Phonology **35** (2018). Supplementary materials

Vowel dispersion and Kazakh labial harmony

Adam G. McCollum
University of California, San Diego

Supplementary materials

Appendix A: List of training and test items

	harmony	*no harmony*		*partition during cross-validation*	*frequency*
root-internal short vowel targets	ʒʏzʏk	ʒʏzɪk	'ring'	1	10/11
	kʏːsʏk	kʏːsɪk	'desert carrot'	2	10/11
	qɔɫɔn	qɔɫən	'colt'	3	16/17
	qɔːzɔ	qɔːzə	'lamb'	2	1/10
root-internal long vowel targets	tʏlʏːk	tʏlɪːk	'graduate'	2	1/11
		tʏːbɪː	'hill'	4	0/11
		qɔɫɑːq	'ear'	4	0/11
		bɔɫɑːt	'steel'	1	0/11
epenthetic targets	kʏl-ʏp	kʏl-ɪp	'laugh-CVB'	4	26/31
	qɔr-ɔp	qɔr-əp	'construct-CVB'	1	13/18
	ʏːl-ʏp	ʏːl-ɪp	'die-CVB'	2	10/18
	qɔːs-ɔp	qɔːs-əp	'add-CVB'	3	1/15
suffix short vowel targets	kʏl-dʏ	kʏl-dɪ	'laugh-3PST'	1	7/22
	ʏːl-dʏ	ʏːl-dɪ	'die-3PST'	2	1/20
	qɔr-dɔ	qɔr-də	'construct-3PST'	3	1/18
	qɔs-tɔ	qɔs-tə	'spit up-3PST'	2	4/30
	qɔːs-tɔ	qɔːs-tə	'add-3PST'	4	1/20
		qɔːj-də	'put-3PST'	4	0/11
	kʏl-dʏm	kʏl-dɪm	'laugh-1SG.PST'	4	4/12
	ʏːl-dʏm	ʏːl-dɪm	'die-1SG.PST'	3	3/11
	qɔr-dɔm	qɔr-dəm	'construct-1SG.PST'	4	3/10
	qɔːs-tɔm	qɔːs-təm	'add-1SG.PST'	1	2/10
	qɔr-dɔŋəz-dɑːr	qɔr-dəŋəz-dɑːr	'construct-2PST.FORM-PL'	3	1/12

	harmony	*no harmony*		*partition during cross-validation*	*frequency*
suffix long vowel targets		kʏl-sɪː	‘laugh-COND’	1	0/15
		ʏːl-sɪː	‘die-COND’	1	0/14
		qɔr-sɑː	‘construct-COND’	3	0/13
		qɔːs-sɑː	‘add-COND’	2	0/13
		kʏl-gɪːn	‘laugh-PFV’	4	0/2
		ʏːl-gɪːn	‘die-PFV’	1	0/1
	kʏn-dʏ	kʏn-dɪ	‘day-ACC’	2	3/10
	ʒʏzʏk-tʏ	ʒʏzʏk-tɪ	‘ring-ACC’	1	1/9
	ʏːt-tʏ	ʏːt-tɪ	‘gall bladder-ACC’	3	2/12
		kʏn-dɪː	‘day-LOC’	2	0/7
		ʏːt-tɪː	‘gall bladder-LOC’	4	0/11
		ʒʏzʏk-tɪːr-dɪ	‘ring-PL-ACC’	3	0/23

Appendix B: Additional analyses

1 Categorical dispersion-based constraints

The scalar alignment constraint used in the MaxEnt analysis in § 5 captures a large amount of variance in the gradient Kazakh data. However, many descriptions of harmony delineate the set of triggers and targets categorically. This is usually accomplished by constraint weighting alone, but is complicated by the use of a scalar constraint. To derive categorical rather than gradient predictions using a scalar constraint, a MaxEnt model would need to produce a sigmoid curve with a steep slope (see McPherson & Hayes 2016 and Hayes 2017 for sigmoids), and additionally, participants and non-participants would need to occur at or near the asymptotes of the curve. Unless positions on the curve are guaranteed to occur near the function’s minimum or maximum, then weighting is insufficient, in and of itself, to differentiate categorical from gradient using scalar constraints in MaxEnt. I sidestep this issue, as it is not central, and demonstrate my point by using strictly ranked constraints enforcing a perceptual distance threshold.

In this section, I use Kaun’s conditional alignment constraints to compel harmony. While Kaun’s constraints drive harmony when the trigger is a member of some featural class, the alignment constraint in (29) motivates harmony if the perceptual distance between the trigger and its harmonic counterpart is less than some threshold, *m*. This threshold is language-specific. This constraint predicts that the set of triggers in a given language will be the set of [+round] vowels which are closest to their harmonic counterparts.

(29) ALIGN-L/R([rd]/Δ_{xy}<m)

Align the feature [+round] to the left/right edge of the word if the distance between the [+round] trigger, x, and its [−round] harmonic counterpart, y, Δ_{xy}, is less than m.

Similarly, I define the set of targets by perceptual distance. Whereas Kaun contends that the set of targets in a language will be those where the feature [round] is most perceptually salient, I argue instead that the set of targets, like the set of triggers, is defined by weakness. This parallels the analysis of triggers – pairs separated by smaller perceptual distances will be more likely to surface via harmony. This is formalised via an IDENT constraint in (30), although it is also possible to encode the preference for minimally salient alternations via *MAP constraints that penalise output–output alternations rather than input–output mismatches (Zuraw 2013). The argument that perceptual weakness determines the set of targets follows directly from Steriade's (2009) P-map proposal in two ways. First, Steriade contends that speakers have access to the relative perceptibility of a set of alternations. Second, she argues that phonotactically illicit forms are repaired by the least salient possible alternation. For instance, a prohibition on word-final voiced obstruents is always realised via devoicing, although numerous other repairs are possible. She argues that devoicing is privileged in this way because a voicing alternations involves a smaller perceptual change than other possible alternations. In a similar vein, the vowels most likely to undergo harmony are those that involve a less salient alternation.

(30) IDENT-IO([rd]/Δ_{vw}<n)

Assign a violation to every input–output [round] pair, v and w, with a perceptual distance, Δ_{vw}, greater than some threshold, n.

Two things are worth noting at this point. First, using input–output correspondence to curtail harmony involves a significant idealisation, since in the vast majority of languages with labial harmony there exists some other harmony pattern. Thus affixes typically undergo four-way alternations, not just the two-way alternations addressed here. For any four-way alternation, evaluating perceptual distances is potentially problematic. Second, input–output correspondence constraints with phonetic detail entail phonetically specified inputs (Flemming ms).

2 Solon

In this subsection I briefly sketch an analysis of Solon with strict ranking. The constraints in (31), both of which use a threshold of 1 (see Table XI), can account for the distribution of non-initial vowels in Solon.

(31) a. Align-R([rd]/$\Delta_{xy}<1$)
Align the feature [round] to the right edge of the word if the distance between the [round] trigger, x, and its [−round] harmonic counterpart, y, Δ_{xy}, is less than 1.

b. Ident-IO([rd]/$\Delta_{xy}>1$)
Assign a violation to every input–output [round] pair, x and y, with a perceptual distance, Δ_{xy}, greater than 1.

The tableau in (32a) shows that Align-R([rd]/$\Delta<1$) must dominate Ident-IO[rd]. This ranking dictates that harmony occurs after perceptually weak triggers. (32b) shows that the dispersion-based Ident constraint must outrank Align-R([rd]/$\Delta<1$), to prevent assimilation of high vowels.

(32) a.

/toʃʃə/	Align-R([rd]/Δ<1)	Ident-IO[rd]
i. toʃʃə	*!	
☞ ii. toʃʃo		*

b.

/xoŋgi/	Ident-IO([rd]/Δ>1)	Align-R([rd]/Δ<1)	Ident-IO[rd]
☞ i. xoŋgi		*	
ii. xoŋgu	*!		*

c.

/uldə/	Ident-IO([rd]/Δ>1)	Align-R([rd]/Δ<1)	Ident-IO[rd]
☞ i. uldə			
ii. uldo			*!

In (32c), /u/ does not trigger harmony on a following non-high vowel, because it does not satisfy the conditional requirement of Align-R([rd]/$\Delta<1$), since $\Delta_{\text{u-i}}$ is 2.7, leaving faithfulness to militate against assimilation.

The general ranking instantiated above, Ident-IO([rd]/$\Delta>n$) ≫ Align-R([rd]/$\Delta<n$) ≫ Ident-IO[rd], holds for all four languages described in this section. The perceptual threshold, n, for each language varies, but the ranking schema remains constant.

Appendix C: Typological predictions

1 Duration and perceptual weakness

I have claimed that the best triggers are also the best targets for harmony, in accordance with Kaun's GesturalUniformity constraint. In all four of the languages discussed in §6, high vowels were always more perceptually distinct than the non-high vowels. This is likely the case in many languages. For this reason, Kaun (1995) claims that [+high] vowels are intrinsically better targets for harmony, because they better signal the [round] feature of the trigger vowel. Kaun's claim accounts for languages like Turkish, where, regardless of trigger height, only high vowels undergo harmony (see also Ultan 1973: 44–47). Using only the perceptual similarity-based constraints introduced thus far, this analysis cannot account for Turkish, since it seems likely that high vowels are more distinct from their harmonic counterparts than non-high vowels (Kiliç & Öğüt 2004).

If the [+high] target pattern does not derive from target salience, then one alternative is that it results from duration. For targets, shorter vowels are better targets for the same reason that auditorily similar vowels are better targets – less noticeable alternations are preferred (Steriade 2009). Moreover, there is evidence that suggests duration plays a significant role in vowel perception, particularly in more crowded regions of the vowel space (Bennett 1968, Ainsworth 1972, Hillenbrand *et al.* 2000). If shorter vowels are more difficult to correctly discriminate, then they may preferentially participate in harmony. As in §6.1, the constraint in (33) establishes a threshold to demarcate the set of undergoers from non-undergoers. (This likely relates to diachronic claims on the emergence of vowel harmony; Ohala 1994, Przezdziecki 2005.)

(33) Ident-IO([rd]/$T_v > m$)

Let v be an output vowel, with duration, T, and v be its input correspondent. If Tv is greater than some threshold, m, assign a violation to every w-v pair that disagree for [round].

Kazakh, like Maltese (Puech 1978), shows that contrastive length may play a role in harmony. I predict, however, that non-contrastive length may also factor into harmony. Many Turkic languages exhibit duration differences between the putative high and non-high vowels that mirror those of Kazakh, although there is no obvious length distinction in many of these languages. Additionally, high (or alternatively, short) vowels undergo elision in many Turkic languages, further suggesting the intrinsic shortness of these vowels (e.g. Poppe 1964, Kavitskaya 2013, Washington 2016). Given the shortness of high vowels in Turkic, a constraint like (33) offers an account for this genetic skew towards high vowel targets in Turkic. This pattern, which motivates Kaun's *[+rd, −hi] constraint, is almost entirely confined to Turkic (see Derbyshire 1979 and Casali 1995 for non-Turkic examples). In addition to defining the set of targets, duration may play a crucial role in determining the set of triggers. In Kazakh, the best triggers are not only the least dispersed vowels. They are also the short vowels. In Kachin Khakass (Korn 1969: 102–103), harmony only obtains when a high vowel triggers assimilation of another high vowel. If, as noted above, the high vowels are far shorter than the non-high vowels in Kachin Khakass, then harmony may be initiated by short vowels, targeting those same short vowels. I define a duration-related alignment constraint in (34). (It is likely that different languages respond in distinct ways to length differences, although in this paper this second dimension is proposed merely as a convenient way to account for length differences that are not well understood.)

(34) ALIGN-L/R([rd]/$T_x<n$)
Align the feature [round] to the left/right edge of the word if the duration of the trigger, *x*, is less than some threshold, *n*.

Before proceeding to the typology, we should note that the analysis curtails harmony by faithfulness. This stands in contrast to Kaun's analysis, which limits harmony by universal markedness constraints, like *[+rd, −hi]. A universal dispreference for non-high round vowels is not consistent with the dispersion-based analysis, because perceptual weakness hinges not on universal properties of each vowel, but rather on system-internal factors. If a systemic notion of perceptual weakness motivates asymmetric trigger relations, then it is more parsimonious if the same force also underlies the relevant target asymmetries. This position cannot be captured by universal markedness. Instead, perceptual similarity must be defined within each system, and therefore requires language-specific factors (though these are not necessarily faithfulness-related) rather than feature co-occurrence restrictions, like *[+rd, −hi], to play a role in the analysis.

Framed in language-specific terms, the set of targets in Kazakh is not defined by a ban on long round vowels, as is encoded by the constraint *[+rd, +long], used in §4 and §5. Like *[+rd, −hi], feature co-occurrence constraints against long round vowels, which are typologically suspect, should be replaced by constraints using language-specific detail. For Kazakh then, long vowels fail to undergo harmony due to IDENT([rd]/$T_v>m$) in (33), set to some value around 50 ms. Alternatively, long vowels may fail to undergo harmony in Kazakh because they are the most dispersed. Either way, using faithfulness to militate against harmony depends on assuming [−round] non-initial inputs, *contra* richness of the base. If a non-initial vowel is underlyingly [+round], as it may be if inputs are not restricted, then an IDENT constraint like (33) will problematically compel that vowel to surface as [+round], and thus avoid a change, rather than compelling unrounding. Perceptual distance- and duration-based IDENT constraints like those used in the analysis can therefore only promote faithfulness, not a dispreference for harmony, unless non-initial inputs are restricted.

The importance of systemic factors in harmony is also supported by recent work on the typology of ATR harmony in African languages. For instance, earlier work argued for a universal dispreference for [+high, −ATR] vowels (e.g. Archangeli & Pulleyblank 1994). This conclusion, though, is undermined by the prevalence of [+high, −ATR] vowels outside of West Africa (Casali 2008, 2014). Casali contends that language-specific inventories dictate the nature of ATR harmony, in parallel with the claim in this paper. What have been construed as universal factors affecting harmony may in reality derive from system-internal considerations.

2 An outline of the typology

In Table XIII I present a typology of labial harmony recast in system-internal terms. The typology was generated using two threshold-based alignment constraints motivating harmony from less dispersed and from shorter [+round] vowels, and two threshold-based faithfulness constraints under strict ranking. First, though, several caveats are necessary. I limit the number of [+round] vowels participating in harmony to four, partly out of convenience, and partly because even in languages with more than four [+round] vowels, like Baarin and Wiliingol, additional vowels are typically marginal, or derive historically from other vowels in the inventory. I additionally assume that a language only makes a two-way distinction for duration, even when length is non-contrastive. This stems partly from convenience, but also from the fact that languages with contrastive length tend to make only one length distinction (Remijsen & Gilley 2008). The numbers 1–4 in the typology in Table XIII represent the least to the most perceptually salient round vowels, while [short] represents the shortest (two) round vowels in a given language, even if length is non-contrastive. The predictions below involve only two harmony-driving constraints, without reference to possible stringency rankings, scalar implementations or similar formal mechanisms.

The typology predicts up to 25 patterns, given the stipulations just noted, omparable to the 24 harmony patterns predicted by Kaun's (1995) set of five constraints. (Kaun 2004: 109 notes the addition of a sixth constraint, *[+rd, −bk], increases the typological space to 36. With a seventh constraint, ALIGN-L/R([rd]/[−long]), the number of distinct patterns increases to 131.) Of the 24 possible patterns in Kaun (1995), nine are attested. Similarly, about nine patterns are attested in the proposed typology. Superficially, this proposal does not provide better empirical coverage than Kaun's constraint set. However, some key generalisations emerge from the typology that are not evident in Kaun's work. Moreover, Kaun assumes the gaps in her typology are accidental (2004: 109), but the typology in Table XIII suggests that certain gaps are very principled.

First, the set of targets is never larger than the set of triggers. There are no known labial harmonies where the set of targets is not a subset of the set of triggers. The unattested Type 16–19 patterns all violate the trigger-target subset relation. Of the six unattested patterns among Types 1–15, Types 4, 8 and 9 also violate this trigger-target subset relation. Further, of the (potentially) unattested Types 20–25, two patterns, Type 25 and potentially Type 20, also violate this relation. These gaps suggest that a theory of labial harmony cannot depend entirely on trigger restrictions (see also Nevins 2010). As argued in McCollum (2017), the typology of labial harmony does appear to depend more on GESTURALUNIFORMITY, the constraint on trigger-target relations, than constraints on triggering vowels alone.

Second, languages tend to use either duration or perceptual distance to restrict harmony, but not both, as in Types 20–25. With this in mind, the Kazakh data is particularly interesting, since triggers are defined by perceptual distance, but targets are definable either by perceptual distance or duration. Like Kazakh, the Type 21 Tofa pattern (Harrison 2000), is analysable without reference to duration, which would render it an example of the Type 3 pattern. So, it is unclear if any languages clearly exhibit a pattern that depends on both perceptual distance and duration. This observation may connect with some speech-perception research that has argued that hearers preferentially attend to only one auditory cue, even when multiple cues are available, in accordance with the claim here (Flege & Hillenbrand 1986, Goudbeek *et al.* 2008).

If these biases exist, then the proposed model fits the typology very well. If not, the typology is at least as empirically adequate as Kaun (1995), and more so than Kaun (2004). Lastly, the gaps in her typology could be construed as accidents, but the gaps in Table XIII fall into several classes, all of which suggest plausible reasons for their non-existence. The proposed typology thus offers new insights into labial harmony to motivate further typological, experimental and psycholinguistic work.

macrotype	type		triggers	targets	examples
unrestricted	unrestricted	1	unrestricted	unrestricted	Akan, Altai Tuvan
same parameter	least dispersed trigger – least dispersed target	2	[1]	[1]	
		3	[1,2]	[1]	Tofa?
		4		[1,2]	
		5	[1,2]	[1,2]	Khalkha, Solon
		6	[1,2,3]	[1]	
		7	[1,2,3]	[1,2]	Kazakh?
		8	[1]	[1,2,3]	
		9	[1,2]	[1,2,3]	
		10	[1,2,3]	[1,2,3]	Kyzyl Khakass
	short trigger – short target	11	[short]	[short]	Kachin Khakass, Yeyi
target restrictions	unrestricted trigger – least dispersed target	12	unrestricted	[1]	Mayak, Meadow Mari
		13	unrestricted	[1,2]	
		14	unrestricted	[1,2,3]	Older Kazakh
	unrestricted trigger – short target	15	unrestricted	[short]	Turkish, Crimean Tatar
trigger restrictions	least dispersed trigger – unrestricted target	16	[1]	unrestricted	
		17	[1,2]	unrestricted	
		18	[1,2,3]	unrestricted	
	short trigger – unrestricted target	19	[short]	unrestricted	
different parameters	least dispersed trigger – short target	20	[1]	[short]	
		21	[1,2]	[short]	Tofa?
		22	[1,2,3]	[short]	Kazakh?
	short trigger – least dispersed target	23	[short]	[1]	
		24		[short]-[1,2]	
		25		[short]-[1,2,3]	

Table XIII

Schematic typology of labial harmony (unexpected patterns are shaded, and languages with multiple possible analyses are entered twice and indicated with ?). For the triggers and targets columns, 1 = least dispersed and 4 = most dispersed. For the example languages cited, see Dolphyne (1988), Harrison (2000), Seidel (2008), Vaysman (2009), Estill (2012) and Kavitskaya (2013).

ADDITIONAL REFERENCES

Archangeli, Diana & Douglas Pulleyblank (1994). *Grounded phonology*. Cambridge, Mass.: MIT Press.

Casali, Roderic F. (1995). Labial opacity and roundness harmony in Nawuri. *NLLT* **13**. 649-663.

Casali, Roderic F. (2008). ATR harmony in African languages. *Language and Linguistics Compass* **2**. 496-549.

Casali, Roderic F. (2014). Assimilation, markedness and inventory structure in tongue root harmony systems. Ms, Trinity Western University & SIL. Available as ROA-1207 from the Rutgers Optimality Archive.

Derbyshire, Desmond C. (1979). *Hixkaryana*. Amsterdam: North Holland.

Estill, Dennis (2012). Revisiting the Meadow Mari vocalic system. *Linguistica Uralica* **48**. 228-237.

Flege, James Emil & James Hillenbrand (1986). Differential use of temporal cues to the /s/–/z/ contrast by native and non-native speakers of English. *JASA* **79**. 508–517.

Flemming, Edward (ms). The realized input. MIT. Available (February 2018) at http://web.mit.edu/flemming/www/paper/RI2.pdf.

Goudbeek, Martijn, Anne Cutler & Roel Smits (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication* **50**. 109-125.

Hayes, Bruce (2017). Varieties of Noisy Harmonic Grammar. In Karen Jesney, Charlie O'Hara, Caitlin Smith & Rachel Walker (eds.) *Proceedings of the 2016 Annual Meeting on Phonology*. http://dx.doi.org/10.3765/amp.v4i0.3997.

Kavitskaya, Darya (2013). Segmental inventory and the evolution of harmony in Crimean Tatar. *Turkic Languages* **17**. 86-114.

Kiliç, Mehmet Akif & Fatih Öğüt (2004). A high unrounded vowel in Turkish: is it a central or back vowel? *Speech Communication* **43**. 143-154.

Poppe, Nicholas (1964). *Bashkir manual: descriptive grammar and texts with a Bashkir–English glossary*. Bloomington: Indiana University & The Hague: Mouton.

Remijsen, Bert & Leoma Gilley (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *JPh* **36**. 318-344.

Vaysman, Olga (2009). *Segmental alternations and metrical theory*. PhD dissertation, MIT.

Zuraw, Kie (2013). *MAP constraints. Ms, University of California, Los Angeles. Available (February 2018) at http://linguistics.ucla.edu/people/zuraw/dnldpprs/star_map.pdf.