlapped gestures, one must be much stronger than the other. Another possible scenario predicted by the model is one in which a harmonizing gesture and an overlapped antagonistic gesture have similar or even identical strengths. When such a scenario arises, the model generates a case of blending resulting in partial transparency and partial undergoing of harmony, as illustrated in Figure (7).

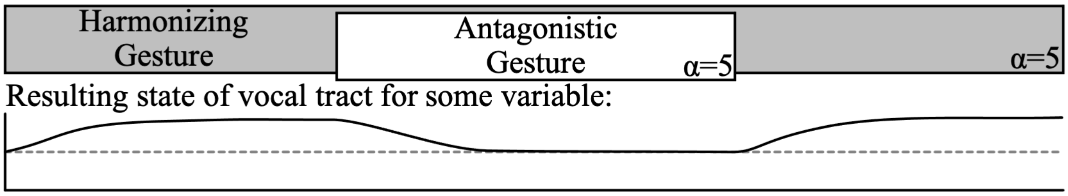


Figure (7): Partial transparency resulting from equal strengths of harmonizing and antagonistic gestures

We propose that partial vowel height harmony represents just such a case of partial transparency. In the following section, we adopt the concept of gestural blending in our analysis of the partial, chain-shifting height harmony evident in Nzebi within the Gestural Harmony Model.

3. A Gestural Analysis of Height Harmony

3.1. Chain-Shifting Harmony in Nzebi

As discussed in section 0, Nzebi exhibits a case of partial, chain-shifting vowel raising harmony. In this harmony pattern, high-mid undergoers fully assimilate to the height of the high front vowel trigger of harmony, while low-mid and low vowels only partially assimilate to the trigger height. The raising pattern in Nzebi is illustrated in Figure (8), repeated from Figure (1) above.

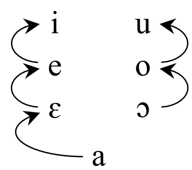


Figure (8): Pattern of vowel raising observed in Nzebi height harmony

We propose that this pattern be treated as a case of partial transparency to harmony and analyzed as the result of gestural blending within the Gestural Harmony Model. In contrast with full transparency, in which the antagonistic gesture of a transparent segment is much stronger than the harmonizing gesture, partial height harmony represents a case in which the two blended gestures are closer in strength.

This analysis relies on blending of the target constriction degrees of tongue body upper surface gestures for vowels of different heights. Each of the four vowel heights observed in Nzebi is represented by a tongue body upper surface gesture with one of four possible constriction degrees: narrow, narrow-mid, wide-mid, and wide. We analyze the vowel raising harmony in Nzebi as the result of overlap by an anticipatory upper surface gesture with narrow constriction degree that is part of the representation of the yotizing suffix high vowel /i/. Under this analysis, vowels that are specified for a wide-mid or wide constriction and that appear to be partially transparent to this vowel raising harmony are able to partially resist the raising effect of the triggering narrow vowel gesture because they are of similar blending strengths. The weaker high-mid vowels, on the other hand, surface as high rather than resisting raising, suggesting that they have a relatively much lower blending strength.

Rather than simply referring to 'relatively strong' and 'relatively weak' gestures, this analysis provides a hand-picked set of precise gestural blending strengths that produce the desired surface vowel constriction degrees for Nzebi raising harmony when input to the blending function in (4). Table (1) provides the vowels of each height with target constriction degrees for their tongue body upper surface gestures, as well as proposed strength values. The precise strength values proposed here are not crucial. Rather, it is the strength ratios between gestures that are important to the analysis.

Vowel	Target Constriction Degree	Strength (α)
/i/, /u/	narrow (4 mm)	10
/e/, /o/	narrow-mid (8 mm)	1
/ɛ/, /ɔ/	wide-mid (12 mm)	10
/a/	wide (16 mm)	20

Table (1): Target constriction degrees and gestural blending strengths for tongue body upper surface gestures of Nzebi vowels[1]

As a result of gestural blending, narrow-mid vowels /e/ and /o/ fully undergo harmony and surface as raised when overlapped by the anticipatory upper surface gesture of suffix /i/. This is achieved by specifying that /e/ and /o/ have a much lower strength than /i/. As shown in (5), the gestural blending function produces a blended target constriction degree of 4.36 mm, very similar to the 4 mm constriction degree of a narrow vowel trigger listed in Table (1).

(5) Gestural blending of narrow and narrow-mid vowels

$$4.36\text{ mm} = \frac{\overbrace{4\text{ mm}\cdot 10}^{/\text{i}/} + \overbrace{8\text{ mm}\cdot 1}^{/\text{e}/,\ /\text{o}/}}{10+1}$$

This blending is illustrated by the gestural score in Figure (9) for the upper surface gestures of a Nzebi word such as [bit-i] 'carry,'[2] in which the first [i] is the result of raising of an underlying /e/. The accompanying time course of tongue body height indicates that the underlying narrow-mid vowel /e/ fully undergoes harmony and surfaces as [i] when it is overlapped by the anticipatory upper surface gesture of a following, stronger narrow vowel [i].

[1] Unlike /i/, the high vowel /u/ is not a trigger of raising harmony and does not overlap other vowels. As a result, we cannot determine the blending strength of /u/. We have opted here to provide it with a blending strength identical to its fellow high vowel /i/, but this assumption is not crucial.
[2] In this and all following gestural scores, to reduce visual clutter we show only the vowels' tongue body upper surface gestures, though we assume that all vowels also include a tongue body back surface gesture and that round vowels include a lip protrusion gesture.

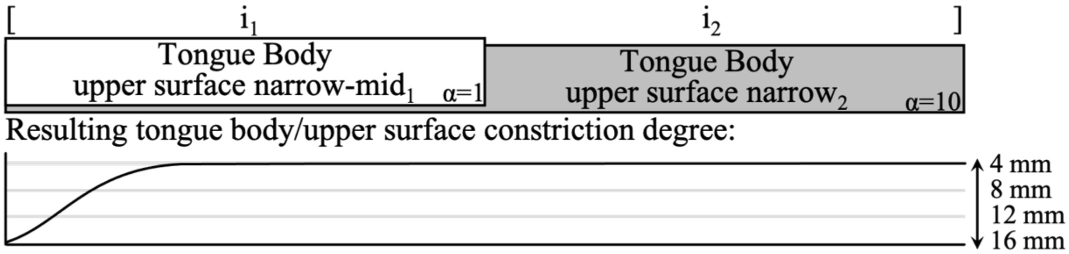


Figure (9): Gestural score for [i-i] sequence illustrating narrow-mid to narrow vowel raising due to gestural blending

Wide-mid and wide vowels, on the other hand, resist fully undergoing harmony, but also do not surface as fully transparent to harmony. Focusing first on the wide-mid vowels, partial transparency is produced by providing narrow trigger /i/ and wide-mid undergoers /ɛ/ and /ɔ/ with equal strengths. The blending function does not favor the target articulatory state of one gesture over the other, but instead returns a target constriction degree of 8 mm, as shown in (6). This blended target is intermediate between the two vowels' antagonistic target constriction degrees and consistent with that of the underlying narrow-mid vowels, producing one-step raising.

(6) Gestural blending of narrow and wide-mid vowels

$$8\text{ mm} = \frac{\overbrace{4\text{ mm} \cdot 10}^{/\text{i}/} + \overbrace{12\text{ mm} \cdot 10}^{/\text{ɛ}/,\ /\text{ɔ}/}}{10 + 10}$$

As a result of this blending calculation, /ɛ/ and /ɔ/ only partially undergo harmony and surface as partially raised. The gestural score in Figure (10) includes the upper surface gestures of a word such as [seb-i] 'laugh,' in which the [e] is the result of raising. The time course of tongue body height shows that the underlyingly wide-mid vowel /ɛ/ only partially undergoes harmony, surfacing as [e] when it is overlapped by the upper surface gesture of a following narrow vowel [i].

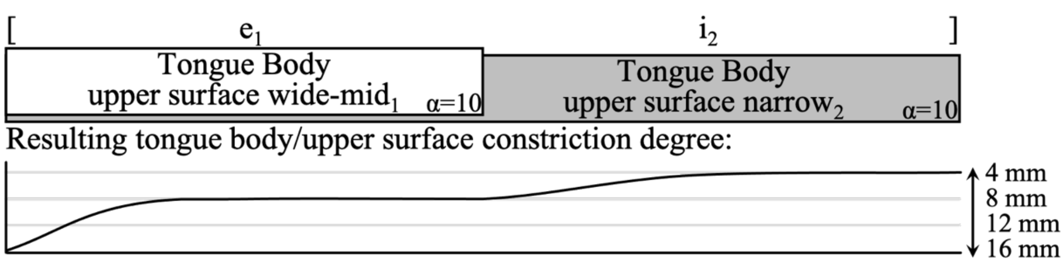


Figure (10): Gestural score for [e-i] sequence illustrating wide-mid to narrow-mid vowel raising due to gestural blending

Finally, the low vowel /a/ also resists fully undergoing harmony and surfaces as partially transparent. However, rather than surfacing with a blended target constriction degree that is halfway between the two antagonistic target constriction degrees, the blending function in (7) favors the target articulatory state of the stronger low vowel. This produces a target constriction degree of 12 mm, consistent with the target of the wide-mid vowels.

(7) Gestural blending of narrow and wide vowels

$$12\text{ mm} = \frac{\overbrace{4\text{ mm}\cdot 10}^{/\text{i}/} + \overbrace{16\text{ mm}\cdot 20}^{/\text{a}/}}{10+20}$$

The somewhat higher blending strength of the upper surface gesture of wide /a/ relative to narrow /i/ is not sufficient to produce full transparency to harmony, but rather one-step raising. This is illustrated by the gestural score and time course of tongue body height in Figure (11) for a word such as [sɛl-i] 'work.'[3]

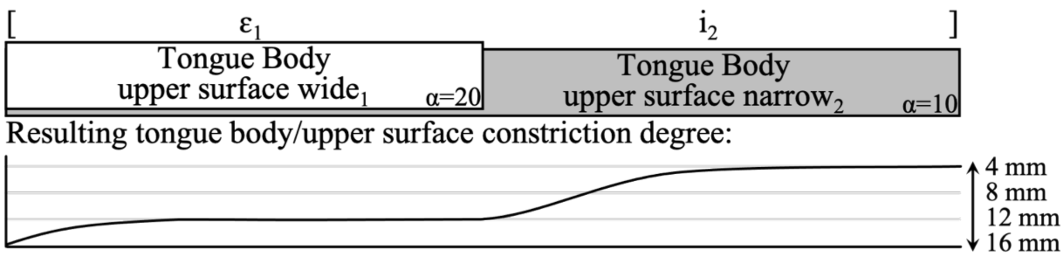
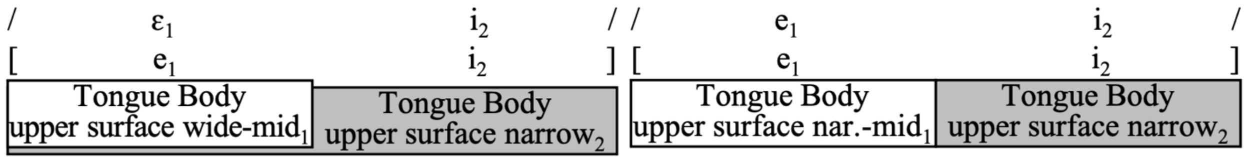


Figure (11): Gestural score for [ɛ-i] sequence illustrating low to wide-mid vowel raising due to gestural blending

This section has demonstrated that when vocalic upper surface gestures of different constriction degrees are provided with the proper strength values, the mechanism of gestural blending can be recruited by the Gestural Harmony Model to produce blended target constriction degrees that are consistent with the chain-shifting vowel raising pattern in Nzebi. Due to overlap by the anticipatory upper surface gesture of a harmony-triggering anticipatory narrow vowel, underlyingly narrow-mid vowels are fully raised and surface as narrow, while underlyingly wide-mid and wide vowels only partially undergo raising. All of these outcomes arise from their upper surface gestures' individual specifications for constriction degree and blending strength.

As discussed in section 1, synchronic chain shifts pose a challenge for most feature-based analyses in output-oriented Optimality Theory and Harmonic Grammar. The success of the Gestural Harmony Model in accounting for the apparently chain-shifting height harmony of Nzebi is due to its not actually being represented as a chain shift, but rather as a derivationally transparent pattern. In Nzebi, gestural blending between trigger and undergoer vowel gestures produces the effect of vowel raising. However, the individual gestures that make up the underlying representation of each vowel are still present and unaltered in the gestural score of a surface form, as shown in Figure (12).

[3]The /a/ → [ɛ] mapping appears to involve not only raising but fronting of /a/. We assume that /a/ is not specified as a back vowel, but rather that its apparent centrality/backness stems from the hinge-like movement of the jaw, with some vowel backing occurring automatically for vowels that involve a lower jaw height.



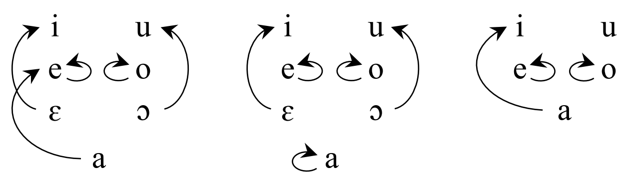
(a) *(b)*

Figure (12): (a) An [e-i] sequence derived from an underlying /ɛ-i/ sequence via gestural overlap; (b) An [e-i] sequence surfacing faithfully from an underlying /e-i/ sequence

While an [e] derived by gestural overlap and blending between /ɛ/ and /i/, as in Figure (12a), and an [e] that surfaces faithfully from an underlying /e/, as in Figure (12b), may be articulatorily and acoustically the same, their gestural makeups are different in both the phonological input and output. Such an approach is unavailable to featural representations of this process. In a feature-based account, a surface [e] that is derived from /ɛ/ is featurally indistinguishable from one that surfaces faithfully from /e/. Therefore, it is difficult to explain why a high-mid to high raising process underapplies to an [e] derived from /ɛ/. In the Gestural Harmony Model, on the other hand, this underapplication opacity is only apparent, and chain-shifting height harmony need not be modeled as a derivationally opaque process.

3.2. Saltatory Height Harmony

With a different set of strength value settings, the Gestural Harmony Model is also able to generate unattested saltatory height harmonies such as those provided in Figure (2), repeated in Figure (13).



(a) *(b)* *(c)*

Figure (13): Unattested two-step vowel raising patterns

Table (2) provides a set of hand-picked gestural strengths that generate a two-step raising process consistent with the one depicted in Figure (13a). As in the analysis of chain-shifting height harmony, it is the strength ratios between the different vowels that are crucial to deriving the pattern, rather than the exact values we provide.

Vowel	Target Constriction Degree	Strength (α)
/i/, /u/	narrow (4 mm)	25
/e/, /o/	narrow-mid (8 mm)	350
/ɛ/, /ɔ/	wide-mid (12 mm)	1
/a/	wide (16 mm)	20

Table (2): Target constriction degrees and gestural blending strengths for upper surface gestures of vowels in a saltatory height harmony system

As the values in Table (2) indicate, the strength values necessary to generate saltatory two-step raising are more extreme than those necessary to generate chain-shifting one-step raising. In Nzebi, the greatest proposed strength necessary to generate chain-shifting height harmony was 20 for the low vowel /a/. In contrast, the greatest proposed strength necessary to generate saltatory harmony in Table (2) is 350 for the high-mid vowels /e/ and /o/. To understand the reason for this difference, we introduce the idea of *overpowering* relations between blended gestures. In order for a segment X to fully assimilate to a segment Y via gestural overlap, the strength of Y's gesture must be exponentially higher than that of X. In other words, Y must *overpower* X. Conversely, in order for segment X to fully resist assimilation to a segment Y, the strength of X's gesture must be exponentially higher than that of Y; X must overpower Y. When one segment only partially assimilates to another due to gestural overlap, no overpowering relation exists between them. Thus, an exponential difference in strength is not necessary, and their gestural strengths may be more similar.

In order to estimate the greatest gestural strength necessary to produce a certain harmony pattern, we can examine chains of overpowering relations that exist between the gestures of vowels in an inventory. Each additional overpowering relation in a chain indicates an order of magnitude of strength that must be reached by the strongest gesture in the vowel inventory. For instance, in the saltatory height harmony in Figure (13a), if the narrow vowels /i/ and /u/ act as triggers of raising harmony, they must be overpowered by narrow-mid /e/ and /o/, as these vowels completely resist assimilation to the narrow vowels. In addition, the narrow vowels must themselves overpower wide-mid /ɛ/ and /ɔ/, as those vowels completely assimilate to the narrow vowels. Therefore, the longest chain of overpowering relations among vowels in the saltatory harmony pattern is two links long, indicating that the strongest vowel(s) in the inventory must include a gesture that is about two orders of magnitude stronger than that of the weakest vowel(s).

In a chain-shifting height harmony triggered by high vowels /i/ and /u/, on the other hand, no vowels completely resist assimilation, and only the narrow-mid vowels /e/ and /o/ assimilate completely to the high vowel triggers of harmony. Therefore, high vowels /i/ and /u/ must only overpower the narrow-mid vowels, and the longest chain of overpowering relations in the chain-shifting harmony pattern is one link long. This indicates that the strongest vowel(s) in the inventory must be approximately one order of magnitude stronger than the weakest vowel(s).

The Gestural Harmony Model, then, is able to generate two types of apparently derivationally opaque height harmony patterns via the settings of the relative strength values of vowel gestures. As with chain-shifting height harmony, the model's ability to generate saltatory height harmony arises from the fact that the process arises from gestural overlap and blending, and is represented as a derivationally transparent pattern, despite its appearance as a case of underapplication opacity.

On its own, the Gestural Harmony Model provides no explanation for the difference in crosslinguistic attestation between these two types of derivationally opaque types of partial height harmony. While chain-shifting height harmony is fairly well-attested, saltatory height harmony is apparently unattested. In the following section, we show that the difference in the extremeness of the gestural strengths necessary to generate saltatory and chain-shifting height harmony has important consequences for the relative learnability, and therefore the robustness of attestation, of chain-shifting and saltatory height harmony patterns.

4. The Gestural Gradual Learning Algorithm

The analysis of chain-shifting Nzebi height harmony presented in section 3.1 depends on a fairly precise set of strength ratios between the tongue body upper surface gestures of vowels at different heights. In this section, we address how a learner of this height harmony system acquires these gestural strength settings. In addition, we examine whether a learner is capable of acquiring a set of gestural strengths that produces an unattested saltatory height harmony system, and whether there is an emergent *learning bias* against saltatory height harmony that could explain its lack of crosslinguistic attestation.

We adopt the view that the attestation of a phonological pattern is not only impacted by which patterns are generable by a given grammatical framework, but also how learnable that pattern is within that framework. Any learning algorithm that efficiently searches a large space of possible phonological grammars is inherently biased toward some over others. According to this view, easier-to-learn patterns are learned more rapidly, requiring exposure to fewer data points in order to be learned accurately. If learners are exposed to a finite number of input data, harder-to-learn patterns that require exposure to more data points are more susceptible to being mislearned, and are more likely to change across generations. This sort of learning bias (or analytic bias) has been argued to be responsible for a variety of asymmetries in phonological typology (Wilson 2006, Moreton 2008, Pater & Moreton 2012, Staubs 2014, Hayes & White 2015, Jarosz 2016, Stanton 2016, Hughto 2020, McCollum et al. 2020, O'Hara 2021). Typologically, we expect harder-to-learn patterns to be less well-attested than their easy-to-learn counterparts, all else held equal.

4.1. Learning Simulation Setup

To address the question of whether a learning bias is responsible for the lack of attestation of saltatory vowel height harmony, we designed a gesture-based computational learning model. In this model, the learner is tasked with setting the constriction degree targets and blending strengths for vowel gestures such that the learner accurately reproduces its teacher's height harmony pattern. In addition, we tasked the learner with setting these same gestural parameter values for the dorsal consonant /g/, whose primary oral gesture uses the tongue body as its primary articulator and is therefore blended with concurrently active vowel gestures. The constriction degrees of /g/ and the vowels are in conflict: while /g/ is specified for closure, all vowels are specified for different degrees of openness. The result of this conflict is blending of these gestures' individual target constriction degrees.

We tested the learner on two types of harmony in languages with a four-height vowel inventory: (1) a Nzebi-like pattern of one-step raising before high vowel triggers /i/ and /u/, and (2) an unattested pattern of two-step saltatory raising before high vowel triggers /i/ and /u/ similar to the one depicted in Figure (13a) in section 3.2. We assume that the learner has already learned a phonological grammar that produces regressive height harmony due to a high vowel's anticipatory tongue body upper surface gesture overlapping the gestures of previous segments. Here, the learning task is to determine the settings of each dorsal segment's constriction degree and blending strength such that that overlap results in either chain-shifting or saltatory height harmony. All words in the training data for both pattern types were of shape (V)CV, with vowels and consonants coming from the inventory in Table (3).

Segment	Primary Articulator	Constriction Location	Constriction Degree
/i/, /u/	tongue body	upper surface	4 mm
/e/, /o/	tongue body	upper surface	8 mm
/ɛ/, /ɔ/	tongue body	upper surface	12 mm
/a/	tongue body	upper surface	16 mm
/g/	tongue body	velum	-2 mm
/b/	lower lip	upper lip	-2 mm

Table (3): Segment inventory of training dataset[4]

The learner utilizes an error-driven learning algorithm that we introduce here: the Gestural Gradual Learning Algorithm, which is defined as in (8).

(8) The Gestural Gradual Learning Algorithm[5]

1. Initialize each gesture in the learner's inventory with a target constriction degree of 16 mm (i.e., all segments start as /a/) and a random strength (between 1 and 20).
2. On each iteration, randomly generate a (V)CV sequence.
3. Check for gestural blending:
   a. If V_2 is a trigger of harmony, it overlaps V_1, resulting in blending.
   b. V_2 overlaps a preceding C. If C is a dorsal /g/, this overlap results in blending.
4. If the learner produces an error (a segment with a constriction degree more than 0.2 mm from the teacher's production):
   a. Update the constriction degree target of the learner's tongue body gestures to produce a constriction degree that better matches the teacher's output.
   b. In cases of blending, update the strength of the learner's tongue body gestures to produce a constriction degree that better matches the teacher's output.

On each training iteration, a random (V)CV sequence is generated based on a uniform distribution of the vowels and consonants in Table (3). Both the learner and the teacher produce this sequence based on their current gestural parameter settings. The learner compares its production to its teacher's, and updates its gestural constriction degree targets and/or strengths when it makes an error. An error is considered to have occurred when a learner produces a constriction degree that does not fall within a set window around the teacher's produced constriction degree for a given segment. Here, we use a window of 0.2 mm.

We illustrate the workings of this algorithm with two sample learning iterations. In the first, depicted in Figure (14), the teacher produces the underlying CVC sequence /abi/ as [ɛbi], with one-step raising resulting from the presence of a harmony-triggering V_2. The learner, on the other hand, produces this form as [a̞bi], with a wider constriction degree than the teacher's during production of V_1. Because this form includes a harmony trigger and therefore involves blending of the two vowel gestures, the learner must update the target constriction degrees and strengths of both vowel gestures such that, when blended, they produce a vowel with a narrower constriction between the tongue body and the upper surface of the vocal tract. Specifically, the learner updates the constriction degree targets for both /a/ and /i/ so that they will each be produced with greater

[4] A negative target constriction degree is often assumed for stop consonants in Articulatory Phonology. While the achievement of such a constriction degree is of course not physically possible, this 'virtual target' allows models of speech to achieve the kinematics and tight closure consistent with the production of stop consonants.

[5] A Python implementation of this algorithm is available at [[web address redacted for anonymity]].

constriction; both of these updates will also produce a higher tongue body position when /a/ and /i/ are blended. In addition, the learner updates the strengths of these gestures, increasing the strength of /i/ and decreasing the strength of /a/. Again, both of these updates produce a higher tongue body position when /a/ and /i/ are blended.

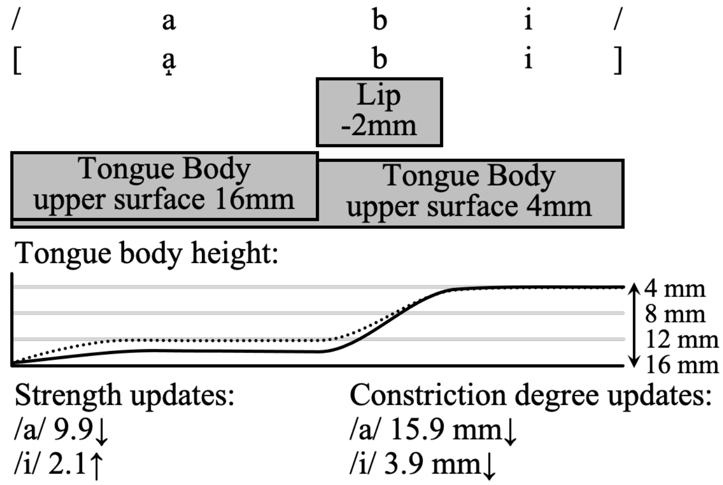


Figure (14): Sample learning trial comparing learner's production (solid) and teacher's production (dashed) of underlying /abi/

In the next example, depicted in Figure (15), the teacher produces the underlying CVC sequence /ɛɡa/ as [ɛɡa], with no raising of the first vowel. The learner produces this form as [ɛ̝ɣa], with a narrower constriction degree than the teacher for V_1 and a wider constriction degree for the dorsal consonant. Because the learner's production of [ɛ̝] depends only on the vowel's constriction degree target and involves no blending between vowel gestures, the learner updates only that target for the tongue body gesture of /ɛ/ by increasing (i.e. widening) its target constriction degree. However, blending does occur between the dorsal consonant /g/ and the second vowel /a/ during the production of the consonant. The result is the velar approximant [ɣ] rather than the stop [ɡ] produced by the teacher. Therefore, the learner updates both the constriction degree targets and strengths of the gestures of /g/ and /a/ such that they produce a narrower constriction degree when blended. The target constriction degrees of both gestures are decreased (i.e. narrowed), the blending strength of /g/ is increased, and the blending strength of /a/ is decreased.

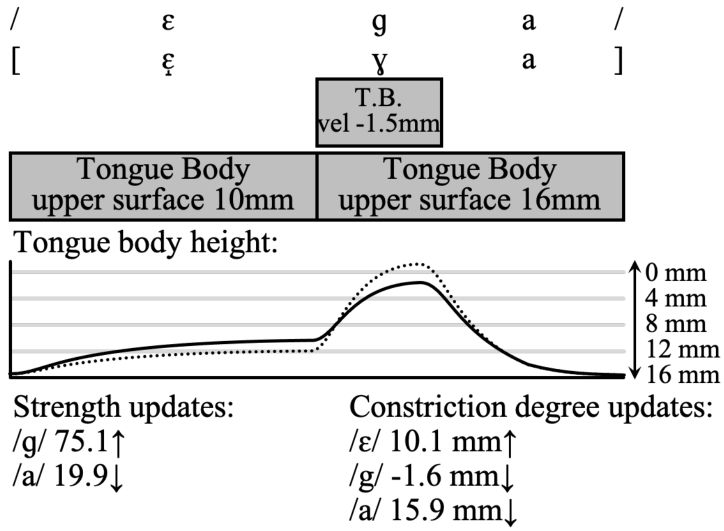


Figure (15): Sample learning trial comparing learner's production (solid) and teacher's production (dashed) of underlying /ɛga/

Using the Gestural Gradual Learning Algorithm, we ran 100 models tasked with learning chain-shifting height harmony and 100 tasked with learning saltatory height harmony. All models were trained until convergence, which occurred when all (V)CV sequences were produced without errors (i.e. with every segment's constriction degree was produced within 0.2 mm of the teacher's production). Results of our learning simulations are reported in the following section.

4.2. Results and Discussion

All of the two hundred models we trained (one hundred per target height harmony pattern) converged upon the target pattern presented to the learner during training. In addition, these models acquired values for gestural blending strength and target tongue body constriction degree that are comparable to our hand-picked values provided in section 3. The mean gestural parameter values acquired by our models at the end of training are provided in Table (4). Note that the overpowering relations between gestures parallel those in our analysis in section 3. In chain-shifting height harmony, triggering /i/ and /u/ overpower fully raising /e/ and /o/, match the strength of partially raising /ɛ/ and /ɔ/, and are doubled in strength by slightly raising /a/. In saltatory height harmony, triggering /i/ and /u/ are overpowered by fully resistant /e/ and /o/ but overpower fully raising /ɛ/ and /ɔ/ and double the strength of partially raising /a/. In both harmony types, the dorsal consonant /g/ resists lenition by overpowering the strongest vowels in the inventory: /a/ in the case of chain-shifting harmony, and /e/ and /o/ in the case of saltatory harmony.

Segment	Chain-Shifting Harmony Blending Strength	Chain-Shifting Harmony Constriction Degree	Saltatory Harmony Blending Strength	Saltatory Harmony Constriction Degree
/i/	11.44	3.84	26.39	3.90
/u/	11.49	3.84	26.39	3.90
/e/	1.02	7.80	343.47	8.10
/o/	1.03	7.80	343.47	8.10
/ɛ/	11.14	12.10	1.02	11.80
/ɔ/	11.14	12.10	1.02	11.80
/a/	22.20	16.10	12.76	16.08
/g/	379.64	-2.00	3125.85	-2.00

Table (4): Mean gestural parameter settings learned for chain-shifting and saltatory vowel height harmony

In order to compare the relative learnability of chain-shifting versus saltatory height harmony when analyzed in the Gestural Harmony Model, we also examined the average number of training iterations necessary for the learners to converge upon each target pattern. The result of this comparison is shown in Figure (16).

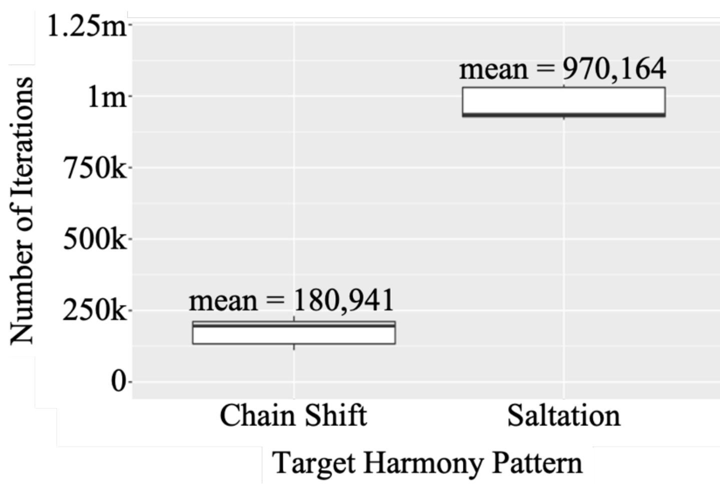
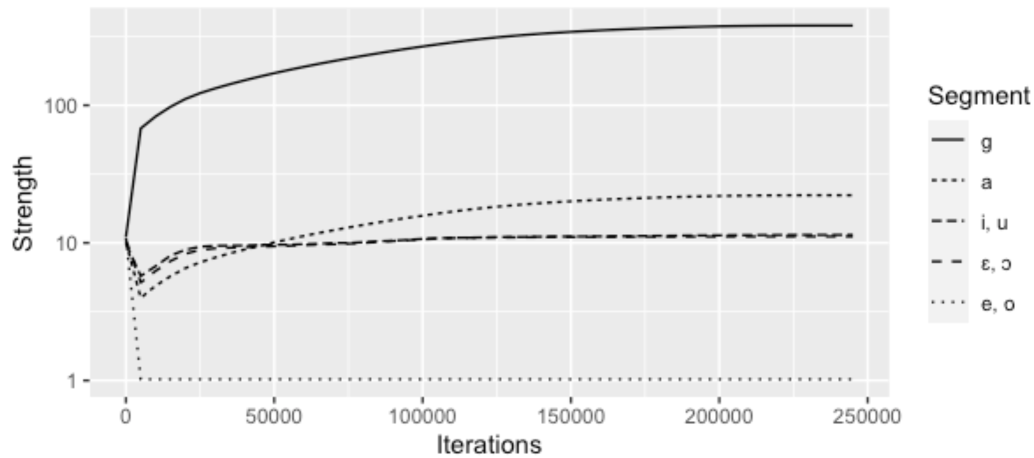


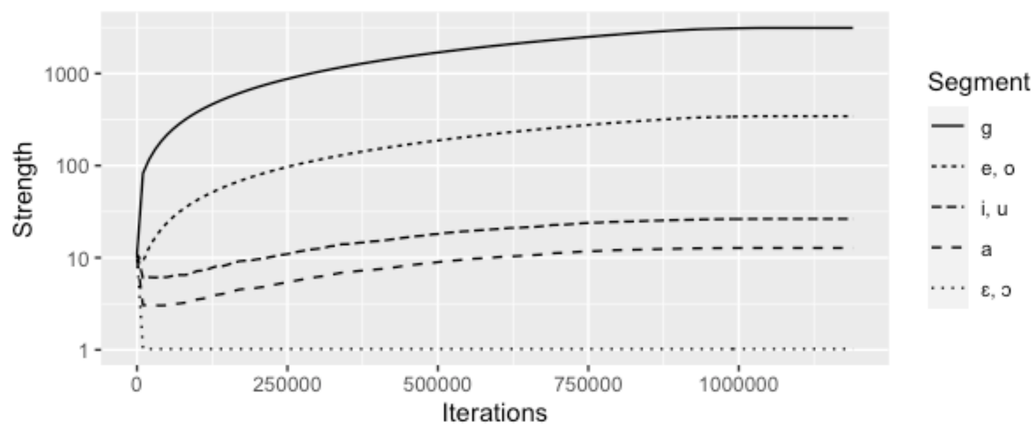
Figure (16): Mean number of training iterations required to learn chain-shifting and saltatory height harmony using the Gestural Gradual Learning Algorithm

We observe that chain-shifting height harmony was learned more than five times faster than saltatory harmony. We interpret this result as an explanation for the attestation of chain-shifting height harmony and the lack of attestation of saltatory height harmony: the acquisition of saltatory height harmony requires much more data, suggesting that it is a more difficult pattern to learn.

The reason why saltatory harmony is more time-consuming and therefore harder to learn lies in the more extreme strength values necessary for the Gestural Harmony Model to generate it. Figure (17) illustrates the average trajectories of segments' gestural strengths throughout training on the two target harmony patterns; (a) depicts trajectories throughout the learning of chain-shifting, while (b) depicts trajectories throughout the learning of saltation. Note the large differences in scale on both the X and Y axes. In both plots, gestural strength is plotted on a logarithmic scale.



(a)



(b)

Figure (17): (a) Trajectory of average learned gestural blending strengths during learning of chain-shifting height harmony; (b) trajectory of average learned gestural blending strengths during learning of saltatory height harmony

These plots show clearly that the blending strength settings for all segments in the learner's inventory take a much longer time to acquire for the saltatory height harmony than for the chain-shifting height harmony. We believe two major factors to be at play here. First, the more extreme strengths necessary to generate saltatory harmony, due to the number of overpowering relations that must exist between different segments in the inventory, require more correct updates to each gesture's blending strength parameter to reach their correct values from their initial values. Recall that all segments are initialized with a gestural blending strength between 1 and 20. With the Gestural Gradual Learning Algorithm's linear parameter update rule, it will simply take more updates for the learner to reach a blending strength value of ~3,125 for /ɡ/ in the saltatory height harmony pattern than a value of ~380 for /ɡ/ in the chain-shifting height harmony pattern.

Note, however, that even less extreme blending strength values take longer to acquire for saltatory harmony learners than for chain-shifting harmony learners. Learners of chain-shifting harmony converge upon values of ~11.5 for the high vowels /i/ and /u/, and they do so on average within 180,000 training iterations. By comparison, learners of saltatory harmony converge upon a similar value of ~12.5 for the low vowel /a/, but take much longer to do so. This is because learning strength ratios involving more extreme strength values involves more incorrect updates to gestures' blending strength parameters, ultimately necessitating more total updates.

In our example learning trial depicted in Figure (14), for instance, the learner blends an /a/ of strength 10 with an /i/ of strength 2, resulting in a degree of raising of /a/ that is not sufficient to match the teacher's production of one-step raising. As a result, the learner updates the strengths of these vowels, making /i/ stronger and /a/ weaker. However, this is an incorrect update for /a/; ultimately, the learner must actually increase the strength of /a/ to ~22 in order to correctly produce the teacher's chain-shifting harmony pattern. It is the too-low strength of /i/ at this point in training that has led the learner to reduce the strength of /a/ after this trial. It is only after training has sufficiently increased the strength of /i/ (in the case of chain-shifting harmony, to twice the strength of /a/) that training items involving the blending of /i/ and /a/ will no longer incorrectly decrease the strength of /a/.

Essentially, gestures that require greater strength values in order to produce a target harmony pattern slow down the acquisition of *all* lesser gestural strength settings. Because generating saltatory height harmony requires much more extreme strength values than chain-shifting harmony, it takes much longer for the strongest gestures to acquire their necessary strength values, leading to more trials with incorrect updates of lower-strength gestures along the way. The result is the slower acquisition of strength settings for all segments in the inventory.

We claim that the difference in learning rate between these two types of height harmony when modeled using the Gestural Gradual Learning Algorithm predicts that the saltatory height harmony pattern should be more likely to be mislearned across generations, thereby becoming less typologically frequent. This learning bias against saltatory height harmony provides an explanation for why such a pattern is not attested, while the relative ease of learning chain-shifting height harmony explains its robust attestation. Therefore, while the Gestural Harmony Model generates both chain-shifting and saltatory height harmony, the relative learnability of these patterns correctly predicts typology.

5. Learning Opaque Height Harmony in Alternative Frameworks

As discussed in section 1, output-oriented, constraint-based frameworks such as Optimality Theory and Harmonic Grammar are unable to generate either chain shifts or saltations using the standard faithfulness constraints assumed within Correspondence Theory (McCarthy & Prince 1995). However, by adopting alternative definitions of faithfulness constraints, it is possible to derive such patterns. In this section, we illustrate these approaches to analyzing the same chain-shifting and saltatory height harmonies modeled in section 4. We also show the results of several sets of learning simulations and compare them to the results of our learning simulations using the Gestural Gradual Learning Algorithm. We show that while the Gestural Gradual Learning Algorithm predicts a learning bias in favor of chain-shifting height harmony and against saltatory height harmony, learning models based within featural phonology show no such bias.

5.1. Chain Shifts and Saltations in Constraint-Based Grammars

Kirchner (1996) characterizes the chain-shifting height harmony of Nzebi as a pattern in which an output is penalized only when a vowel changes two or more of its height features. For instance, underlying /ɛ/ may surface as [e], violating faithfulness to its [±ATR] specification, but not as [i], violating faithfulness to both its [±ATR] and [±high] specifications. A common method of modeling this sort of constraint cumulativity is via constraint ganging in Harmonic Grammar. In constraint ganging, one candidate's violation of two lower-weighted constraints results in a greater penalty than another candidate's violation of a single higher-weighted constraint, allowing the two low-weighted constraints to 'gang up' on and overrule the higher-weighted constraint. However, Albright et al. (2008) and Farris-Trimble (2008) show that the ganging of IDENT constraints in Harmonic Grammar is unable to produce chain shifts.

Łubowicz (2002) and McCarthy (2003) characterize a phonologically-derived environment effect, a class of patterns that includes saltation, as one in which an output is penalized only when it contains a structure that is both unfaithful and marked. Outputs may either surface as faithful with respect to some feature, or having fully resolved violation of a relevant markedness constraint. In our hypothetical saltatory two-step height harmony, underlying /e/ may surface faithfully as [e], despite not fully harmonizing with a high trigger vowel, while underlying /ɛ/ may not surface as [e]; it is both unfaithful to the input for [±ATR] and fails to fully harmonize with a high trigger. However, saltation cannot be generated in Harmonic Grammar via the ganging of a markedness and a faithfulness constraint, as observed by White (2013), Hayes & White (2015), and Smith (to appear).

Constraint ganging in Harmonic Grammar, then, is unable to generate either chain shifts or saltations, both examples of underapplication derivational opacity. However, this does not mean that Harmonic Grammar is wholly unable to generate such patterns. Tesar (2013) and Magri (2018a, 2018b) show that whether a constraint-based phonological grammar can produce derivationally opaque chain shifts and saltations depends not on constraint interaction but rather on properties of the grammar's individual faithfulness constraints.

Constraints from the IDENT family as defined in Correspondence Theory assign violations linearly: each feature change between input and output incurs one IDENT violation. Therefore, any input-output mapping involving multiple feature changes incurs all of the violations incurred by more faithful mappings involving a single feature change. Tesar (2013) and Magri (2018b, 2018a) show that to generate a chain shift, what is needed is a faithfulness constraint that assigns more violations to a more unfaithful mapping than it does to two somewhat unfaithful mappings combined. In the case of vowel raising harmony, for instance, there must be a faithfulness constraint that penalizes two-step raising of /ɛ/ → [i] more greatly than the one-step mappings of /ɛ/ → [e] and /e/ → [i] combined. The inequality in (9), the complement of Magri's Faithfulness Idempotency Condition, provides the violation profile necessary for a faithfulness constraint $\mathbb{C}$ to be able to generate chain-shifting height harmony, where $\mathbb{C}$(mapping) indicates the number of violations that $\mathbb{C}$ assigns to that mapping.

(9) Constraint violation profile necessary to generate chain-shifting vowel raising

$$\mathbb{C}(/\varepsilon/\rightarrow[\mathrm{i}]) > \mathbb{C}(/\varepsilon/\rightarrow[\mathrm{e}]) + \mathbb{C}(/\mathrm{e}/\rightarrow[\mathrm{i}])$$

Conversely, in saltatory height harmony, input-output mappings involving two-step raising such as /ɛ/ → [i] surface while mappings involving one-step raising such as /ɛ/ → [e] or /e/ → [i] do not. Therefore, there must be a faithfulness constraint that assigns fewer violations to the less faithful mapping than to both of the more faithful mappings combined. The inequality in (10), the complement of Magri's Faithfulness Output-Drivenness Condition, provides the violation profile necessary for a faithfulness constraint $\mathbb{S}$ to be able to generate a saltation by which /ɛ/ → [i] while /e/ → [e].

(10) Constraint violation profile necessary to generate saltatory vowel raising

$$\mathbb{S}(/\varepsilon/\rightarrow[\text{i}]) < \mathbb{S}(/\varepsilon/\rightarrow[\text{e}]) + \mathbb{S}(/\text{e}/\rightarrow[\text{i}])$$

Tesar (2013) and Magri (2018a, 2018b) show that the typical versions of faithfulness constraints IDENT, MAX, and DEP assumed within Correspondence Theory do not meet the violation profile conditions in (9) and (10) above under most circumstances, echoing an earlier proposal by Moreton (2004). As a result, they are both *output-driven* and *idempotent* (i.e., incapable of generating either chain shifts or saltations). However, other types of faithfulness constraints have been proposed that do meet these conditions and therefore can generate such derivationally opaque patterns; these are the subjects of sections 5.2 and 5.3.

5.2. Scalar and Categorical Faithfulness

One method of deriving both chain-shifting and saltatory phonological patterns in a constraint-based grammar is to adopt feature *scales*, as proposed by Gnanadesikan (1997). Under this approach, specific feature values are represented by positions along these scales. For instance, the features for voicing, nasality, and sonority are reconceptualized as the Inherent Voicing scale, with values Voiceless Obstruent = 1, Voiced Obstruent = 2, and Sonorant = 3. Gnanadesikan also proposed a ternary Vowel Height scale, though she acknowledges that this scale may need to admit four or more values in order to sufficiently represent languages with greater numbers of vowel heights. For Nzebi's four-height vowel inventory, we expand Gnanadesikan's Vowel Height scale such that it comprises the values High = 1, High-Mid = 2, Low-Mid = 3, and Low = 4. Chain-shifting height harmony in this framework is modeled as a one-step shift along this Vowel Height scale, while saltatory height harmony is modeled as a two-step shift along the scale.

Crucial to this framework is the use of both scalar and categorical markedness and faithfulness constraints. Gnanadesikan (1997) proposes that faithfulness constraints from the IDENT family come in two versions: one, IDENT(X), that penalizes any changes in a segment's feature specification, and another, IDENT-ADJ(ACENT)(X) that penalizes changes of more than one step along a feature scale. These IDENT constraints are defined in (11).

(11) IDENT constraints in Ternary Feature Scales theory (Gnanadesikan 1997, p. 78)

a. IDENT(X): Given an input segment A and its correspondent output segment B, then A and B have identical values on the scale X.

b. IDENT-ADJACENT(X): Given an input segment A and its correspondent output segment B, then A and B must have related values on scale X, where the defined relations are identity and adjacency.

Because we have expanded the Vowel Height scale to include four values rather than three, we also assume a third constraint, IDENT-PART(IAL)(X), that penalizes changes of more than two steps along a feature scale. These faithfulness constraints are in a stringency relationship: IDENT(X) is stricter than IDENT-ADJ(X), which is stricter than IDENT-PART(X). As such, they are not equivalent to the IDENT constraints of Correspondence Theory, which collectively assign one violation per feature change between input and output and generate only derivationally transparent patterns. The IDENT constraints in (11), made possible by the adoption of features scales, are able to generate derivationally opaque phonological patterns. The high weighting of IDENT-ADJ(Height) is responsible for chain-shifting height harmony, while the high weighting of IDENT(Height) is largely responsible for saltatory height harmony.[6]

We first demonstrate how IDENT-ADJ(Height) makes it possible to generate chain-shifting height harmony. This constraint is satisfied not only by candidates that realize an input vowel's height faithfully in the output, but also by those candidates exhibiting single-step vowel raising; only more drastic shifts along the Vowel Height scale incur a violation. IDENT-ADJ(Height) therefore conforms to the violation profile in (9) that is necessary for a faithfulness constraint to be able to generate a chain shift, as shown in the inequality in (12).

(12) IDENT-ADJ(Height) is capable of deriving chain-shifting height harmony

$$\underset{1}{\text{IDENT-ADJ(Ht.)}(/\varepsilon/\to[\text{i}])} > \underset{0}{\text{IDENT-ADJ(Ht.)}(/\varepsilon/\to[\text{e}])} + \underset{0}{\text{IDENT-ADJ(Ht.)}(/\text{e}/\to[\text{i}])}$$

The ability of IDENT-ADJ(Height) to generate a chain-shifting height harmony pattern is illustrated by the tableau in (13). In Gnanadesikan's (1997) framework, harmony is driven by constraints from the ASSIM(ILATE) family, which are defined similarly to constraints from the AGREE family (Lombardi 1999; Baković 2000). Like IDENT, Gnanadesikan proposes that the constraint ASSIM comes in multiple versions: ASSIM(ILATE)(X) requires total agreement between two segments with respect to some feature, while ASSIM(ILATE)-ADJ(ACENT)(X) requires only that two segments have feature scale values that are at least adjacent on the feature scale. Again, we assume a third constraint, ASSIM(ILATE)-PART(IAL)(X), which requires that two segments have feature scale values that are no more than two positions away from one another on our quaternary scale of vowel height. In (13), the constraints ASSIM(Height), ASSIM-ADJ(Height), and ASSIM-PART(Height) are all assigned a weight higher than that of IDENT(Height) in order to capture the fact that some amount of raising is preferable to full faithfulness to underlying vowel height. The constraint IDENT-ADJ(Height), on the other hand, is assigned the highest weight in order to prevent feature scale changes of more than one.

[6] For two-step raising harmony in a four-height vowel system, IDENT-PART(Height) is also necessary in order to prevent three-step mapping of /a/ → [i]; however, we will not discuss this portion of the process here.

(13) IDENT-ADJ(Height) penalizes two- and three-step raising, not one-step raising

Input: /a-i/	IDENT-ADJ(Ht.) w=4	ASSIM-PART(Ht.) w=2	ASSIM-ADJ(Ht.) w=2	ASSIM(Ht.) w=2	IDENT(Ht.) w=1	IDENT-PART(Ht.) w=1	$\mathcal{H}$
a. [a-i]		-1	-1	-1			-6
☞ b. [ɛ-i]			-1	-1	-1		-5
c. [e-i]	-1			-1	-1		-7
d. [i-i]	-1				-1	-1	-6
Input: /ɛ-i/	IDENT-ADJ(Ht.) w=4	ASSIM-PART(Ht.) w=2	ASSIM-ADJ(Ht.) w=2	ASSIM(Ht.) w=2	IDENT(Ht.) w=1	IDENT-PART(Ht.) w=1	$\mathcal{H}$
e. [ɛ-i]			-1	-1			-4
☞ f. [e-i]				-1	-1		-3
g. [i-i]	-1				-1		-5
Input: /e-i/	IDENT-ADJ(Ht.) w=3	ASSIM-PART(Ht.) w=2	ASSIM-ADJ(Ht.) w=2	ASSIM(Ht.) w=2	IDENT(Ht.) w=1	IDENT-PART(Ht.) w=1	$\mathcal{H}$
h. [e-i]				-1			-2
☞ i. [i-i]					-1		-1

For input /a-i/ in (13), candidate (a) [a-i] is faithful to the input, with the first vowel having not undergone any raising. As a result, it violates all of the ASSIM constraints. Winning candidate (b) [ɛ-i] undergoes one-step raising, eliminating a violation of ASSIM-PART. Because this one-step raising is penalized only by low-weighted IDENT(Height) and not by high-weighted IDENT-ADJ(Height), this candidate is chosen as the winner. Candidates (c) [e-i] and (d) [i-i] both involve multi-step raising, incurring fewer violations of the ASSIM constraints but also fatal violations of IDENT-ADJ(Height). These constraints work similarly for inputs /ɛ-i/ and /e-i/. For input /ɛ-i/, faithful candidate (e) [ɛ-i] and one-step raising candidate (f) [e-i] both satisfy high-weighted IDENT-ADJ(Height); however, candidate (f) [e-i] eliminates a violation of ASSIM-ADJ(Height) and is therefore selected as the winner. Candidate (g) [i-i] undergoes full raising, satisfying all ASSIM constraints at the expense of fatally violating IDENT-ADJ(Height). For input /e-i/, neither candidate violates IDENT-ADJ(Height). As a result, it falls to the ASSIM constraints to determine the winner. While faithful candidate (h) [e-i] violates ASSIM(Height), in candidate (i) [i-i] both vowels harmonize fully for height and this candidate is chosen as the winner.

Feature scale theory also makes it possible to generate saltatory height harmony. Rather than individual faithfulness constraints penalizing changes to each of the vowel height features [±high], [±low], and [±ATR], the single constraint IDENT(Height) crucially penalizes a change of any magnitude along the Vowel Height scale equally. This means that a mapping such as /ɛ/ → [i], involving two-step raising, incurs the same penalty from IDENT(Height) as one-step raising mappings /ɛ/ → [e] and /e/ → [i]. Its violation profile therefore fits the violation profile inequality necessary to generate saltation in (10) above, as shown in (14).

(14) IDENT(Height) is capable of deriving saltatory height harmony

$$\underset{1}{\text{IDENT(Height)}(/\varepsilon/\rightarrow[\text{i}])} < \underset{1}{\text{IDENT(Height)}(/\varepsilon/\rightarrow[\text{e}])} + \underset{1}{\text{IDENT(Height)}(/\text{e}/\rightarrow[\text{i}])}$$

With IDENT(Height) weighted high, the more unfaithful mapping /ɛ/ → [i] incurs no more violation than the somewhat unfaithful mappings /ɛ/ → [e] and /e/ → [i]. However, the /ɛ/ → [i] mapping resolves all violations of the harmony-driving ASSIM constraints. Therefore, the more drastic /ɛ/ → [i] mapping is favored, as shown by the tableau in (15). In order to focus on the portion of the two-step raising process that results in saltation, we do not include an account of the two-step raising of /a/ → [ɛ] here.

(15) IDENT(Height) does not favor one-step raising over two-step raising

Input: /ɛ-i/	IDENT(Height) w=3	ASSIM-ADJ(Height) w=3	ASSIM(Height) w=2	IDENT-ADJ(Height) 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(Height) w=3	ASSIM-ADJ(Height) w=3	ASSIM(Height) w=2	IDENT-ADJ(Height) w=1	$\mathcal{H}$
d. [i-i]	-1				-3
☞ e. [e-i]			-1		-2

In (15) with input /ɛ-i/, IDENT(Height) is violated equally by the two-step raising candidate (a) [i-i] and the one-step raising candidate (b) [e-i]. Therefore, whether /ɛ/ raises fully to [i] or only to [e] is dependent on the relative weights of ASSIM(Height) and IDENT-ADJ(Height). Because ASSIM(Height) is weighted higher, candidate (a) [i-i] is the winner. The faithful candidate (c) [ɛ-i] satisfies both IDENT constraints at the expense of violating both ASSIM constraints. With input /e-i/, no amount of raising is favored, as candidate (d) [i-i] violates high-weighted IDENT(Height) while faithful candidate (e) [e-i] violates only lower-weighted ASSIM(Height).

In this section, we have shown that the scalar and categorical faithfulness constraints of Feature Scales theory are able to generate both chain-shifting and saltatory harmony. Scalar faithfulness constraints generate chain shifts by assigning violations based on the magnitude of a feature change along a scale and not based on changes to individual binary features. Categorical faithfulness constraints generate saltations by assigning equal violations regardless of the magnitude of a feature change along a scale.

5.3. Distinct Faithfulness

Another approach to generating derivationally opaque phonological patterns involves the use of distinct faithfulness constraints for each input-output mapping using constraints from the *MAP family (Zuraw 2007). These *MAP constraints penalize specific input-output mappings rather than featural mismatches between input-output correspondents. White (2013) provides the definition in (16) for this family of constraints.

(16) *MAP constraint definition (White 2013, p. 44)

*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.

For example, the constraint *MAP([+low],[+high]) assigns a single violation to a candidate in which an underlying /a/ corresponds with a surface [i] or [u], and the constraint *MAP([+ATR,-high],[+high]) assigns a single violation to an underlying /e/ or /o/ that surfaces as [i] or [u]. Because there is assumed to be a distinct *MAP constraint for every input-output mapping, there is a *MAP constraint that penalizes a mapping involving two or more feature changes, but not one involving a single feature change. This constraint can generate a chain shift. Conversely, there is also a *MAP constraint that penalizes a mapping involving a single feature change, but not one involving two or more feature changes; this constraint can generate a saltation.

In the case of height harmony in a language with a four-height vowel inventory, we assume the *MAP constraints in (17). We assume a *MAP constraint penalizing input-output mappings between each vowel height, using the following abbreviations for conciseness: /a/ for [+low], /ɛ/ for [-ATR, -low], /e/ for [+ATR, -high], and /i/ for [+high].

(17) *MAP constraints penalizing changes in vowel height

a. *MAP(e,i): Assign a violation when a [+ATR, -high] segment is in correspondence with a [+high] segment.

b. *MAP(ɛ,e): Assign a violation when a [-ATR, -low] segment is in correspondence with a [+ATR, -high] segment.

c. *MAP(a,ɛ): Assign a violation when a [+low] segment is in correspondence with a [-ATR, -low] segment.

d. *MAP(ɛ,i): Assign a violation when a [-ATR, -low] segment is in correspondence with a [+high] segment.

e. *MAP(a,e): Assign a violation when a [+low] segment is in correspondence with a [+ATR, -high] segment.

f. *MAP(a,i): Assign a violation when a [+low] segment is in correspondence with a [+high] segment.

Recall that in order to produce a chain shift in which /ɛ/ → [e] and /e/ → [i], there must be a constraint that assigns a greater number of violations to the mapping /ɛ/ → [i] than it does to the mappings /ɛ/ → [e] and /e/ → [i] combined, following (9) above. The constraint *MAP(ɛ→i) fits this violation profile condition, as shown in (18): it assigns one violation to the two-step raising mapping /ɛ/ → [i], and no violations to any other mapping.

(18) *MAP(ɛ,i) is capable of deriving chain-shifting height harmony

$$\underset{1}{*\text{MAP}(\varepsilon,i)(/\varepsilon/\rightarrow[i])} > \underset{0}{*\text{MAP}(\varepsilon,i)(/\varepsilon/\rightarrow[e])} + \underset{0}{*\text{MAP}(\varepsilon,i)(/e/\rightarrow[i])}$$

The ability of this constraint to generate chain-shifting height harmony is demonstrated by the tableau in (19). To drive harmony, we utilize constraints we simply call Harmony(high), Harmony(ATR), and Harmony(low); whether these constraints motivate harmony via feature agreement or feature spreading is unimportant. Several crucial constraint weightings are at play in the generation of chain-shifting height harmony. First, to drive one-step raising, each of the Harmony constraints must outweigh one of the set of *Map constraints penalizing one-step vowel raising: Harmony(high) must outweigh *Map(e,i), Harmony(ATR) must outweigh *Map(ɛ,e), and Harmony(low) must outweigh *Map(a,ɛ). In addition, multistep vowel raising is prevented by assigning high weights to all of the *Map constraints that penalize two- or three-step raising (*Map(a,e), *Map(a,i), and *Map(ɛ,i)). To prevent two-step raising of /ɛ/ → [i] and /ɔ/ → [u], *Map(ɛ,i) must be provided a weight higher than the combined weights of harmony-driving Harmony(high) and one-step raising-penalizing *Map(ɛ,e). Similarly, to prevent two-step raising of /a/ → [e], *Map(a,e) must have a weight higher than the combined weights of Harmony(ATR) and *Map(a,ɛ). Finally, to prevent full raising of /a/ → [i], the constraint *Map(a,i) must have a greater weight than the combined weights of three lower-weighted constraints: harmony-driving Harmonize(high) and Harmonize(ATR) and two-step raising-penalizing *Map(a,ɛ).

(19) *Map(a,i), *Map(a,e), and *Map(ɛ,i) penalize multi-step vowel 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	*Map (e,i) w=1	*Map (ɛ,e) w=1	*Map (a,ɛ) w=1	$\mathcal{H}$
a. [i-i]	-1									-6
b. [e-i]		-1		-1						-6
☞ c. [ɛ-i]				-1	-1				-1	-5
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	*Map (e,i) w=1	*Map (ɛ,e) w=1	*Map (a,ɛ) w=1	$\mathcal{H}$
e. [i-i]			-1							-4
☞ f. [e-i]				-1				-1		-3
g. [ɛ-i]				-1	-1					-4
Input: /e-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	*Map (e,i) w=1	*Map (ɛ,e) w=1	*Map (a,ɛ) w=1	$\mathcal{H}$
☞ h. [i-i]							-1			-1
i. [e-i]				-1						-2

In (19) with input /a-i/, multistep raising candidates (a) [i-i] and (b) [e-i] are each ruled out due to violation of a high-weighted *Map constraint. Winning candidate (c) [ɛ-i] exhibits one-step raising, violating only low-weighted *Map(a,ɛ) in order to satisfy Harmony(low). Finally, faithful candidate (d) [a-i] satisfies all of the *MAP constraints, but violates all of the HARMONY constraints. With input /ɛ-i/, two-step raising is again prevented due to the high weighting of *MAP(ɛ,i), which is violated by candidate (e) [i-i]. Winning candidate (f) [e-i] violates low-weighted *MAP(ɛ,e) in order to satisfy HARMONY(ATR), while faithful candidate (g) [ɛ-i] violates HARMONY(high) and HARMONY(ATR). With input /e-i/, candidate (h) [i-i] is able to satisfy all HARMONY constraints with only one-step raising, violating only low-weighted *MAP(e,i), while faithful candidate (i) [e-i] violates HARMONY(high).

There is also a *MAP constraint with the violation profile necessary to generate two-step, saltatory height harmony. In order to produce the saltatory pattern in which /ɛ/ → [i] and /e/ → [e], there must be a constraint that assigns fewer violations to the more unfaithful mapping /ɛ/ → [i] than to the mappings /ɛ/ → [e] and /e/ → [i] combined. The constraint *MAP(e,i) fits this violation profile, as shown in (20).

(20) *MAP(e,i) is capable of deriving saltatory height harmony

$$\underset{0}{\text{*MAP(e,i)(/ɛ/→[i])}} < \underset{0}{\text{*MAP(e,i)(/ɛ/→[e])}} + \underset{1}{\text{*MAP(e,i)(/e/→[i])}}$$

We illustrate this constraint's ability to generate saltatory height harmony in the tableau in (21). In order to enforce the faithful mapping of /e/ → [e], the constraint *MAP(e,i), which penalizes one-step raising of high-mid vowels to high, must be provided a higher weight than the harmony-driving markedness constraint HARMONY(high). In order to drive the saltatory two-step raising of /ɛ/ → [i] and /ɔ/ → [u], the constraint *MAP(ɛ,i) must be provided a weight lower than the combined weights of HARMONY(high) and HARMONY(ATR). There is no crucial weighting of one-step raising-penalizing *MAP(ɛ,e); therefore, we provide it with a low weight here. Again, in order to focus only on the portion of the two-step raising process that results in saltation, we do not include an account of the two-step raising of /a/ → [e].

(21) *MAP(e,i) penalizes one-step raising, not two-step raising

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**	$\mathcal{H}$
d. [i-i]	-1					-2
☞ e. [e-i]		-1				-1

In (21) with input /ɛ-i/, winning candidate (a) [i-i] exhibits full raising, satisfies both HARMONY constraints while violating only low-weighted *MAP(ɛ,i). The more-faithful candidate (b) [e-i], meanwhile, violates HARMONY(high) as well as *MAP(ɛ,e). The fully-faithful candidate (c) [ɛ-i] violates no *MAP constraints but both HARMONY constraints. By contrast, with input /e-i/ the full raising candidate (d) [i-i] is not selected as the winner due to its violation of high-weighted *MAP(e,i). Winning candidate (e) [e-i] satisfies this constraint and violates only lower-weighted HARMONY(high).

Both the scalar and categorical faithfulness approach and the distinct faithfulness approach, then, are capable of deriving both chain-shifting and saltatory height harmony. In the following section, we test the learnability of chain-shifting and saltatory height harmony within frameworks assuming either scalar features or distinct faithfulness constraints and compare the results of those learning simulations to those of the Gestural Harmony Model.

5.4. Learning Simulations

5.4.1. Setup

To determine the relative learnability of chain-shifting and saltatory height harmony in the frameworks described in sections 5.2 and 5.3, we used the MaxEnt Generational Stability Model (O'Hara 2021). This model is an iterated agent-based learning model based on those developed by Kirby & Hurford (2002), Griffiths & Kalish (2007), Culbertson & Kirby (2016), and others. The use of such models has been extended to phonological acquisition by Pater & Moreton (2012), Staubs (2014), Hughto (2020), and others.

The MaxEnt Generational Stability Model simulates the learning and transmission of phonological patterns across multiple generations of learners tasked with acquiring a Maximum Entropy (henceforth MaxEnt) Harmonic Grammar (Goldwater & Johnson 2003; Jäger 2007) using the Perceptron learning algorithm (Rosenblatt 1958; Boersma & Pater 2016). In one generation, one model agent, the learner, is exposed to a large, but limited, number of forms produced by the grammar of another model agent, the teacher. For each learning trial, an input is randomly sampled, and the learner then samples an output form for that input given its grammar's current constraint weights. If the learner makes an error, i.e. selects an output candidate different from the one selected by the teacher, the learner's grammar is updated, increasing the weights of constraints violated by the learner and decreasing the weights of constraints violated by the teacher. In the next generation, the learner becomes the teacher, and its grammar produces forms to which the next generation's learner is exposed. This process is repeated for a number of generations.

In this learning model, patterns that are learned more quickly and accurately are more likely to remain stable across generations, while those that are learned less quickly and accurately are more likely to change. At the end of the last generation, a learning simulation is classified according to the pattern produced by the grammar learned by the final generation.[7] The *stability* of a pattern can be quantified by examining the percentage of simulations in which the last generation's grammar is classified as the same type of pattern as the original teacher's pattern. The stability of a pattern across generations serves as a metric for the learnability of that pattern, with more stable patterns being learned more quickly and easily (O'Hara 2021).

Using the MaxEnt Generational Stability Model, we ran two sets of simulations of the learning of both chain-shifting and saltatory height harmony: one using the scalar and categorical

[7] Because MaxEnt is a probabilistic model of grammar, no learner ever learns a categorical target pattern. Here, we classify a MaxEnt grammar according to the categorical pattern generated by the set of constraints and weights in its non-probabilistic Harmonic Grammar equivalent.

faithfulness constraints of Feature Scales theory, and one using distinct *MAP faithfulness constraints. The full list of constraints used is provided in Table (5).

	Markedness Constraints	Faithfulness Constraints
Scalar/Categorical Faithfulness	ASSIM-IDENT(Height) ASSIM-ADJ(Height) ASSIM-PART(Height)	IDENT(Height) IDENT-ADJ(Height) IDENT-PART(Height)
Distinct Faithfulness	HARMONIZE(high) HARMONIZE(ATR) HARMONIZE(low)	*MAP(e,i) *MAP(ɛ,e) *MAP(a,ɛ) *MAP(ɛ,i) *MAP(a,e) *MAP(a,i)

Table (5): Constraint sets used in MaxEnt Generational Stability Model simulations

In our scalar/categorical faithfulness simulations, each learner was initialized with all harmony-driving markedness constraints (those from the ASSIM family) weighted at 50, and all faithfulness constraints (those from the IDENT family) weighted at 1. This initial weighting of markedness constraints above faithfulness constraints has precedence in the field of phonological acquisition modeling, having been shown to ensure learning of restrictive patterns (Tesar & Smolensky 1998) and to better match stages of child language acquisition (Gnanadesikan 2004; Jesney & Tessier 2011).

The initial constraint weighting conditions for our simulations utilizing *MAP constraints are more involved. White (2013) argues for a substantive bias in the relative weights of *MAP constraints during learning that is consistent with the P-Map (Steriade 2008), a speaker's knowledge of the perceptual similarities between all pairs of segments. Such a bias makes it more difficult to learn saltation and other patterns that favor less faithful mappings over more faithful ones. White implements the P-Map bias using the MaxEnt Grammar Tool[8] through a persistent Gaussian prior on the weights of *MAP constraints based on the P-Map. This prior term leads the MaxEnt Grammar Tool to prefer specific weightings of constraints that are similar to those specified by the user. Here, we make use of gradual learning simulations rather than the MaxEnt Grammar Tool's batch learning simulations, and therefore we bias the constraint weights via initial constraint weightings instead of a persistent prior term. Gradual learning algorithms with non-persistent biases have been argued to better parallel natural language learning and to perform better than batch learners with persistent priors on matching some human behavior (O'Hara 2020).

In our simulations, we approximated the P-Map bias by providing initial weightings to our *MAP constraints as follows. The constraint *MAP(a,i) penalizing changes to three vowel height features was assigned three times the initial weight of those *MAP constraints penalizing changes to only one height feature (*MAP(a,ɛ), *MAP(ɛ,e), and *MAP(e,i)), and those *MAP constraints penalizing changes to two height features (*MAP(a,e) and *MAP(ɛ,i)) were assigned twice the initial weight of the one-step penalizing *MAP constraints. These linear weight ratios were chosen to remove any potential bias favoring either chain-shifts or saltations.

Our distinct faithfulness simulations were also divided into three subcategories based on different initial relative weightings of the markedness and faithfulness constraints. Because

[8] Available at: https://linguistics.ucla.edu/people/hayes/MaxentGrammarTool/

constraints weight hierarchies for *MAP constraints have not been previously implemented in online learning simulations, we ran simulations with multiple initial weightings. In one, harmony-driving markedness constraints were provided weights much greater than all *MAP faithfulness constraints (abbreviated M > F); this is comparable to the initial weighting condition used for our scalar and categorical faithfulness simulations, as well as previous work on phonological acquisition. In another, markedness constraints were provided weights equal to the lowest-weighted *MAP constraints (abbreviated M ≈ F). In the third initial weighting condition, markedness constraints were provided weights much lower than all *MAP constraints (abbreviated M < F). This condition is most similar to the batch learner proposed by White (2013), in which markedness constraints have a prior term biasing their weights toward zero and *MAP constraints have a prior term biasing their weights to be consistent with the P-Map. The exact initial weightings of constraints in each of these conditions are provided in Table (6).

Condition	Markedness Constraints		Faithfulness Constraints	
M > F	HARMONIZE(high)	50	*MAP(e,i)	5
	HARMONIZE(ATR)	50	*MAP(ɛ,e)	5
	HARMONIZE(low)	50	*MAP(a,ɛ)	5
			*MAP(ɛ,i)	10
			*MAP(a,e)	10
			*MAP(a,i)	15
M ≈ F	HARMONIZE(high)	50	*MAP(e,i)	50
	HARMONIZE(ATR)	50	*MAP(ɛ,e)	50
	HARMONIZE(low)	50	*MAP(a,ɛ)	50
			*MAP(ɛ,i)	100
			*MAP(a,e)	100
			*MAP(a,i)	150
M < F	HARMONIZE(high)	1	*MAP(e,i)	17
	HARMONIZE(ATR)	1	*MAP(ɛ,e)	17
	HARMONIZE(low)	1	*MAP(a,ɛ)	17
			*MAP(ɛ,i)	34
			*MAP(a,e)	34
			*MAP(a,i)	51

Table (6): Initial constraint weightings for simulations utilizing distinct faithfulness constraints

In most of our simulations, at each generation a learner was exposed to 2,000 forms produced by its teacher (in all but the first generation, the learner from the previous generation). The only exception was the set of simulations utilizing *MAP constraints and the M > F initial constraint weighting condition. Under this initial condition, both chain-shifting and saltatory height harmony are learned significantly slower. To account for this, for these simulations we increased the number of forms each learner was exposed to per generation from 2,000 to 3,600. All simulations ran for 20 generations with a learning rate constant of 0.05. For each target height harmony pattern (chain-shifting or saltatory), constraint set (scalar/categorical faithfulness or distinct faithfulness), and initial constraint weighting (markedness over faithfulness for scalar/categorical; M > F, M ≈ F, and M < F for distinct) we ran 100 simulations.

5.4.2. Results and Discussion

Our results indicate that across all of our learning simulations, saltatory height harmony is predicted to be either more stable than or as stable as chain-shifting height harmony across generations of phonological acquisition, and therefore at least as typologically frequent as chain-shifting height harmony. This is in direct contrast with both the typological facts and the results of our simulations utilizing the Gestural Gradual Learning Algorithm reported in section 4.2.

The stability of each height harmony pattern in the simulations utilizing the scalar and categorical faithfulness constraints is reported in Table (7).

Target Pattern	Stability Across Generations
Chain Shift	67%
Saltation	93%

Table (7): Rates of stability for chain-shifting and saltatory height harmony using scalar and categorical faithfulness constraints (IDENT, IDENT-ADJACENT, and IDENT-PARTIAL)

With this set of constraints, the saltation pattern is more stable than the chain-shifting pattern, and therefore easier to learn than the chain-shifting pattern. From these results, we would expect that saltatory height harmony would be better represented crosslinguistically than chain-shifting height harmony, contra the typological evidence.

Figure (Figure) shows the results of our simulations utilizing distinct faithfulness constraints. In each of the three initial weighting conditions, the chain shift pattern (shown in white) is less stable than the saltatory pattern (shown in gray).

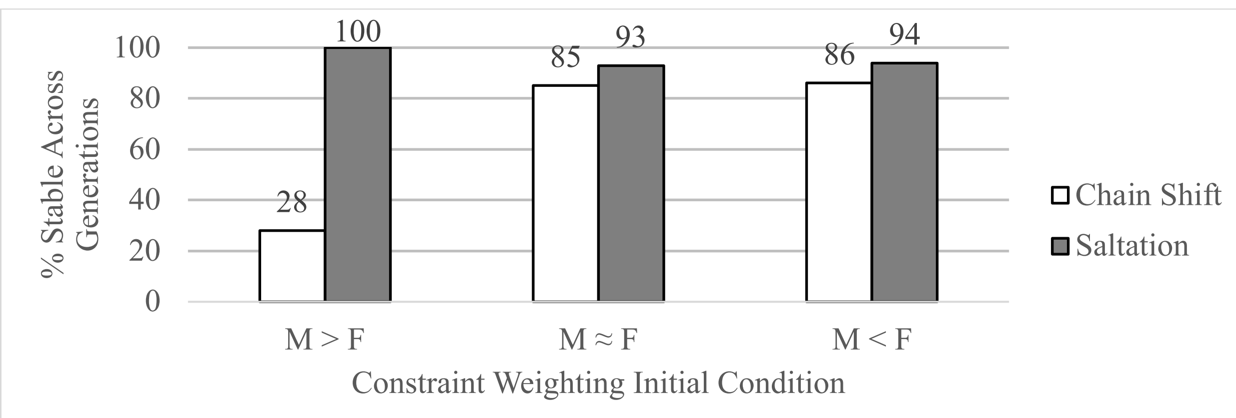


*Figure (8): Rates of stability for chain-shifting and saltatory height harmony using distinct faithfulness constraints (*MAP family) with three different initial constraint weighting conditions*

Together, these three sets of simulations suggest that learners of a grammar that includes faithfulness constraints from the *MAP family are not biased in favor of chain-shifting height harmony over saltatory height harmony. This finding is consistent across a variety of possible initial weightings, suggesting that this effect is not the result of any particular initial weighting condition. In fact, one of these initial weighting conditions appears to substantially favor saltatory over chain-shifting height harmony.

Across both of the examined featural approaches to generating derivationally opaque height harmony (scalar and categorical faithfulness and distinct faithfulness), our simulations showed no learning bias in favor of attested chain-shifting harmony and against unattested saltatory harmony. We argue that this result arises because the saltatory harmony pattern makes greater use of *consistent* constraints. Constraint consistency is a measure of how often across learning trials a constraint is violated by intended winner candidates (as determined by the teacher's grammar), candidates chosen by the learner in error, or a combination of the two.

The consistency of several abstract constraints is illustrated in the sample comparative tableau in (22). Each row compares violations of the intended winning candidate to violations of some other candidate that the learner could erroneously choose as the winner. Each cell shows the difference vector between violations of the intended winning candidate and the learner's erroneous winning candidate. If the difference vector for a constraint is positive, that constraint is *winner-favoring*, indicated by a W. If the difference vector is negative, the constraint is *loser-favoring*, indicated by an L. A zero indicates no difference in number of violations between the candidates.

(22) Sample comparative tableau illustrating constraint consistency

			CON 1	CON 2	CON 3
/input/	a.	[intended winner] ~ [error 1]	0	1W	-1L
	b.	[intended winner] ~ [error 2]	-1L	0	0
	c.	[intended winner] ~ [error 3]	-1L	1W	1W

A consistent constraint is one that is either never winner-favoring or never loser-favoring. In (22), CONSTRAINT 1 is consistent as it is only loser-favoring or neutral, never winner-favoring. CONSTRAINT 2 is also consistent as it is only winner-favoring or neutral, never loser-favoring. However, CONSTRAINT 3 is inconsistent because one comparison between an intended winner and a learner's erroneous winner is winner-favoring, while another is loser-favoring.

Constraint consistency plays an important role in online learning via the Perceptron algorithm (Rosenblatt 1958; Boersma & Pater 2016) for constraint weight updates. The Perceptron algorithm updates a constraint's weight whenever the learner makes an error by comparing the violations of that constraint by the teacher's winning candidate and the learner's erroneous winning candidate. If the constraint is violated more by the learner's erroneous winning candidate (i.e., if it is loser-favoring), the constraint's weight is increased. However, if it is violated more by the teacher's winning candidate (i.e., if it is winner-favoring), its weight is decreased. A maximally consistent constraint's weight will only ever be updated in one direction, while for an inconsistent constraint some errors made by the learner during training will increase the constraint's weight and others will decrease it, leading the weight to oscillate rather than monotonically increase or decrease. These inconsistent constraints ultimately update their weights more slowly than consistent constraints as different training items' updates cancel each other out.

A general finding across work on the learning of constraint-based grammars is that learners more easily learn patterns with high-weighted constraints that are consistent. Staubs (2014) and Stanton (2016) show that stress systems generated by consistent constraints are learned faster and are more typologically frequent. Moreton et al. (2017) show that patterns supported by more consistent constraints (which they call *valid* constraints) are learned more quickly in a MaxEnt framework with a conjunctive constraint schema, and O'Hara (2021) shows that consistency impacts the update speed of constraints in the MaxEnt Generational Stability model and is a major factor in determining which phonological patterns are more or less stable across generations of learners. We claim that it is constraint consistency that leads to saltatory height harmony being more stable than chain-shifting harmony in some of our sets of simulations.

In all of our simulations, the grammar that produces saltatory height harmony includes more consistent markedness constraints compared to the grammar that produces chain-shifting height harmony. We exemplify the difference in consistency using the HARMONY family of constraints used in our *MAP simulations in section 5.4.[9] The consistency of each markedness constraint in our chain-shifting and saltatory height harmony patterns can be seen in the comparative tableaux below. We show only faithful candidates and those involving vowel raising; those candidates involving vowel lowering are harmonically bounded and therefore not seen as errors in training.

The comparative tableau in (23) shows the consistency of the markedness constraints when chain-shifting height harmony is the learner's target pattern. We see that only HARMONY(low) is consistent, as it is winner-favoring when comparing the mapping /a-i/ → [ɛ-i] with the mapping /a-i/ → [a-i], and neither winner- nor loser-favoring elsewhere. HARMONY(ATR) and HARMONY(high), by contrast, are both inconsistent.

(23) HARMONY(ATR) is not consistent in chain-shifting height harmony

			HARMONY(high)	HARMONY(ATR)	HARMONY(low)
I. /a-i/	a.	[ɛ-i] ~ [a-i]	0	0	1W
	b.	[ɛ-i] ~ [e-i]	0	-1L	0
	c.	[ɛ-i] ~ [i-i]	-1L	-1L	0
II. /ɛ-i/	d.	[e-i] ~ [ɛ-i]	0	1W	0
	e.	[e-i] ~ [i-i]	-1L	0	0
III. /e-i/	f.	[i-i] ~ [e-i]	1W	0	0

On the other hand, with saltatory height harmony as the learner's target pattern, HARMONY(ATR) is consistent, as shown in the comparative tableau in (24). None of the teacher's winning candidates violate HARMONY(ATR), so it is always either winner-favoring or neutral. HARMONY(low) is equally consistent for both chain-shifting and saltatory height harmony, while HARMONY(high) is inconsistent for both patterns.

(24) HARMONY(ATR) and HARMONY(low) are consistent in saltatory height harmony

			HARMONY(high)	HARMONY(ATR)	HARMONY(low)
I. /a-i/	a.	[e-i] ~ [a-i]	0	1W	1W
	b.	[e-i] ~ [ɛ-i]	0	1W	0
	c.	[e-i] ~ [i-i]	-1L	0	0
II. /ɛ-i/	d.	[i-i] ~ [ɛ-i]	1W	1W	0
	e.	[i-i] ~ [e-i]	1W	0	0
III. /e-i/	f.	[e-i] ~ [i-i]	-1L	0	0

Taken together, these tableaux provide a potential explanation for the lack of a learning bias against saltatory height harmony and in favor of chain-shifting harmony in feature- and constraint-

[9] For the four-height vowel inventory used in our simulations, the ASSIM family of constraints have the same violation profiles as the HARMONY constraints: ASSIM-IDENT(Height) corresponds to HARMONY(high), ASSIM-ADJ(Height) to HARMONY(ATR), and ASSIM-PARTIAL(Height) to HARMONY(low).

based accounts of these patterns. While two of the three markedness constraints implicated in these patterns are inconsistent for a learner attempting to set the constraint weights necessary to produce chain-shifting harmony, only one of these constraints is inconsistent for a learner of saltatory harmony. The learner of saltatory harmony, then, is able to update its constraint weights more quickly and efficiently, with fewer training items canceling out each other's updates. We can compare this to the results of our simulations using the Gestural Gradual Learning Algorithm reported in section 4.2. In those simulations, it was learners of saltatory harmony who made inconsistent updates to segments' blending strength values, resulting in ultimately slower setting of those values compared to learners of chain-shifting height harmony.

In sum, under both the scalar and categorical faithfulness approach and the distinct faithfulness approach, unattested saltatory height harmony was learned and passed on to future generations of learners with at least as much stability as attested chain-shift harmony in simulations using the MaxEnt Generational Stability Model. These results indicate no presence of a learning bias against saltatory height harmony, and in some cases even indicate a bias in favor of such a pattern. We conclude that these approaches to capturing cases of underapplication opacity in vowel harmony are unable to predict the typological asymmetry by which chain-shifting height harmony is well-attested and saltatory height harmony is not.

6. Conclusion

Derivationally opaque patterns pose a question for phonological theory: how can a phonological model be powerful enough to generate attested derivationally opaque patterns without generating a wide variety of unattested opaque patterns? In this paper, we have shown that in the Gestural Harmony Model, attested chain-shifting height harmony and unattested saltatory height harmony can both be straightforwardly modeled as the result of gestural overlap and blending between a trigger and an undergoer. While both types of opaque harmony patterns can be modeled in the Gestural Harmony Model, we showed via simulations utilizing the Gestural Gradual Learning Algorithm that unattested saltatory height harmony was much harder to learn, providing an explanation for its lack of attestation. We also showed that alternative constraint-based approaches that can generate both of these derivationally opaque patterns cannot turn to learnability to explain the lack of attestation of saltatory height harmony; in all cases, saltation was as easy or easier to learn than a chain shift. We take this as evidence in favor of the Gestural Harmony Model's representation of derivationally opaque vowel harmony patterns.

Looking forward, synchronic chain shifts can be found in many phonological processes outside of height harmony, including consonant lenition, tone sandhi, and segment deletion. Evidence for similar synchronic saltatory patterns has also been found (White 2013; Smith to appear). Future work will examine whether the mechanism of gestural blending utilized here to represent chain-shifting and saltatory height harmonies can be similarly used to model cases of underapplication opacity more generally, and whether these additional cases exhibit learning biases that might inform the typological picture.

References

Albright, Adam, Giorgio Magri, & Jennifer Michaels (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 (2000). *Harmony, Dominance and Control*. Ph.D. Dissertation, Rutgers, The State University of New Jersey.

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, & Joe Pater (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., & Louis Goldstein (1986). Towards an Articulatory Phonology. *Phonology Yearbook*, 3, 219–252.

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

Camilli, Amerindo (1929). Il dialetto di Servigliano. *Archivum Romanicum*, 13, 220–271.

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.

Culbertson, Jennifer, & Simon Kirby (2016). Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in Psychology*, 6.

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

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

Gnanadesikan, Amalia (2004) Markedness and Faithfulness Constraints in Child Phonology. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in Phonological Acquisition* (pp. 73–108). Cambridge University Press.

Gnanadesikan, Amalia (1997). *Phonology with Ternary Scales*. Ph.D. 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).

Griffiths, Thomas L., & Michael L. Kalish (2007). Language Evolution by Iterated Learning With Bayesian Agents. *Cognitive Science*, 31(3), 441–480.

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

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

Hughto, Coral (2020). *Emergent Typological Effects of Agent-Based Learning Models in Maximum Entropy Grammar*. Ph.D. 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.

Jarosz, Gaja (2016). Learning Opaque and Transparent Interactions in Harmonic Serialism. In G. Ó. Hansson, A. Farris-Trimble, K. McMullin, & D. Pulleyblank (Eds.), *Proceedings of the 2015 Annual Meeting on Phonology*. Linguistic Society of America.

Jesney, Karen, & Anne-Michelle Tessier (2011). Biases in Harmonic Grammar: the road to restrictive learning. *Natural Language and Linguistic Theory*, 29, 251–290.

Kaze, Jeffery W. (1989). *Metaphony in Italian and Spanish dialects revisited*. Ph.D. Dissertation, University of Illinois at Urbana-Champaign.

Kirby, Simon, & James R. Hurford (2002). The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model. In A. Cangelosi & D. Parisi (Eds.), *Simulating 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, Yoshiro Miyata, & Paul Smolensky (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.

Lombardi, Linda (1999). Positional Faithfulness and Voicing Assimilation in Optimality Theory. *Natural Language and Linguistic Theory*, 17, 267–302.

Łubowicz, Anna (2002). Derived environment effects in Optimality Theory. *Lingua*, 112, 243–280.

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.

Mascaró, Joan (2011). An analysis of stress-dependent harmony in Servigliano. *Probus*, 23(1), 21–55.

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

McCarthy, John J. (2003). Comparative markedness. *Theoretical Linguistics*, 29, 1–51.

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

McCollum, Adam G., Eric Baković, Anna Mai, & Eric Meinhardt (2020). Unbounded circumambient patterns in segmental phonology. *Phonology*, 37(2), 215–255.

Moreton, Elliott (2004). Non-computable functions in Optimality Theory. In J. J. McCarthy (Ed.), *Optimality Theory in Phonology* (pp. 141–164).

Moreton, Elliott (2008). Analytic bias and phonological typology. *Phonology*, 25, 83–127.

Moreton, Elliott, Joe Pater, & Katya Pertsova (2017). Phonological Concept Learning. *Cognitive Science*, 41, 4–69.

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, Charlie (2020). Frequency Matching Behavior in On-line MaxEnt Learners. In A. Ettinger, G. Jarosz, & M. Nelson (Eds.), *Proceedings of the Society for Computation in Linguistics: Vol. 3* (pp. 461–463).

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

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

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

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

Roca, Iggy (Ed.) (1997). *Derivations and Constraints in Phonology*. Clarendon Press.
Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6), 386–408.
Saltzman, Elliot L., & Kevin G. Munhall (1989). A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology*, 1(4), 333–382.
Schmidt, Deborah S. (1996). Vowel raising in Basaa: A synchronic chain shift. *Phonology*, 13(2), 239–267.
Smith, Caitlin (2018). *Harmony in Gestural Phonology*. Ph.D. Dissertation, University of Southern California.
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, Jennifer L. (to appear). Some Formal Implications of Deletion Saltation. To appear in *Linguistic Inquiry*.
Smolensky, Paul, & Géraldine Legendre (Eds.) (2006). *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. The MIT Press.
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*. Ph.D. Dissertation, University of Massachusetts Amherst.
Steriade, Donca (2008). The Phonology of Perceptibility Effects: The P-Map and its Consequences for Constraint Organization. In K. Hanson & S. Inkelas (Eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky* (pp. 151–179). MIT Press.
Tesar, Bruce (2013). *Output-Driven Phonology: Theory and Learning*. Cambridge University Press.
Tesar, Bruce, & Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry*, 29(2), 229–268.
Walker, Rachel (2011). *Vowel Patterns in Language*. Cambridge University Press.
White, James (2013). *Bias in Phonological Learning: Evidence from Saltation*. Ph.D. Dissertation, University of California Los Angeles.
Wilson, Colin (2006). Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science*, 30(5), 945–982.
Zuraw, Kie (2007). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog infixation. *Language*, 83(2), 277–316.