Perceptual Motivations for Parasitic Restrictions in Vowel Harmony

Wendell Kimper
University of Massachusetts, Amherst

1. Introduction

In languages with parasitic harmony, agreement with respect to some (primary) feature is dependent on agreement with respect to another (secondary) feature. For example, in Yawelmani, rounding harmony is parasitic on height (Cole and Kisseberth 1995, and others); harmony occurs between stem and suffix vowels that agree in height, but not between vowels of differing height (1).

In Yawelmani, both high vowels (1b) and non-high vowels (1d) are able to trigger rounding harmony, and both are able to undergo harmony, but a non-high vowel cannot trigger harmony on a high vowel (1f), and a high vowel cannot trigger harmony on a non-high vowel (1h). The restriction, then, is parasitic — rounding harmony only occurs among vowels that are of the same height. Rounding harmony is also parasitic on height in a variety of Turkic languages; see Kaun (1995) for a survey. Additionally, ATR harmony may be parasitic on height, as it is in Vata (Kaye 1982) and Phuthi (Donnelly 2000).

Kaun explains the motivation for parasitic restrictions in articulatory terms. Citing Linker (1982), she notes that the rounding gesture in high vowels is both quantitatively and qualitatively different than the rounding gesture in non-high vowels. She therefore proposes a constraint, UNIFORMITY, requiring a feature to be implemented with a uniform gesture. In this paper, I propose that parasitic harmony is perceptually motivated — the perceptual benefits of harmony are undermined by secondary-feature disagreement — and present the results of a discrimination study supporting that proposal.

The paper is organized as follows. Section 2 provides an overview of the perceptual advantages of harmony. Section 3 presents the design and methods of the current experi-
ment, Section 4 presents the results, and 5 discusses the implications of those results for the typology of harmony processes.

2. Background: Perception and Harmony

Harmony has traditionally been analyzed as an articulatorily motivated phenomenon — it increases articulatory ease by minimizing the number of gestural transitions a speaker makes throughout the course of producing a word. However, there are also perceptual benefits to harmony.

Suomi (1983) proposes that harmony reduces the perceptual burden on the listener by rendering contrasts predictable in non-initial positions. Because initial syllables are psycholinguistically prominent, the likelihood of accurate perception there is higher than in non-initial syllables; if non-initial syllables are predictable, the listener does not need to carefully attend to subsequent cues. Furthermore, because subsequent vowels agree with the initial vowel, the cues in those vowels can serve to reinforce the initial hypothesis that a listener has made about the feature value of the initial vowel, further increasing the likelihood of an accurate identification.

Kaun (1995), building on Suomi’s proposal, focuses on the hypothesis that extending the duration of realization of a particular feature enhances its perceptibility. She notes, following Steriade (1995), that perceptually difficult contrasts that are subject to positional neutralization tend to be licensed in positions of greater duration (e.g. word edges and prosodically strong positions). This permits not only a more complete articulatory realization, but also a more robust opportunity for the listener to correctly identify the cue.

In vowel harmony, rather than simply limiting a difficult contrast to a position of greater duration, the cues are realized across multiple segments — extending the duration of a feature’s realization and thus increasing the likelihood of a correct identification. Like Suomi, Kaun proposes that in harmony domains, cues to the quality of later vowels serve to reinforce a listener’s initial hypothesis about the feature value in question.

Gallagher (2010) discusses the perceptual advantages of harmony not in terms of feature identification but in terms of maximizing the distinctness of lexical items. She argues that assimilatory co-occurrence restrictions on laryngeal features are driven by the pressure for roots to be as perceptually discriminable as possible. In a language with laryngeal assimilation, words with different specifications for laryngeal features consistently differ with respect to those features on all relevant consonants. Words with multiple differences are more perceptually distinct from one another than words with only a single difference, so laryngeal assimilation aids in perception.

Gallagher supports this claim experimentally, presenting results of several discrimination studies in which subjects were presented with pairs of disyllabic (CVCV) nonce words and asked to make a same/different judgment. In the first study, each nonce word contained either zero, one, or two ejectives; in the second study, each nonce word contained either zero, one, or two aspirates. Results from both studies patterned similarly — subjects were significantly more accurate at discriminating words that differed by laryngeal features in both consonants than words that differed by laryngeal features in only one. On other words, a contrast between zero and two ejectives/aspirates was more readily per-
ceptible than either a contrast between zero and one or a contrast between one and two ejectives/aspirates.

Gallagher’s results are based on the perception of consonantal rather than vocalic features, but the broader effect should be general — in a language with harmony, words are consistently more different from each other than in a language without harmony, and are therefore more easily distinguished.

Suomi (1983), Kaun (1995), and Gallagher (2010) each take a slightly different perspective on the perceptual advantages of harmony, but the central claims are similar: harmony is advantageous because it enhances the perceptual salience of difficult contrasts.

If harmony is perceptually advantageous because realizing a feature across multiple segments aids in identifying that feature’s value, then listeners must be interpreting those segments as reflexes of a single feature. This interpretation is more likely if the realization is consistent across segments, as it is when rounding is spreading among vowels that already agree in height. When rounding is spreading among vowels which disagree in height, however, the acoustic effects of that rounding will differ throughout the course of the feature’s realization. While Kaun (1995)’s discussion is limited to rounding, the same reasoning applies to cases where ATR is parasitic on height.

This non-uniformity among segments manifesting a feature should interfere with a listener’s interpretation of the harmonized segments as manifesting a single contrast, therefore diminishing the perceptual advantage of harmony. Under this account, parasitic restrictions represent a strategy for deploying harmony only where it will be the most beneficial — among segments that agree with respect to some secondary feature.

3. Methods

In this experiment, subjects are presented with pairs of disyllabic nonce words and asked to make a same/different judgment. There are two distinct yet closely related hypotheses under consideration. The first hypothesis is that harmony is perceptually advantageous because it maximizes the discriminability of words — subjects should be faster and more accurate at distinguishing pairs of words where both vowels are different (as is the case in languages with harmony) than pairs of words where only one vowel differs (as is the case in languages without harmony). The second hypothesis is that this advantage is diminished when harmony occurs among vowels which disagree with respect to a secondary feature — subjects should be faster and more accurate at distinguishing pairs of words where each word is internally harmonic for both the primary and secondary feature than pairs of words where each word is internally harmonic for one feature but not the other.

3.1 Stimuli

The primary harmonizing feature used in this experiment is ATR, and the secondary feature is height. The stimuli were disyllabic nonce words, of the shape CVCV; the consonants [h] and [g] were chosen because of their low coarticulatory effect on surrounding vowels, and vowels were manipulated for height (high or mid) and ATR. A list of the resulting nonce words can be found in the left-hand column in Table 1.
<table>
<thead>
<tr>
<th>Items</th>
<th>Same</th>
<th>Different (1)</th>
<th>Different (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hege</td>
<td>hege--hege</td>
<td>hege--hige</td>
<td>hege--hege</td>
</tr>
<tr>
<td>hegE</td>
<td>hege--hege</td>
<td>hege--hige</td>
<td>hege--hege</td>
</tr>
<tr>
<td>hEge</td>
<td>hege--hege</td>
<td>hege--hige</td>
<td>hege--hege</td>
</tr>
<tr>
<td>hegI</td>
<td>hegI--hegi</td>
<td>hige--hige</td>
<td>hegI--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
<tr>
<td>hEgi</td>
<td>hEgi--hEgi</td>
<td>hEgi--hEgI</td>
<td>hEgi--hEgi</td>
</tr>
</tbody>
</table>

Table 1: Pairs of nonce-word stimuli.

Each CV syllable was recorded separately, in a neutral frame sentence, read by a phonetically trained native speaker of North American English in a sound-attenuated booth. For [+ATR] vowels, the diphthong portion was removed. Using Praat (Boersma and Weenink 2008), syllables were equalized for $F_0$ (220hz), consonant duration (50ms), and vowel duration (150ms), then spliced together to form the nonce words. Based on information from a pilot study, a small amount of pink noise was added.

While there are some differences between the tense/lax contrast used in languages like English and the ATR contrast in languages with tongue root harmony (see e.g. Lindau-Webb 1987), the manipulations above resulted in vowel stimuli that closely resembled the acoustic properties typical of ATR languages (see Starwalt 2008 for further discussion of those acoustic characteristics).

### 3.2 Subjects and Task

The experiment used an AX discrimination task. Subjects were 36 native speakers of North American English, students in introductory linguistics courses recruited by offering extra credit points. They were presented with the nonce word stimuli in pairs, with an ISI of 500ms between words, and asked to make a same/different judgment. In the pairs, given in Table 1, the two nonce words differed with respect to the ATR value of one of the vowels, both of the vowels, or neither of the vowels (the “same” condition).

Stimuli were presented in blocks, with a 96 trials per block. There were 6 blocks, and each subject saw each “same” pair a total of 18 times, and each “different” pair a total
Figure 1: Accuracy and response time as a function of the number of differing vowels between members of stimulus pairs, averaged across all subjects and all items. Error bars represent 95% confidence intervals.

of 6 times, resulting in an equal number of trials for which “same” and “different” were the correct responses. Responses were cut off at 1500ms, and no subject failed to respond on more than 10% of the trials.

4. Results

4.1 One Difference vs. Two

As seen in Figure 1, subjects were significantly faster and more accurate to distinguish pairs of stimuli where the members of the pair differed by the ATR values of both vowels (e.g. hege ~ hEgE) than when the members of the pair differed by the ATR value of only one member of the pair (e.g. hege ~ hGe).

A Linear Mixed Effects Model\(^1\) with log-transformed response time as the dependent variable and random effects for subject and item found the difference in response time to be significant (t = –4.83, p < 0.0001). A Mixed Logit Model with correct responses as the dependent variable and random effects for subject and item found the difference in accuracy to be significant (z = –4.84, p < 0.0001).

Only the “different” pairs were of interest; the “same” pairs were excluded from the analysis, as they were primarily included in the design as distractors to balance the responses.

\(^1\)Mixed models were done using the lme4 package (Bates and Maechler 2010) in R (R Development Core Team 2009). For non-logit models, p-values were obtained via Markov Chain Monte Carlo sampling using the languageR package (Baayen 2010).
Figure 2: Accuracy and response time as a function of height harmony and ATR harmony, averaged across all subjects and all items, for pairs where both vowels differed. Error bars represent 95% confidence intervals.

4.2 Height and ATR Interaction

Among pairs where the members of the pair differed by the ATR value of both vowels, there was a significant interaction between ATR harmony and height harmony. For pairs where each nonce word was internally harmonic for height, subjects were slightly (but not significantly) faster to respond to pairs where each nonce word was also internally harmonic for ATR (e.g. hege ~ hEgE) than pairs where each nonce word was internally disharmonic for ATR (e.g. hegE ~ hEge). However, for pairs where each nonce word was internally disharmonic for height, subjects were significantly faster to respond to pairs where each nonce word was also internally disharmonic for ATR (e.g. hegI ~ hEgi) than pairs where each nonce word was internally harmonic for ATR (e.g. hegi ~ hEgI).

Accuracy was high across the board, and a Mixed Logit Model with correct responses as the dependent variable and random effects for both subject and item found no significant differences among conditions. Because accuracy was so high, this lack of difference is likely attributable to a ceiling effect.

A Linear Mixed Effects Model with log-transformed response time as the dependent variable and random effects for both subject and item found that, within pairs of height-disharmonic words, there was a significant effect of ATR harmony (t = 2.61, p < 0.01). Within pairs of height-harmonic words, the effect of ATR harmony failed to reach significance (t = 0.9, p > 0.05). There was a significant interaction between height harmony and ATR harmony (t = –2.48, p < 0.05).

In light of this interaction, a comparison was made between height-disharmonic pairs whose members differed by the ATR value of one vowel (e.g. hegi ~ hEgi) and height-disharmonic pairs whose members differed by the ATR value of both vowels and were harmonic for ATR (e.g. hegi ~ hEgI). Figure 3 shows that subjects were faster and more accurate in the ATR harmonic pairs differing by two vowels than the pairs differing...
Perceptual Motivations for Parasitic Restrictions in Vowel Harmony

Figure 3: Accuracy and response time for pairs of height-disharmonic words, averaged across all subjects and all items, comparing pairs whose members differed by one vowel to pairs whose members differed by both vowels and were ATR harmonic. Error bars represent 95% confidence intervals.

by only one vowel.

A Linear Mixed Effects Model with log-transformed response time as the dependent variable and random effects for both subject and item found this difference in response time to be significant ($t = -3.95, p < 0.001$). A Mixed Logit Model with correct responses as the dependent variable and random effects for both subject and item also found the difference in accuracy to be significant ($z = -3.95, p < 0.0001$).

5. Discussion

The results of this experiment support the claim that harmony is perceptually advantageous because it maximizes the distinctness of words, and that this advantage is diminished in contexts where listeners are less likely to interpret multiple segments bearing the same feature as representing a single contrast.

The central result (shown in Figure 1) that words differing by multiple segments are distinguished faster and more accurately replicates one of the results of Gallagher (2010)’s study of laryngeal contrasts, and should hardly come as a surprise. Words differing by multiple segments are more recognizable as different because they are, in fact, more different — indeed, it would be surprising if this wasn’t the case.

However, the fact that this result is unsurprising should not be mistaken for triviality. Gallagher shows how the pressure to maximize the distinctness of words can be used to explain both assimilatory and dissimilatory co-occurrence restrictions — either type of restriction ensures an overall system where distinct words consistently differ by the feature values of multiple segments.

Of present interest is the specific advantage of harmony as a strategy for maximiz-
ing perceptual distinctness. The results in Figure 2 can be seen as addressing the tradeoff between assimilatory and dissimilatory restrictions. Pairs that differ by two vowels and are ATR harmonic represent forms that obey an assimilatory co-occurrence restriction, and pairs that differ by two vowels and are ATR disharmonic represent forms that obey a dissimilatory co-occurrence restriction.

In Figure 2 we see that, when the vowels in a word agree with respect to a secondary feature (height), there is no particular advantage of either assimilation or dissimilation — both are more or less equally effective at easing the perceptual burden to the listener, because both ensure that the words in the pair differ by both vowels. However, when the vowels of a word disagree with respect to a secondary feature, harmony is at a clear disadvantage.

Parasitic harmony, then, is a way of deploying harmony only when it will be most perceptually beneficial — when both vowels agree with respect to some secondary feature. Crucially, non-parasitic harmony is still advantageous — in Figure 3, we see that even when words disagree with respect to the secondary feature, an assimilatory co-occurrence restriction that ensures multiple differences between words would limit the set of possible comparisons to those which are faster and more accurate (the ATR harmonic pairs) than the comparisons in a language with no co-occurrence restriction (the pairs differing by only one vowel).

The pattern of perceptual advantage seen in this experiment reflect the attested typology of harmony systems — harmony is beneficial even when it is not parasitic, but it is even more beneficial when there is secondary-feature agreement. There are languages which use harmony across both secondary-agreement contexts and secondary-disagreement contexts, and languages which make use of harmony only in secondary-agreement contexts, but crucially there are no languages where harmony is anti-parasitic — no languages where vowels assimilate only if they disagree with respect to some other feature.

This does raise a question, though, about the advantages of dissimilatory co-occurrence restrictions for vowels. While vowel harmony is robustly attested, vowel dissimilation is strikingly rare — it occurs in Woleian Sohn (1975) and other Oceanic languages (see Lynch 2003 for further discussion). Dissimilation is, however, robustly attested for consonantal features (Alderete 1997, Gallagher 2010, Odden 1994, Yip 1988, and many others). One explanation for this difference concerns articulation — harmony is advantageous both perceptually and articulatorily; dissimilation is only advantageous perceptually, and is articulatorily marked.

The prevalence of consonant dissimilation may be due to a difference between consonant articulation and vowel articulation — vowel articulations overlap considerably, while consonant articulations (particularly those for e.g. laryngeal features) overlap substantially less (or not at all). An additional prediction of the results of this experiment is that dissimilation should be anti-parasitic — because dissimilation is of the most benefit in contexts of secondary-feature disagreement, we expect to find languages which dissimilate only in those contexts. While there are too few languages with vowel dissimilation to know if this prediction is borne out, it is attested in consonant co-occurrence restrictions — see Gallagher (2010) for a discussion of languages with “identity effects”, where consonants dissimilate for laryngeal features only if they are non-homorganic.
6. Conclusion

In this paper, I presented the results of a discrimination study designed to test the hypothesis that harmony is perceptually advantageous because it maximizes the perceptual distinctness of words. Furthermore, the experiment tested the hypothesis that this advantage is diminished in cases of secondary-feature disagreement, where listeners would be less likely to interpret multiple segments bearing the same feature as representing a single contrast.

Subjects were indeed faster and more accurate to discriminate words which differed by feature values of multiple segments than words which differed by only a single segment. Furthermore, this advantage was diminished in words with secondary-feature disagreement, supporting the claim that parasitic harmony patterns are perceptually motivated.

References


Bates, Douglas, and Martin Maechler. 2010. lme4: Linear mixed-effects models using s4 classes. URL http://CRAN.R-project.org/package=lme4, r package version 0.999375-33.


