Not Crazy After All These Years?

Perceptual Grounding for Long-Distance Vowel Harmony[*]

Wendell Kimper

University of Manchester

Draft: November 12, 2016

Abstract

Long-distance (or 'transparent') vowel harmony systems have frequently been considered 'unnatural' and analysed as 'crazy rules' (Bach and Harms, 1972) because they violate the principle of *strict locality*. Articulatory explanations for the phonetic grounding of vowel harmony are unable to extend to non-local processes, and attempts to re-analyse cases of transparent harmony as strictly local have been largely unsuccessful. In this paper, I present experimental evidence suggesting that vowel harmony may be perceptually (as well as articulatorily) grounded, and that this source of phonetic grounding does in fact extend to long-distance as well as local harmony. In a series of four experiments, subjects were presented with a nonsense word followed by an isolated vowel, and asked to report whether the isolated vowel had occurred in the preceding word. Subjects were consistently faster and more accurate in nonsense words which exhibited vowel harmony along the relevant feature dimension, regardless of locality. A fourth experiment included a task requiring subjects to identify whether the vowel occurred in a specific syllable, and here too they showed better performance on items with vowel harmony along the relevant feature dimension. I argue that strict locality is

not a necessary component of a phonetically grounded theory of vowel harmony, suggesting that long-distance harmony can be analysed as an explicitly non-local process without abandoning phonetic grounding.

1   Introduction

Vowel harmony — the requirement that all vowels within a prosodic or morphological domain must agree with respect to some feature — is commonly cited as an example of a *phonetically grounded* or *natural* process (see in particular Archangeli and Pulleyblank 1994; Ohala 1994). There is no general consensus regarding the precise criteria against which a process can be judged to be natural, but there is broad agreement that phonetically grounded processes either confer some advantage (usually to the speaker, in the form of reduced articulatory effort) or arise through sound changes that are plausibly attributable to channel bias in generational transmission.

Harmony processes are typically considered natural because their prevalence is attributable to articulatory factors — assimilation reduces the number of articulatory gestures necessary to execute a phonological form. For example, harmony along the Front/Back dimension in Finnish (1) means that only a single fronting or backing gesture is required for the entire word.

(1)   **Finnish**: Front/Back Harmony (Local)
    a.   pøytæ-næ   'table-ESS'   *pøytæ-na
    b.   pouta-na   'fine weather-ESS'   *pouta-næ

This assumption has been implicit in much of the literature on vowel harmony (Archangeli and Pulleyblank, 1994; Bakovic, 2000; Gafos, 1999; Ni Chiosain and Padgett, 2001, and others), but see e.g. Boersma (1998); Riggle (1999) for explicit analyses based on articulatory ease. Another perspective on the same source of explanation can be found in Evolutionary Phonology (Blevins, 2004; Ohala, 1994, and others), where the genesis of harmony is attributed to listeners' misinterpretation of gradient coarticulation as categorical assimilation.

An explanation of harmony as an articulatorily grounded process necessarily entails a commitment to the principle of *strict locality* — an inviolable restriction which limits phono-

logical dependencies to adjacent segments. This follows from articulatory grounding because assimilation only reduces gestural transitions among articulatorily adjacent segments, so agreement among distant segments (without assimilation of intervenors) provides no articulatory advantage. Likewise, gradient vowel-to-vowel coarticulation is a local phenomenon — so the effects it introduces should also be local.

Strict locality is an attractive principle, and has enjoyed a privileged analytical status in phonological theory. Recent work, however, has questioned its axiomatic status, showing that there are in fact instances where harmony results in non-adjacent dependencies — as in Finnish (2), where harmony takes place across 'neutral' vowels [i] and [e] (see Section 2.1 for further discussion of non-adjacent dependencies). Consequently, a number of different approaches to harmony have adopted explicitly non-local representations — see in particular Agreement by Correspondence (Hansson, 2001; Rose and Walker, 2004; Rhodes, 2012, and others) and Trigger Competition (Kimper, 2011) for theoretical developments in this area.

(2)   **Finnish**: Front/Back Harmony (Non-Local)
    a.   puhe̲-han   'speech-EMPH'   *puhe-hæn
    b.   tsaari̲-na   'car-ESS'   *tsaari-næ

This creates something of an impasse — the existence of non-local dependencies in harmony seems to be at odds with phonetic grounding, which requires strict locality. One analytical strategy to resolve the impasse has been to introduce theoretical machinery which produces surface non-locality via strictly local phonological operations, either through derivational opacity or tier- or contrast-based definitions of locality (Clements, 1977; Kiparsky, 1981, and others). These approaches, however, ultimately lose contact with the phonetic grounding that motivates the strict locality they seek to maintain, and still fall short of full empirical coverage. See Section 2.1 for further discussion.

It is possible, of course, to abandon phonetic grounding entirely — to treat non-local harmony as a *crazy rule* (Bach and Harms, 1972), an accident of diachrony in which a sequence of separate phonetically motivated sound changes has resulted in an unnatural synchronic alternation. However, non-local harmony is robustly typologically attested across a genetically diverse range of languages and with a number of different features[1] — treating

---

[1]Including, but not limited to: Hungarian (Uralic, palatal harmony; Vago 1976), Wolof (Niger-Congo,

it as aberrant misses important generalisations about the nature of harmony itself.

However, if non-locality in harmony is to be reconciled with phonetic grounding, the first requirement is a plausible source of grounding which does not depend on strict locality. In this paper, I argue that it may in fact be possible to maintain both phonetic grounding and explicitly non-local representations. I present experimental evidence demonstrating that harmony confers a *perceptual* advantage — even when it is non-local. The paper is organised as follows. Section 2 provides a discussion of the empirical difficulties faced by strict locality, as well as an overview of previous work discussing perceptual approaches to vowel harmony; Sections 3–7 present a series of experiments designed to probe the perceptual effects of harmony among both adjacent and non-adjacent vowels; and Section 8 discusses potential connections between the perceptual asymmetries found in those experiments and phonetically grounded theories of phonology.

2   Background

2.1   Against Strict Locality

In this context, the *strict* in strict locality refers to its inviolability. This applies both to the set of possible phonological representations — e.g. the "no line crossing constraint" in autosegmental phonology (Goldsmith, 1976), or Archangeli and Pulleyblank's (1994) "precedence principle" — and to the possible computations performed on those representations, preventing e.g. CON in Optimality Theory (Prince and Smolensky, 1993/2004) from containing constraints which assign violation marks in ways that involve dependencies between non-adjacent segments.

There are a number of ways to define adjacency for the purpose of assessing strict locality. The definition most closely tied to phonetic grounding — *articulatory* adjacency (Gafos, 1998) — serves as a useful starting point for the discussion. According to this definition of locality, vowel harmony is free to disregard consonants, since vowel gestures overlap articulatorily regardless of intervening consonants. It should not, however, apply across intervening

tongue root harmony; Archangeli and Pulleyblank 1994), Khalkha Mongolian (Altaic, rounding harmony; Kaun 1995), and Menominee (Alogonquian, height harmony; Bloomfield 1962).

vowels.

A number of potential examples of long-distance harmony can be understood as still maintaining strict articulatory locality. In Kinande tongue root harmony, the minus [−ATR] vowel [a] lacks a contrastive [+ATR] counterpart; it has been described as transparent, allowing harmony to take place across it (Schlindwein, 1987). However, Gick et al. (2006) provide ultrasound evidence showing that [a] in Kinande covertly undergoes tongue root harmony and surfaces as [ʌ] in [+ATR] harmonic domains. Additionally, Gafos (1998) notes that, in apparent cases of long-distance consonant harmony, coronal features like anteriority can be manifested articulatorily on intervening vowels and non-coronals without interfering with their realisation. Both of these seemingly long-distance processes, then, can credibly be interpreted as articulatorily local.

However, while some cases may yield to reanalysis, a strong case can be made that this version of strict locality is not empirically tenable. One well-studied case of harmony resulting in non-adjacent surface dependencies is Hungarian palatal harmony (Vago, 1976; Kiparsky, 1981; Hayes and Londe, 2006; Hayes et al., 2009, and many others), which may apply across certain intervening vowels. The data in (3) illustrate the general process; suffix vowels alternate to agree with the stem, with front stems selecting a front suffix vowel and back stems selecting a back suffix vowel.

(3) **Hungarian**: Suffix vowels agree with back/front stems

    a.   hɔd   hɔd-nɔk   'army(-DAT)'

    b.   kuːt   kuːt-nɔk   'well(-DAT)'

    c.   tœk   tœk-nɛk   'pumpkin(-DAT)'

    d.   fyst   fyst-nɛk   'smoke(-DAT)'

The non-low, front, unrounded vowels [i] and [e] have no back counterparts in the inventory, and do not alternate in harmony. When these vowels intervene between another stem vowel and a suffix vowel, they behave as transparent — they are skipped over by harmony, and the suffix vowel is determined by the preceding stem vowel. This is illustrated in (4).

(4) **Hungarian**: High front [i,iː] is transparent

a. ɟumi ɟumi-nɔk 'rubber(-DAT)' (*ɟumi-nɛk)

b. pɔpiːr pɔpiːr-nɔk 'paper(-DAT)' (*pɔpiːr-nɛk)

While phonemic descriptions of the relevant data are not in dispute, the actual transparency of these intervening vowels should not be taken for granted from these descriptions — [i] and [e] lack contrastive back counterparts, and it's conceivable that they could be realised as something closer to [ɯ] and [ɤ] in back harmonic domains. Can a case be made that Hungarian, like Kinande, does not involve true transparency? Benus and Gafos (2007); Benus (2005) provide articulatory data (electromagnetic articulography and ultrasound) which suggests that front [i] and [e] are realised with less extreme articulatory gestures in a back-harmonic context — however, their results show that these segments are still articulatorily front. This is consistent with acoustic evidence from similar transparent vowels in Finnish (Gordon, 1999). Likewise, consonant harmony processes cannot all be explained by appealing to gestural locality. See Hansson (2001) for a discussion of a broader range of harmony processes than is addressed by Gafos (1998), and see Gallagher (2010) for a careful acoustic analysis demonstrating the non-locality of laryngeal co-occurrence restrictions in Quechua.

Benus and Gafos argue that the gradient coarticulation seen in Hungarian is sufficient to predict the backness of a following suffix vowel. In their analysis, long-distance harmony consists of two processes — one where a back vowel triggers gradient retraction on a following front vowel (in particular [i] and [e], which are not able to surface as fully back), and one where a slightly retracted front vowel triggers categorical backness on a following vowel. It is this latter process, needed to maintain strict locality, which loses touch with the phonetic grounding it seeks to preserve. What's missing from this account is an explanation of how following a slightly retracted front vowel with a categorically back vowel constitutes an articulatory improvement over following it with another categorically front vowel. Szeredi (2012) also provides experimental evidence that the size of the effect that Benus and Gafos found falls below the Just Noticeable Difference threshold, meaning that Hungarian speakers are not able to meaningfully perceive or make use of this level of acoustic detail.

Furthermore, Kimper (2011) notes that splitting long-distance harmony into separate processes in this way is typologically undesirable, since they would then be predicted to occur independently. This means that a front vowel which is gradiently retracted for reasons

having nothing to do with vowel harmony — the coarticulatory effects of a uvular consonant, for example — could potentially trigger categorical backness on a following vowel. To my knowledge, no language exists which instantiates this predicted pattern.

Articulatory adjacency is not, of course, the only way of defining locality, and rejection of the principle on the basis of the most stringent proposal alone would constitute an argument against a straw man. However, as we will see, adopting alternative definitions of adjacency does not particularly improve the situation.

One common approach is to assess adjacency in terms of phonological *tiers*, usually based on contrastive features (Kiparsky, 1981, and others).[2] Dependent segments must be adjacent on their tier, and material that is *not* part of that tier is disregarded. In Kiparsky's (1981) analysis of Hungarian, for example, the relevant tier consists of those segments which are contrastively specified for [±BACK] — because [i] and [e] lack back counterparts, and their feature value is fully predictable, they are unspecified. Since unspecified segments are not part of the relevant tier, they can be skipped — no association lines are crossed, and these representations do not run afoul of strict locality.

The central predictions of this approach are that (a) only unspecified segments can be skipped, and (b) specified segments which do not undergo harmony will necessarily block it. However, Kimper (2013) notes that there are exceptions to both of these generalisations. In Khalkha Mongolian (Kaun, 1995), for example, high vowels do not undergo colour harmony — [i] and [u] contrast, but [i] is nonetheless transparent to harmony. In Finnish, vowels which would undergo harmony in native stems do not alternate in loanwords — despite being fully contrastive in the inventory, these vowels are variably transparent. Furthermore, the very similar dialects of Ifẹ and Ọyọ Yoruba illustrate the fact that identical vowel inventories can exhibit either transparency or opacity — high vowels [i] and [u] lack [-ATR] counterparts in both dialects, but they are transparent in Ifẹ and opaque in Ọyọ (Pulleyblank, 1996). Regardless of how phonological contrast is determined, Ifẹ and Ọyọ should behave identically — the fact that one dialect exhibits transparency and the other exhibits opacity suggests that defining locality in terms of feature tiers is unsuccessful.

---

[2]Specific implementations of this approach vary in their criteria for identifying tiers and/or assigning contrastive features, but the fundamental mechanisms are the same.

An alternative approach attempts to maintain strict locality (at least at the phonological level) by appealing to derivational opacity.[3] Transparent segments first undergo strictly local harmony, and subsequent neutralisation reverts their feature specification, resulting in disharmony (Clements, 1977; Walker, 1998; Bakovic, 2000). This falls victim to the same problems as contrast-based approaches — because a derivationally opaque analysis of non-locality requires a process of absolute neutralisation, only segments which do not contrast for the spreading feature can be transparent. As we have just seen, this is not the case.

It seems, then, that the empirical predictions of strict locality are not borne out — segments which are non-adjacent, however adjacency is defined, may nonetheless exhibit phonological dependency in vowel harmony. This is, of course, not to suggest that locality plays no role in phonological processes; the choice here is not between strict locality and 'anything goes'. While non-local vowel harmony does exist, it is subject to an implicational hierarchy: any language with non-local vowel harmony also has local harmony, but not vice versa, and local harmony is more typologically frequent. A successful locality restriction, then, will need to predict that locality is preferred, but violable; a number of current proposals, including Agreement by Correspondence (Hansson, 2001; Rose and Walker, 2004) and Trigger Competition (Kimper, 2011), build this into phonological theory. It may also be possible to attribute a preference for locality to processes that occur in learning, for example an attentional bias privileging adjacent dependencies.

The central question at issue in this paper is whether the arguments against strict locality also constitute an argument against phonetic grounding — does a representational system that can encode non-local harmony necessarily divorce itself from the phonetic precursors to or motivations for local harmony? The following section discusses the potential role of perceptual (rather than exclusively articulatory) grounding factors in harmony, which are not necessarily limited to strictly local application.

---

[3]As with tier- or contrast-based locality, there are a number of different implementations of this — the arguments here concern the fundamental approach, abstracting away from implementational details.

2.2 Perceptual Grounding for Harmony

The idea that harmony might be beneficial to listeners is not novel. Suomi (1983) proposes that palatal harmony is a way of facilitating the perception of F2 contrasts, as it renders feature values predictable outside of psycholinguistically prominent positions (in his examples, initial syllables). Kaun (1995), building on Suomi (1983)'s proposal, argues that extending the duration of realisation of a particular feature (by realising it across multiple segments) provides the listener with greater opportunity to correctly identify that feature value. Kaun builds on an argument from Steriade (1995a), who suggests that positional neutralisation processes affecting vowels serve to provide the listener with more robust cues to vowel contrast. Limiting contrasts to prosodic positions with greater duration gives speakers better opportunity to hit the relevant articulatory target, resulting in more robust cues across a longer time-span for the listener. Similarly, Walker (2005) argues that harmony targeting prominent positions (in her examples, stressed syllables) can serve as a form of feature licensing.

Gallagher (2010) discusses the perceptual advantages of harmony in terms of maximising the distinctness of lexical items, drawing connections to Dispersion Theory (Flemming, 1995, 2004, 2006). She argues that assimilatory co-occurrence restrictions on laryngeal features are driven by a pressure for roots to be as discriminable as possible — in a language with laryngeal agreement, words with different feature specifications will consistently differ with respect to that feature on *all* relevant segments. Since words with multiple segmental differences are more perceptually distinct from one another than words which differ by single segments, maximising differences by means of assimilation aids in perception.

Gallagher supports this claim experimentally, presenting results of several discrimination studies in which subjects made same/different judgements about pairs of CVCV nonce words. Subjects were more accurate when the pairs differed with respect to the laryngeal features of *both* consonants than when they differed by the features of only one consonant. Kimper (2013) presents results of a similar study, using vocalic tongue root contrasts rather than consonantal laryngeal contrasts. Subjects were faster and more accurate when discriminating between words which differed by both vowels than words which differed by only a single vowel.

In both Gallagher's (2010) and Kimper's (2013) studies, however, *dissimilation* was also found to be advantageous — dissimilation is, after all, another way of ensuring that words consistently exhibit multiple differences for a particular feature contrast. This mirrors typological patterns for consonantal features, where dissimilatory co-occurrence restrictions are robustly attested — vowel dissimilation is, however, extremely rare[4]. Differences between long-distance consonantal and vocalic assimilation/dissimilation processes is beyond the scope of the present study — however, it is important to investigate whether assimilation presents a perceptual advantage that is distinct from the 'number of differences' effects found in the discrimination studies just above.

The experiments that follow aim to address the question of the perceptual advantage of harmony directly by using a modified identification task — subjects hear a nonsense word, and are subsequently asked to identify whether a particular vowel occurred in that nonsense word. It is hoped that this task, rather than involving a general assessment of distinctness or difference between two perceived stimuli, will probe more specific aspects of the mapping between an incoming signal and the corresponding abstract features or segmental categories.

3 Experiment 1

The primary aim of Experiment 1 is to investigate whether or not harmony confers a perceptual advantage, and (if so) the extent to which that advantage is locally restricted. If harmony is perceptually advantageous, subjects should be faster and more accurate in identifying whether or not a nonsense word contained a particular target vowel if the nonsense word contains another vowel which shares phonological features with that target. If this advantage is not strictly local, the effect should be found even when the target and 'supporting' vowel are non-adjacent. The net has been cast somewhat broadly here — there are a number of possible forms that a perceptual advantage could take, more than one of which would result in performance differences in the tasks in this (and subsequent) experiments;

---

[4]See Itô (1984) for a possible case of dissimilation in Ainu — however, it should be noted that an analysis as the pattern as dissimilation is somewhat dependent on theory-specific assumptions. Other cases, found primarily in Oceanic languages (see e.g. Blust 1996a,b) involve low vowels and could be plausibly analysed as either dissimilation or as sonority-related processes under alternating stress.

further discussion can be found in Section 8.2.

This second hypothesis — that harmony is perceptually advantageous even among non-adjacent segments — is of primary interest because the articulation-only approach to phonetic grounding, which entails commitment to strict locality, can offer no principled explanation of its robust typological attestation. However, there is also an opportunity to examine the potential perceptual basis for other typological asymmetries; in particular, the tendency for harmony to be facilitated among already-similar segments (parasitic harmony; see e.g. Cole and Trigo 1988; Cole and Kisseberth 1995; Kaun 1995). It was also possible to conduct some post-hoc investigation into the relationship between perception and the tendency for particular feature values (in this case [+ROUND]) to be phonologically active in harmony processes (Steriade, 1995b, and others). Both of these tendencies have alternative explanations, either in terms of articulation or in terms of formal representation, but may also be bolstered by perceptual factors. These predictions involve interactions — if similarity-sensitivity in harmony is perceptually grounded, the advantage of harmony should be greater among already-similar vowels; if the phonological activity of [+ROUND] is perceptually grounded, the advantage of harmony should be greater when the target is round.

The results from this experiment show that all four predictions are borne out — subjects were faster and more accurate in identifying target vowels in harmonic words, even when harmony was non-adjacent. This effect was greater with harmony among already-similar vowels, and greater for trials where the target was round.

3.1 Methods

3.1.1 Subjects

The data analysed for this study came from 33 native speakers of North American English (students in introductory linguistics courses at the University of Massachusetts Amherst who received course credit for their participation). Subjects who took part in the study were excluded from analysis if they reported speech or hearing disorders, were not native speakers, or registered no response on more than 10% of trials. In addition to the 33 subjects who met the criteria for inclusion, 7 subjects participated for course credit but were excluded (4

non-native speakers and 3 insufficient responders).[5]

Native speakers of English were chosen for this study — as in Gallagher's (2010) and Kimper's (2013) studies (discussed in Section 2.2 above) —precisely because there is no process of vowel harmony in English. Hypotheses about sources of phonetic grounding are by necessity independent of language-particular learned dependencies, and the effect should therefore be present in naïve listeners. This means that native language phonotactic legality is held constant across experimental conditions,[6] elimiting this potential confound.

3.1.2 Stimuli

Stimuli consisted of trisyllabic (CVCVCV) nonsense words, as well as target vowels uttered in isolation. The consonants {h g k} were chosen for the nonsense words because of their low coarticulatory effect; target vowels were given a [ʔ] onset. 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. Using Praat (**?**), these syllables were equalised for F_0 (221Hz), intensity (approx. 80dB), consonant duration (60ms) and vowel duration (250ms) before being spliced together to form the nonsense words. This equalisation meant that the stimuli sounded somewhat robotic; prior to the experiment, several native speakers listened to samples of the stimuli and reported that they nonetheless sounded like speech. There was no formal debriefing following the experimental task, but in informal debriefing subjects indicated that the stimuli sounded like speech.

For this experiment, only [i] and [u] were used as target vowels. The intention here is for subjects' judgements to reflect the feature dimension relevant to harmony — in other words, for errors to indicate a failure to identify the back/rounding feature that is spreading, in order to examine the effect that harmony has on the recognition of that feature. The vowels {i e a o u} were used to form the nonsense words; a list of the vowel combinations is given in Table 1. Every word contained the low back vowel [a] — this is meant to stand in

---

[5]After subjects with more than 10% null responses were excluded, the mean null response rate for subjects was 3.7%, with a standard deviation of 2.6%. Null responses were distributed more or less evenly across conditions.

[6]The stimuli used violate English phonotactics by lacking stress cues — which would normally include reduction of unstressed vowels — but this is true across all experimental conditions.

for non-contrastive transparent vowels found in many non-local harmony systems.[7] Among the remaining vowels, adjacency and agreement for colour[8] and height features were fully crossed.

In the local conditions, the colour-contrastive vowels were adjacent to each other (followed by [a]); in the non-local conditions, [a] intervened. The order in which the colour-contrastive vowels appeared, and whether they were of the same or different height, was counterbalanced. Of the words with vowels of the same height, only those with [i] or [u] were used (words with only [e] and [o] were excluded, because they contained neither of the target vowels). This resulted in 24 different vowel sequences; each appeared with every possible order of the onset consonants {h g k}, resulting in 144 items total.

	Disharmonic		Harmonic		
Local	u e a	e u a	i e a	e i a	Diff. Height
	i o a	o i a	u o a	o u a	
	i u a	u i a	i i a	u u a	Same Height
Non-Local	u a e	e a u	i a e	e a i	Diff. Height
	i a o	o a i	u a o	o a u	
	i a u	u a i	i a i	u a u	Same Height

Table 1: Vowel combinations used in nonce-word stimuli, Experiment 1.

3.1.3 Task

A nonsense word was presented auditorily, at a comfortable volume followed by a target vowel (after an ISI of 750ms). Subjects were asked to indicate by pressing a button whether or not the target vowel had been in the preceding word — they were told that they should answer quickly, and responses were cut off after 1500ms. Stimuli were presented in pseudorandom

---

[7]It's worth noting here that low vowels are typologically dispreferred as transparent vowels; Finley (2015) found better learning of harmony systems with transparent [i] than with transparent [a], and Kimper (2011) points out that languages with multiple transparent vowels exhibit in implicational relationship between low and high transparent vowels. Since the goal here is to determine whether the perceptual benefits of harmony extend to non-local arrangements, the use of [a] represents in some sense a useful worst-case-scenario for examining that prediction.

[8]Here and throughout this paper I use 'colour' as a way of grouping rounding and backness, since the two cannot be clearly distinguished in this task. While this term is borrowed from feature geometry, no specific theoretical commitments accompany its present use.
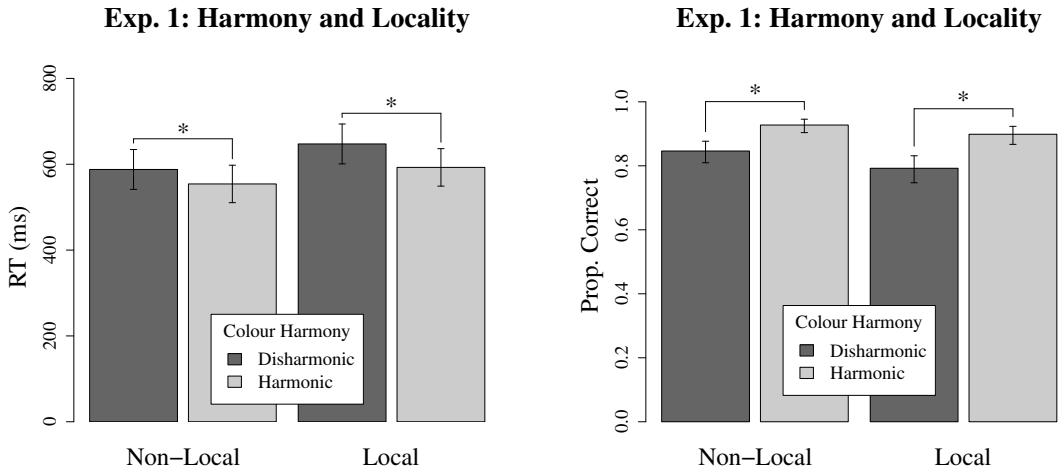
Figure 1: Accuracy and response time as a function of colour harmony and locality, Experiment 1. Error bars represent 95% Confidence Intervals.

order using Superlab 4, organised into two blocks of 288 trials each. Brief within-block breaks were offered every 72 trials, with a longer obligatory break between blocks. Subjects heard each item a total of four times, twice with [i] as the target and twice with [u] as the target.

Both the format of the task and the length of the ISI are designed to probe judgements related to category membership rather than acoustic detail (see e.g. Pisoni 1973; Gerrits and Schouten 2004). Because subjects listened to the entire word before discovering which target they must identify, the task is sensitive to the effects of vowels that follow the target in the word as well as those which precede it. It is assumed that subjects learned from the early trials that [i] and [u] were the two possible targets; responses are taken then to be indicative of identification of the features that mark the difference between the two, namely backness and rounding.

3.2 Results

Mixed effects models (with random intercept for subject, and random slope for subject $\times$ colour harmony) were fitted to the results by adding relevant factors — and their interactions — one at a time until there was no further improvement in model fit.[9] Linear models were

---

[9]Mixed effects were calculated using the `lme4` package (version 1.1.7, Bates et al. 2014) in R (version 3.1.2, R Core Team 2014). For linear mixed effects models, p-values were supplied by the `lmerTest` package

Exp. 1: Harmony and Height

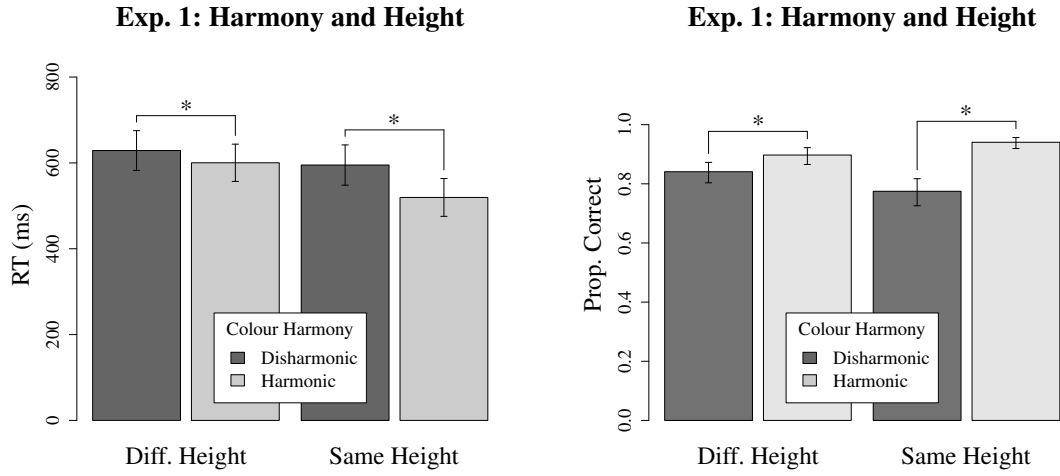**Exp. 1: Harmony and Height**

Figure 2: Accuracy and response time as a function of colour harmony and height agreement, Experiment 1. Error bars represent 95% Confidence Intervals.

Exp. 1: Harmony and Trial Type
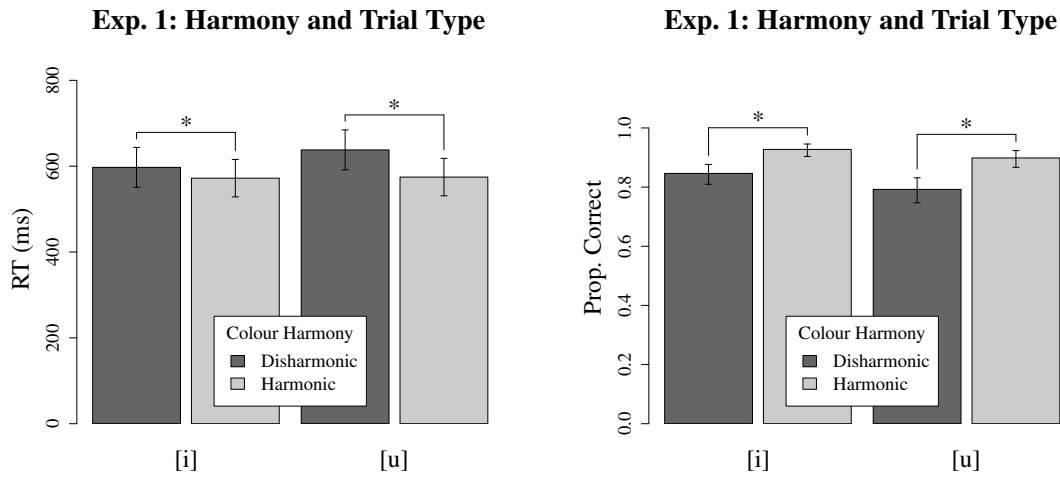

Exp. 1: Harmony and Trial Type

Figure 3: Accuracy and response time as a function of colour harmony and trial type, Experiment 1. Error bars represent 95% Confidence Intervals.

fitted for response time;[10] logit models were fitted on correct responses.[11] Deviation contrast coding was used throughout; models are given in Table 2. Where simple effects are reported, these are the result of re-running models with the relevant subsets of the data.[12]

Figure 1 shows that subjects were faster (p < 0.001) and more accurate (p < 0.001) with stimuli whose contrastive vowels were harmonic for colour features than those whose contrastive vowels were disharmonic — for example, [u] was better identified in *hugoka* than in *hugeka*. For response time, there was a significant interaction with locality (p < 0.01) — the difference between *hugoka* and *hugeka* was greater than the difference between *hugako* and *hugake* — but subjects were faster on harmonic items in both local (p < 0.001) and non-local (p < 0.001) conditions. For accuracy, there was no significant interaction between harmony and locality. On both measures, subjects performed better in non-local conditions across the board; this appears to be an artifact of timing in the task (see Sections 3.3 and 6.3 for further discussion).

Figure 2 shows the interaction between colour harmony and height agreement. For both response time (p < 0.001) and accuracy (p < 0.001), there was a significant interaction — the effect of harmony was greater when accompanied by height agreement (i.e. the difference between *huguka* and *hugika* was greater than the difference between *hugoka* and *hugeka*). For both response time and accuracy, the effect of harmony was still significant for both height-agreeing (p < 0.001) and height-disagreeing (p < 0.001) items. Additionally, subjects were faster overall (p < 0.001) on items where the colour-contrastive vowels agreed in height than those where they disagreed, though the main effect for height agreement was marginal (p = 0.096) for accuracy.

Finally, Figure 3 shows the interaction between harmony and trial type. The effect of

(version 2.0.11, Kuznetsova et al. 2014) and coefficients and standard errors used in graphs were retrieved from the fitted models using the `effects` package (version 3.0.1, Fox 2003). Here and throughout, non-significant effects which nonetheless improved model fit are reported in model tables, while non-significant effects which did not improve model fit are omitted.

[10]Here and throughout, response times are measured from the stimulus offset.

[11]Proportion correct was chosen as the relevant measure of accuracy rather than detection-theoretic measures like d′ (Macmillan and Creelman, 2004) because certain experimental conditions consisted of only 'signal' trials — in particular, same-height colour-disharmonic words necessarily contained both targets. Because of the absence of 'noise' trials in these conditions, calculation of d′ would require pooling of conditions in a way that would preclude the relevant comparisons.

[12]Reported p-values are uncorrected, but correcting for familywise error in planned simple effects does not result in a qualitative change in interpretation.

Response Time

	Estimate	Std. Err.	df	t value	Pr(> |t|)	
(Intercept)	587.419	22.566	32.000	26.031	<2e-16	***
Colour Harmony	26.045	3.268	36.000	7.969	1.95e-09	***
Locality	-25.006	2.361	18396.000	-10.589	<2e-16	***
Height Agreement	28.635	2.361	18396.000	12.126	<2e-16	***
Block	76.850	2.227	18396.000	34.501	<2e-16	***
Trial	-10.801	2.227	18396.000	-4.851	1.24e-06	***
Colour Harm. × Loc.	-7.637	2.361	18397.000	-3.234	0.00122	**
Colour Harm. × Height Agr.	-11.753	2.361	18396.000	-4.977	6.52e-07	***
Loc. × Height Agr.	1.512	2.361	18396.000	0.640	0.52213	
Colour Harm. × Trial	-9.548	2.227	18397.000	-4.288	1.81e-05	***
Colour Harm. × Loc. × Height Agr.	7.242	2.361	18396.000	3.067	0.00217	**

Accuracy (correct responses)

	Estimate	Std. Err.	z value	Pr(> |z|)	
(Intercept)	1.95551	0.13698	14.276	<2e-16	***
Colour Harmony	-0.50686	0.04364	-11.614	<2e-16	***
Height Agreement	-0.04070	0.02446	-1.664	0.096126	.
Locality	0.18309	0.02148	8.522	<2e-16	***
Block	-0.07464	0.02137	-3.493	0.000477	***
Trial	0.07941	0.02207	3.598	0.000320	***
Colour Harm. × Height Agr.	0.25561	0.02446	10.449	<2e-16	***
Colour Harm. × Trial	0.06965	0.02207	3.156	0.001600	**

Table 2: Mixed effects models for accuracy and response time, Experiment 1

harmony was greater in trials where [u] was the target vowel than for [i]-trials, for both response time ($p < 0.001$) and accuracy ($p < 0.001$). The effect of harmony remained significant in both [u]-trials ($p < 0.001$) and [i]-trials ($p < 0.001$) for both accuracy and response time. Trial type did not interact significantly with any factor besides colour harmony, for either response time or accuracy.

3.3   Discussion

The central prediction — that harmony confers a perceptual advantage — is supported by the results of Experiment 1: subjects did indeed show a performance advantage in words with harmony. Furthermore, this advantage was present even when harmonic vowels were non-adjacent, lending support to the claim that perceptual grounding covers long-distance as well as local instances of harmony. This provides support to the claim that long-distance harmony is a grounded phenomenon rather than a 'crazy rule', as it offers a potential source of phonetic grounding.

The interaction between harmony and locality — that is, the finding that the advantage

17

of harmony is somewhat diminished among non-adjacent vowels (which reached significance for response time but not for accuracy) — is consistent with a strong typological preference for local harmony. This finding, while consistent with expectations based on typology, is not crucial — since perceptual factors are not being pursued as the *sole* source of grounding for harmony, this result simply adds to the already existing articulatory factors motivating the existence of some form of locality preference. Additionally, the fact that subjects performed better overall — regardless of harmony — in non-local conditions should not be taken as indicative of the superiority of non-adjacency in perception. This is likely due to a difference in temporal proximity between the most recent informative cue and the presentation of the isolated target — by necessity, all words in the non-local condition contained a potentially-informative vowel in final position. Words in the local condition, however, did not — adding an additional latency (the duration of the final [a]) between the last informative cue and the presentation of the isolated target. As locality is of interest here not *a priori* but specifically in relation to perceptual benefits conferred by harmony, this does not interfere with the main result of this study.

Similarly, the finding that harmony was more advantageous among already similar vowels — those which agreed in height — provides an additional source of explanation for phenomena attributed to Kaun's (1995) *gestural uniformity* constraint. It should be noted that distinguishing between the role of *similarity* and the role of *complete identity* is difficult, and in an experimental design with a relatively scarcely-populated vowel inventory the two are conflated — vowels which agree in height and are colour-harmonic are in fact identical. In this case, then, it should be no surprise that harmony is advantageous in the height-agreement condition — it should of course be easier to detect [u] in a word which contains two instances of that very vowel. It is the height disagreeing conditions that are more informative here: that the task of identifying [u] was easier in a word which also contained [o] suggests that it's not merely the frequency of the target that aids in identification — sub-components of the target (either phonological features or more general phonetic properties) are relevant in the decision that subjects are making.

Finally, the finding that harmony was more advantageous for target [u] than target [i] may provide a source of grounding for the tendency for [+ROUND] to be phonologically

active — indeed, it has been claimed that [–ROUND] is never active, and this has been built into the architecture of phonological feature systems (see e.g. Steriade 1995b). It would be unwise to make too much of this experimental effect — while significant, it is quite small, and (foreshadowing a bit) not consistently replicated. Additionally, because backness and rounding cannot be distinguished here, the effect cannot be definitively attributed to [+ROUND]. It's also possible that [i] and [e] are more easily confused than [o] and [u], which may or may not bear a relationship to phonological activity of feature values. Nonetheless, it holds out a curious clue for future inquiry.

4   Experiment 2

Experiment 2 was designed to further investigate the effect of distance on the perceptual advantages of harmony. Experiment 1 found that non-adjacent segments showed a slightly diminished but still robust effect of harmony, suggesting both that long-distance harmony may be phonetically grounded and that a preference for locality is supported by perceptual as well as articulatory asymmetries.

The previous experiment only compared vowels in adjacent and non-adjacent syllables, but natural languages distinguish at least one additional degree of distance. In Hungarian, for example, harmony across a single [i] as in (4) is obligatory; however, if multiple non-undergoers stand between trigger and target, harmony becomes variable (see Hayes and Londe 2006 and Hayes et al. 2009 for further discussion).

This distinction is difficult to explain in articulatory terms — the number of gestural transitions is the same, regardless of the number of non-undergoers involved, so these situations should be equivalent. Can a perceptual account of phonetic grounding in harmony provide an explanation? Experiment 2 largely replicates the previous design, but with the inclusion of an additional non-local condition with two intervenors, to investigate whether increasing degrees of distance further diminishes the advantage of harmony.

The results suggest that this is in fact the case — the main effect of harmony was replicated robustly, and while the interaction was not significant at one degree of distance, it was significant with two intervenors, suggesting that gradient distance effects in harmony

may be grounded in the same perceptual factors that motivate long-distance harmony.

4.1 Methods

Data for Experiment 2 come from 38 native speakers of North American English (students in introductory linguistics courses at the University of California, Santa Cruz, who received course credit for their participation). As in the previous study, subjects were excluded from analysis if they reported speech or hearing disorders, were not native speakers, or registered no response on more than 10% of trials. In addition to the 38 subjects who met the criteria for inclusion, 28 subjects participated for course credit but were excluded from analysis (7 reported speech or hearing disorders, 17 non-native speakers and 4 insufficient responders).

Stimuli consisted of quadrisyllabic (CVCVCVCV) nonsense words, using the same vowel inventory as in Experiment 1. As before, two of the vowels in each word contrasted for colour features, and these either agreed or disagreed with each other. These were either adjacent to each other (preceded or followed by two [a] syllables), separated by one [a] syllable (preceded or followed by another), or separated by two [a] syllables. Nonsense words were formed by concatenating the same isolated syllables from Experiment 1, with the addition of [sk] as a possible syllable onset (these were recorded and normalised at the same time and in the same way as the other syllables). Likewise, the target vowels [i] and [u] were the same as those in Experiment 1.

Vowel combinations were combined with randomly selected subsets of the possible consonant combinations to create a stimulus set with equal numbers of items across each level of distance, height agreement, and harmony. This resulted in a total of 288 items.

The task and presentation of stimuli were the same as in Experiment 1 — subjects heard a nonsense word followed by a target vowel, and were asked to indicate whether or not the target had been in the preceding word. Subjects heard each nonsense word twice, once with [i] and once with [u]. Trials were distributed across two blocks of 288 trials each, with brief within-block breaks every 72 trials and a longer break between blocks.

Exp2: Harmony by Height and Distance
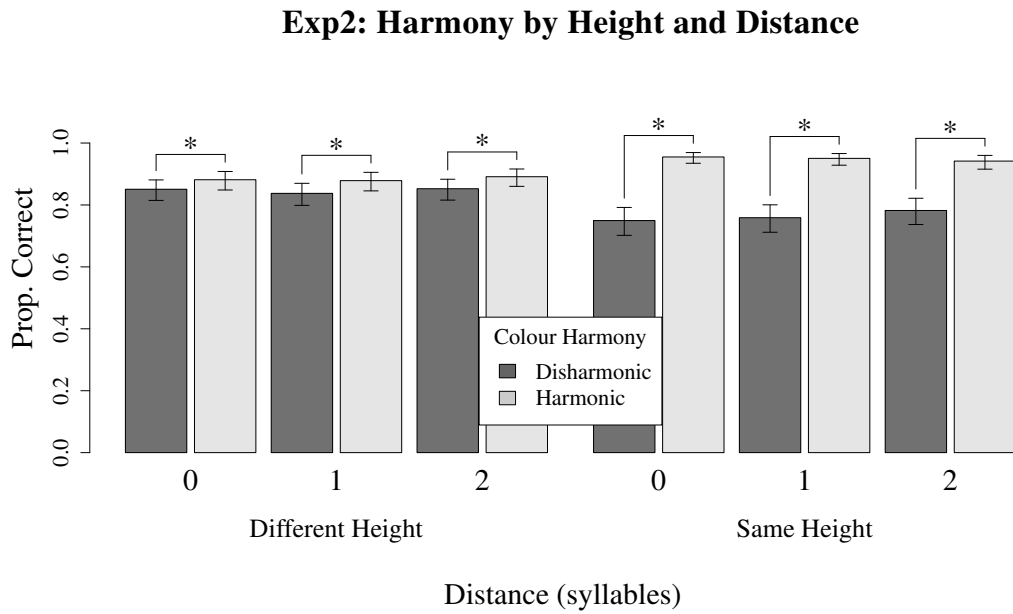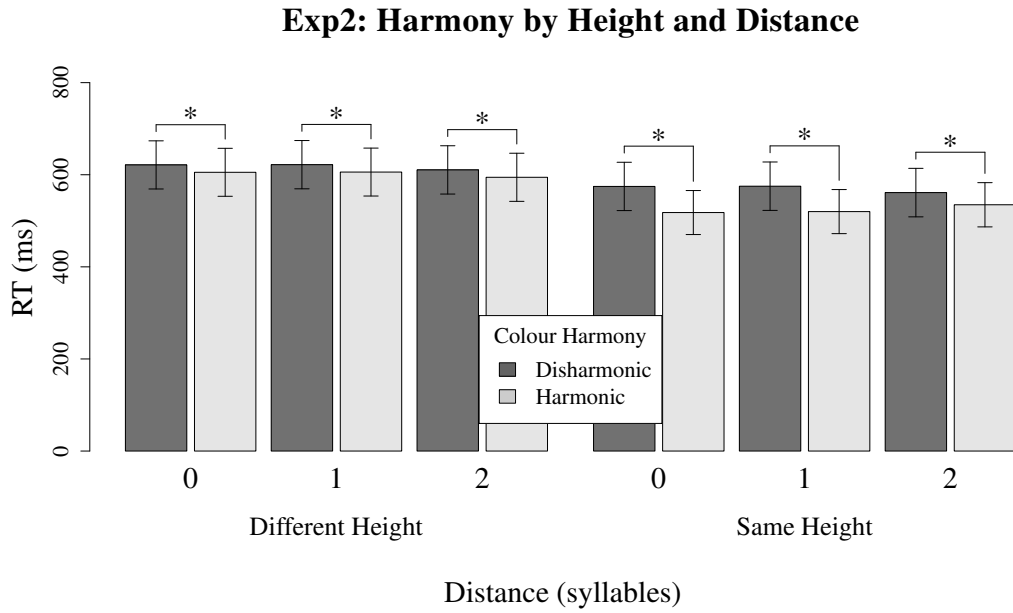

Exp2: Harmony by Height and Distance

Figure 4: Accuracy and response time as a function of colour harmony, height agreement and locality, Experiment 2. Error bars represent 95% confidence intervals.

	Harmonic		Disharmonic		
	aaii	aauu	aaiu	aaui	Same Height
	iiaa	uuaa	iuaa	uiaa	
Distance: 0	aaie	aaei	aaio	aaoi	
	ieaa	eiaa	ioaa	oiaa	Diff. Height
	aauo	aaou	aaeu	aaue	
	uoaa	ouaa	euaa	ueaa	
	aiai	auau	aiau	auai	Same Height
	iaia	uaua	iaua	uaia	
Distance: 1	aiae	aeai	aiao	aoai	
	iaea	eaia	iaoa	oaia	Diff. Height
	auao	aoau	aeau	auae	
	uaoa	oaua	eaua	uaea	
	iaai	uaau	iaau	uaai	Same Height
Distance: 2	iaae	eaai	iaao	oaai	Diff. Height
	uaao	oaau	eaau	uaae	

Table 3: Vowel combinations used in nonce-word stimuli, Experiment 2. Distance reflects the number of syllables separating colour-contrastive vowels.

4.2   Results

As in Experiment 1, mixed effects models were fitted for accuracy and response time (with random intercept for subject, and random slope for subject × colour harmony); here Helmert coding was used for distance, and deviation coding was used for all other factors. The full models can be seen in Figure 4. The effects seen in Experiment 1 were, broadly speaking, replicated — subjects were faster ($p < 0.001$) and more accurate ($p < 0.001$) with stimuli whose contrastive vowels were harmonic for colour features; there was a significant interaction with height agreement for both accuracy ($p < 0.001$) and response time ($p < 0.001$), with the effect of harmony greater among vowels with height agreement. Trial type also interacted with harmony for both accuracy ($p < 0.001$) and response time ($p < 0.01$), with [u]-trials showing a greater effect of harmony.

The three-way interaction between height agreement, distance, and colour harmony was significant for accuracy ($p < 0.05$) and response time ($p < 0.05$) at a distance of two syllables,

but not at a distance of one syllable. For items where colour-contrastive vowels disagreed in height (e.g. *skuhogaka* vs *skuhegaka*) colour harmony had a significant main effect for both accuracy ($p < 0.001$) and response time ($p < 0.01$) but no interactions were significant for either measure. For items with height agreement, there was a significant interaction between distance and colour harmony at a distance of two syllables for both accuracy ($p < 0.01$) and response time ($p < 0.05$).[13] The effect of harmony was diminished at this distance — the difference between *skuhugaka* and *skuhigaka* was greater than the difference between *skuhagaku* and *skuhagaki* — but the effect remains significant for both accuracy ($p < 0.001$) and response time ($p < 0.01$). Unlike in Experiment 1, the interaction does not reach significance at a distance of one syllable, but numerical trends are in the expected direction.

4.3   Discussion

The main effect from Experiment 1, that harmony confers a perceptual advantage, was replicated in the current experiment. As before, too, this advantage obtained regardless of distance — the effect of harmony was statistically significant, even in the least conducive set of conditions.

The prediction at hand in this particular version of the experiment — that the effect of harmony diminished with increasing distance — was borne out, with somewhat of a caveat: the effect was only detectable among identical vowels. As before, harmony had a greater effect in the same-height conditions than in different-height conditions; the small size of the main effect in the different-height conditions may be responsible for the failure of the interaction to reach statistical significance.

Perhaps surprisingly, there was not a significant diminution of the effect of harmony at a distance of one syllable — this is not entirely unexpected, given the mixed results from that interaction from the previous experiment. However, there was a significant diminution of the effect at a distance of two syllables; harmony was less advantageous at this distance. This is consistent with distance effects found in languages like Hungarian.

---

[13]Correction for familywise error in simple effects jeopardises the result for response time, but not for accuracy.

Response Time

	Estimate	Std. Err.	df	t value	Pr(> |t|)	
(Intercept)	587.4117	25.1239	37.0000	23.38	<2e-16	***
Colour Harmony	15.0326	2.6497	37.0000	5.67	1.7e-06	***
Distance (1)	-0.0456	2.5827	21291.0000	-0.02	0.98590	
Distance (2)	-1.5514	1.7006	21290.0000	-0.91	0.36162	
Height Agreement	31.8057	2.1067	21290.0000	15.10	<2e-16	***
Right Edge	23.7885	2.6033	21290.0000	9.14	<2e-16	***
Trial	-17.0328	2.1034	21290.0000	-8.10	4.4e-16	***
Colour Harm× Dist. (1)	3.4645	2.5852	21291.0000	1.34	0.18021	
Colour Harm. × Dist. (2)	-1.5102	1.4903	21291.0000	-1.01	0.31092	
Colour Harm. ×Height Agr.	-7.9154	2.1093	21291.0000	-3.75	0.00018	***
Dist. (1) × Height Agr.	-0.9063	2.5806	21291.0000	-0.35	0.72544	
Dist. (2) × Height Agr.	0.6801	1.4894	21290.0000	0.46	0.64796	
Colour Harm. × Trial	-5.5900	2.1034	21291.0000	-2.66	0.00788	**
Colour Harm. × Dist. (1) × Height Agr.	3.8204	2.5854	21291.0000	1.48	0.13950	
Colour Harm. × Dist. (2) × Height Agr.	3.4266	1.4903	21290.0000	2.30	0.02150	*

Accuracy (correct responses)

	Estimate	Std. Err.	z value	Pr(> |z|)	
(Intercept)	1.85018	0.12114	15.27	<2e-16	***
Colour Harmony	-0.50210	0.03909	-12.85	<2e-16	***
Distance (1)	-0.02111	0.02520	-0.84	0.40222	
Distance (2)	0.01018	0.01726	0.59	0.55544	
Height Agreement	-0.04401	0.02100	-2.10	0.03613	*
Right Edge	-0.21626	0.02375	-9.11	<2e-16	***
Trial	0.07797	0.02060	3.79	0.00015	***
Colour Harm. × Dist. (1)	0.00548	0.02522	0.22	0.82787	
Colour Harm. × Dist. (2)	0.02514	0.01508	1.67	0.09556	.
Colour Harm. × Height Agr.	0.31072	0.02105	14.76	<2e-16	***
Dist. (1) × Height Agr.	-0.00824	0.02518	-0.33	0.74342	
Dist. (2) × Height Agr.	0.00369	0.01507	0.24	0.80649	
Colour Harm. × Trial	0.08334	0.02060	4.05	0.000052	***
Colour Harm. × Dist. (1) × Height Agr.	-0.03002	0.02522	-1.19	0.23389	
Colour Harm. × Dist. (2) × Height Agr.	-0.03680	0.01509	-2.44	0.01472	*

Table 4: Mixed effects models for accuracy and response time, Experiment 2

This result, however, should be interpreted with some caution. The size of the interaction was small, and detectable only among identical vowels. This offers the potential for a perceptual explanation for distance effects, but such an explanation is not the only plausible account — for example, it may very well be the case that biases in learning systematically disadvantage patterns that require associations across increasing distances (see Moreton 2012; Moreton and Pater 2012 for more on structural biases in learning).

5   Experiment 3

The previous experiments have used only high vowels [i] and [u] as targets; in this version, additional blocks with mid-vowel [e] and [o] targets are included. This means that, in addition to investigating the effect of similarity (or identity) on the advantage posed by harmony, it is possible to look more closely at the impact of height itself. In particular, the typology of rounding harmony suggests that mid vowels tend to be preferred as triggers over high vowels; Kaun (1995) argues that this is attributable to the fact that low and mid vowels are less articulatorily and perceptually rounded than their high counterparts (Linker, 1982; Terbeek, 1977). If this is the case, we expect that subjects should perform worse overall when identifying the front/back contrast in mid-vowel blocks than in high-vowel blocks.

In addition, inclusion of high-only stimuli in mid-target blocks (and vice versa) — e.g. *hogeka* in blocks where [i] and [u] are the targets — can help shed light on the extent to which identification is sensitive to either featural or phonetic sub-components of the target segments; this will be further explored in a discussion of error types across all experiments in Section 7.

5.1   Methods

Data for Experiment 3 come from 36 native speakers of North American English (students in introductory linguistics courses at the University of California, Santa Cruz, who received course credit for their participation). As in the previous two studies, subjects were excluded from analysis if they reported speech or hearing disorders, were not native speakers, or registered no response on more than 10% of trials. In addition to the 36 subjects who

met the criteria for inclusion, 30 subjects participated for course credit but were excluded from analysis (7 reported speech or hearing disorders, 17 non-native speakers, and 6 non-responders).

Stimuli consisted of trisyllabic (CVCVCV) nonsense words, the items from Experiment 1 plus several additional vowel sequences. Local conditions were included where the colour-contrastive vowels appeared *following* the neutral [a] (cf. Experiment 1, where they always preceded [a]).[14] The full list of vowel combinations used can be seen in Table 5. Each combination was paired with four different consonant sequences, for a total of 136 items. In addition to target vowels [i] and [u], [e] and [o] were also used. All material used to create the additional stimuli was recorded at the same time and normalised in the same way as the materials for the previous experiments.

	Disharmonic		Harmonic		
Local	u e a	e u a	i e a	e i a	Diff. Height
	i o a	o i a	u o a	o u a	
	i u a	u i a	i i a	u u a	Same Height
	e o a	o e a	e e a	o o a	
	a e o	a o e	a e e	a o o	
Non-Local	u a e	e a u	i a e	e a i	Diff. Height
	i a o	o a i	u a o	o a u	
	i a u	u a i	i a i	u a u	Same Height
	e a o	o a e	e a e	o a o	

Table 5: Vowel combinations used in nonce-word stimuli, Experiment 3.

The task and presentation of the stimuli were the same as in the previous experiments — subjects heard a nonsense word followed by a target vowel, and were asked to indicate whether or not the target had been in the preceding word. Subjects heard each nonsense word a total four times, once with each target vowel. Trials were organised into two blocks of 272 trials each, with brief within-block breaks every 72 trials and a longer break between blocks. Within each block, the height of the target vowels was consistent — [i] and [u] shared

---

[14]This was to investigate the role of latency artifacts in driving the main effect of locality, but will not be discussed here; Experiment 4 provides a more illustrative demonstration of the relevant artifacts

a block, with [e] and [o] in a separate block, in order to ensure that the judgement made for each trial primarily represents information about the colour features of the stimulus. The order of the blocks was pseudorandomised for each subject.

5.2   Results

For analysis, data from Experiment 3 was divided between block-matched and block-mismatched conditions. Stimuli were block-mismatched if they contained none of the target vowels relevant to a particular block — e.g. stimuli with all high vowels in a mid-vowel block, and vice versa. As mentioned above, this was done to focus more directly on the role of colour harmony on detection of the feature in question — a 'no' response on a trial like *hogeka ... u* reflects a decision about the height of the vowels in the stimulus and may or may not reflect the detection of the relevant BACK/ROUND feature value. Results for block-matched data are presented here; see Section 7 for a discussion of the block-mismatched data.

As in the previous experiments, mixed effects models were fitted for accuracy and response time (with random intercept for subject, and random slope for subject × colour harmony); deviation contrast coding was used throughout. Full models can be seen in Table 6. Once again the main effects seen in the preceding experiments were broadly replicated. There was a robust main effect of colour harmony for both accuracy ($p < 0.001$) and response time ($p < 0.001$), and significant interactions with height agreement for both accuracy ($p < 0.001$) and response time ($p < 0.05$). While the effect of harmony was diminished among stimuli whose colour-contrastive vowels disagreed in height, the effect remained significant for both accuracy ($p < 0.01$) and response time ($p < 0.001$). There were no interactions with locality, which departs from the findings of Experiment 1 but is consistent with the results from Experiment 2 (at a distance of one syllable). Unlike in previous experiments, there was no interaction between the colour of the trial ([i] vs. [u] or [e] vs. [o]) and the effect of harmony.

As can be seen in Figure 5, the height of the trials in a block ([i]/[u] blocks vs. [e]/[o] blocks) exhibited a main effect for accuracy ($p < 0.001$) and response time ($p < 0.01$), with subjects performing better overall on high stimuli. There was also significant interaction between block height and colour harmony for both accuracy ($p < 0.001$) and response time ($p < 0.05$), with the effect of harmony greater in high blocks than in mid blocks. An

27

Exp. 3: Harmony by Block Type   **Exp. 3: Harmony by Block Type**
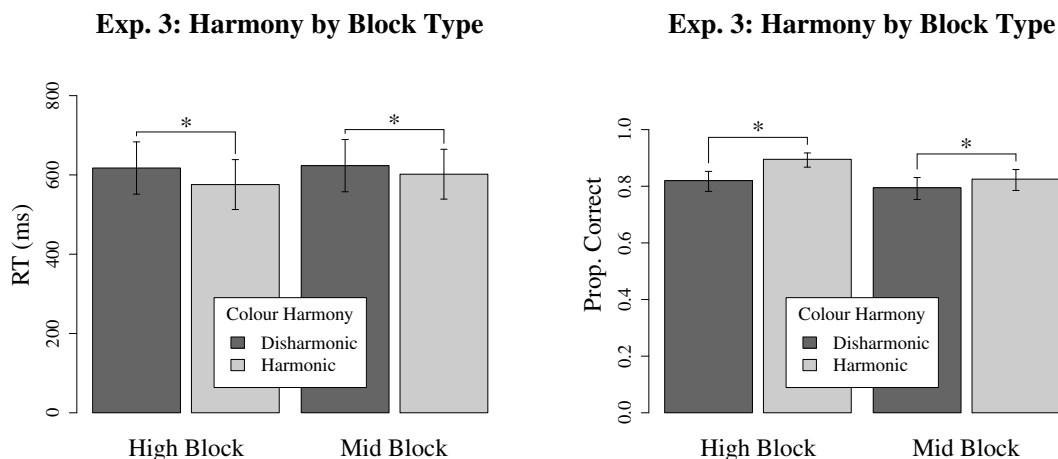

Figure 5: Accuracy and response time as a function of colour harmony, height agreement and block type, Experiment 3. Error bars represent 95% confidence intervals.

analysis of just the disharmonic items suggests that the main effect of block type is not driven solely by differences in the benefits of harmony; even with disharmonic words, subjects were somewhat faster ($p < 0.05$) and more accurate ($p < 0.05$) in high blocks than in mid blocks. For accuracy there was a three-way interaction with height agreement ($p < 0.001$) — the increase in the advantage of harmony under height agreement was exaggerated in high blocks, though this interaction was not significant for response time.

5.3   Discussion

The results of this experiment again replicate the main result of experiments 1 and 2, which is that harmony confers a perceptual advantage. In addition, findings from the previous experiments — that this advantage still obtains among non-adjacent segments, and is greater among vowels which share other features — were also replicated.

This experiment also allows more direct comparison of harmony involving vowels of different height. The first finding here is that subjects performed better across the board in blocks where the targets were high vowels. This provides support for Kaun's (1995) explanation of mid vowels' behaviour as preferential triggers of harmony — subjects poorer performance in identifying back/round contrasts with mid vowels in this task bolsters the

Response Time

	Estimate	Std. Err.	df	t value	Pr(> |t|)	
(Intercept)	595.523	32.522	35	18.311	< 2e-16	***
(Intercept)	596.01	32.56	35.00	18.31	<2e-16	***
Colour Harmony	17.38	3.00	39.00	5.79	9.8e-07	***
Height Agreement	31.48	2.61	13916.00	12.08	<2e-16	***
Locality	-18.67	2.50	13915.00	-7.47	8.3e-14	***
Block	46.43	2.56	13940.00	18.16	<2e-16	***
Block Height	-8.05	2.57	13917.00	-3.14	0.0017	**
Colour Harm. × Height Agr.	-5.49	2.61	13916.00	-2.11	0.0351	*
Colour Harm. × Block Height	5.04	2.50	13924.00	2.02	0.0437	*

Accuracy (correct responses)

	Estimate	Std. Err.	z value	Pr(> |z|)	
(Intercept)	1.6346	0.1187	13.77	<2e-16	***
Colour Harmony	-0.2532	0.0344	-7.36	1.8e-13	***
Height Agreement	0.0551	0.0247	2.23	0.02559	*
Locality	0.1996	0.0231	8.64	<2e-16	***
Block	0.0356	0.0232	1.54	0.12464	
Block Height	0.2176	0.0248	8.77	<2e-16	***
Colour Harm. × Height Agr.	0.1632	0.0247	6.60	4.1e-11	***
Colour Harm. × Block	-0.0472	0.0231	-2.04	0.04123	*
Colour Harm. × Block Height	-0.1338	0.0248	-5.39	6.9e-08	***
Height Agr. × Block Height	-0.1033	0.0247	-4.18	2.9e-05	***
Colour Harm. × Height Agr. × Block Height	0.0935	0.0247	3.78	0.00015	***

Table 6: Mixed effects models for accuracy and response time, block-matched trials, Experiment 3

claim that these vowels are in most need of the boost in perceptual salience that comes with harmony. The finding that the advantage of harmony is greater for high vowels than for mid vowels, while it fails to provide any additional support for this line of reasoning, does not necessarily contradict it — while high vowels may stand to gain more from harmony, they have less need for it.

6 Experiment 4

In the previous three experiments, subjects were tasked with recognising and recalling the target vowel regardless of its location in the word. In the course of real-world speech perception, however, phoneme/feature recognition is generally localised to specific linear positions. For example, (Astheimer and Sanders, 2009, 2011) provide evidence from Event-Related Potential (ERP) studies showing that listeners modulate auditory attention over the time-course of speech perception, and that attention is increased in contexts where phonological

content is unpredictable.

As discussed in Section 2.2, one of the possible advantages of vowel harmony is that it renders contrasts predictable across a variety of positions — this in turn reduces the burden on listener's auditory attention. However, predictability is not a unique feature of harmony: dissimilation provides an equal degree of predictability with no increase in computational complexity. Furthermore, there is no difference in predictability between transparent harmony and articulatorily-preferred opaque harmony.[15] The previous studies suggest that perception of a target vowel is facilitated by the presence of another vowel bearing a relevant feature value; does this effect still obtain in situations more closely resembling the type of attentional modulation required of real-world speech perception?

Experiment 4 extends the previous studies to include localised attention, expanding the set of tasks to include attending specifically to initial and final syllables. Initial-syllable identification in particular bears a closer resemblance to the task of speech perception, since initial syllables are psycholinguistically prominent, selectively attended, and license a greater degree of contrast than non-initial syllables (Beckman, 1998) — even in languages where vowel harmony is anticipatory, initial syllables serve as the listener's first cue to the spreading feature value. If harmony represents a true perceptual advantage, this effect should be present in the more realistic task of identifying an initial-syllable vowel.

6.1 Methods

Data for Experiment 4 come from 25 native speakers of British English (students in introductory linguistics courses at the University of Manchester who received course credit for their participation). As in the previous studies, subjects were excluded from analysis if they reported speech or hearing disorders, were not native speakers, or registered no response on more than 10% of trials. In addition to the subjects who met the criteria for inclusion, and additional 16 subjects participated for course credit but were excluded from analysis (10 non-native speakers, 1 insufficient responder, and 5 due to data loss resulting from

---

[15]Equality of predictability between transparency and opacity depends on opaque segments propagating their own feature value, rather than permitting a contrast in subsequent vowels; fortunately, this appears to be the norm for opaque harmony systems.

equipment failure). Subjects participated in a sound-attenuated booth in the University of Manchester Psycholinguistics Laboratory; stimuli were presented at a comfortable volume through circumaural headphones, using E-Prime 2.0 Professional (Schneider et al., 2002).

Stimuli were the same as in Experiment 3. As before, subjects heard a nonsense word followed (after an ISI of 750ms) by a vowel in isolation; as in experiments 1 and 2, only [i] and [u] served as the target vowels. Trials were organised into blocks by task, with brief within-block breaks and a longer between-blocks break as in the previous studies. In *Anywhere* blocks, subjects were instructed to respond 'yes' if they heard the target vowel anywhere in the nonsense word; in *First Syllable* blocks, subjects were instructed to respond 'yes' if they heard the target vowel in the first syllable of the nonsense word; in *Last Syllable* blocks, subjects were instructed to respond 'yes' if they heard the target vowel in the last syllable of the nonsense word. The *First Syllable* condition is of primary concern; the others are included for reference. Each subject received two blocks; each was selected pseudorandomly from the set of three, sampled with replacement. Each task was distributed equally between Block 1 and in Block 2, though 'Last Syllable' blocks appeared more often than the other two.

6.2   Results

As with Experiment 3, data for analysis was divided into block-matched and block-mismatched trials. For *Anywhere* blocks, trials were considered block-matched if they contained at least one of the target vowels ([i] and [u]); for *First Syllable* blocks, trials were considered block-matched if they contained one of the targets in initial position, and for *Last Syllable* blocks, trials were considered block-matched if they contained one of the targets in final position. Results for block-matched data are presented here, and block-mismatched data are discussed in Section 7.

As in the previous experiments, mixed effects models (with random intercept for subject, and random slope for subject $\times$ colour harmony); deviation contrast coding was used throughout. Full models can be seen in Table 7. As in the previous experiments, there was an overall main effect of colour harmony for both accuracy ($p < 0.001$) and response time ($p < 0.001$), and a significant interaction between colour harmony and height agreement for
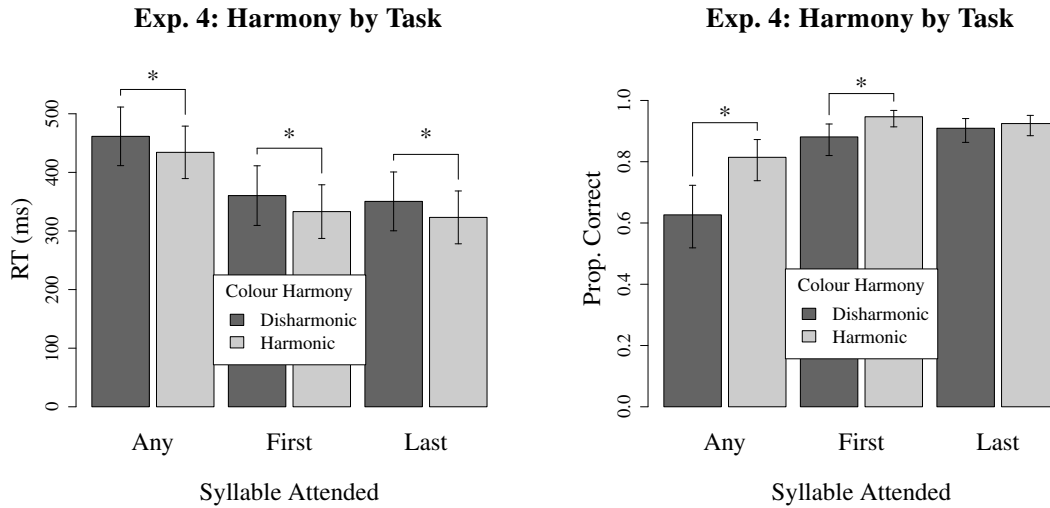
31

Figure 6: Accuracy and response time as a function of colour harmony and task, Experiment 4. Error bars represent 95% Confidence Intervals.

both accuracy ($p < 0.001$) and response time ($p < 0.05$), with the effect of colour harmony greater for same-height than different-height items. For response time, this was the only statistically significant interaction term — for accuracy, there were also significant interactions for task ($p < 0.05$) and block ($p < 0.05$) and a marginal interaction for trial ($p = 0.0696$). There was also significant main effect of task, for both accuracy ($p < 0.001$) and response time ($p < 0.001$).

Figure 6 shows the effect of colour harmony across the three tasks. The *Any Syllable* blocks repeat the conditions of the previous experiments, and replicate the main effect. In the *First Syllable* blocks, the effect remains significant for both accuracy ($p < 0.01$) and response time ($p < 0.01$). In the *Last Syllable* blocks, participants were at or close to ceiling across the board; the effect of colour harmony is significant for response time ($p < 0.05$), but not for accuracy.

Figure 7 takes a closer look at the *First Syllable* blocks, showing the effect of harmony for both same-height and different-height items. For accuracy, we see the combined action of the interactions between colour harmony and height agreement and between colour harmony and task: different-height items in the *First Syllable* task are doubly disadvantaged. Regardless, the effect of colour harmony remains significant even in these items ($p < 0.05$).

Exp. 4: Harmony by Height, First Syllable Attended

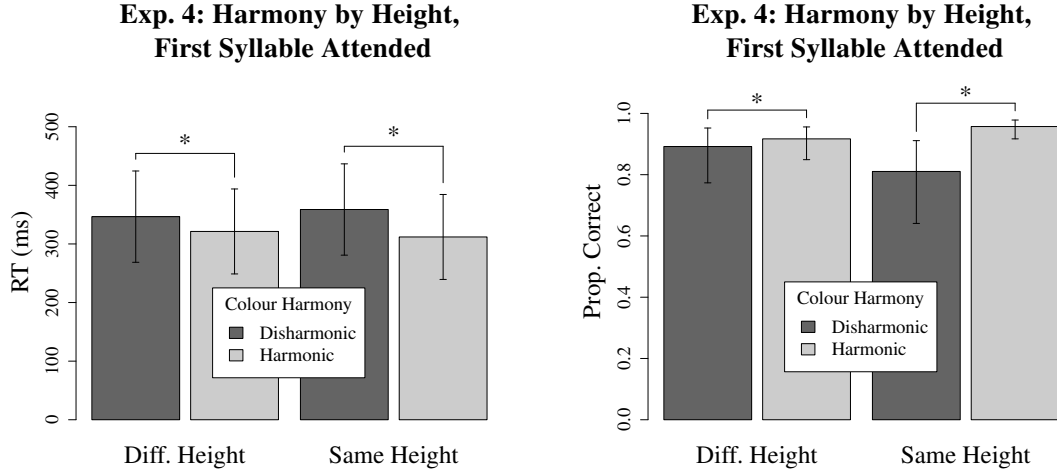**Exp. 4: Harmony by Height, First Syllable Attended**

Figure 7: Accuracy and response time as a function of colour harmony, block-matched trials only, first syllable attended. Error bars represent 95% Confidence Intervals.

Response Time

	Estimate	Std. Err.	df	t value	Pr(> |t|)	
(Intercept)	378.39	23.41	24.00	16.17	2.0e-14	***
Colour Harmony	14.73	3.14	25.00	4.70	8.2e-05	***
Height Agreement	4.93	2.84	8788.00	1.74	0.082	.
Locality	-7.09	3.26	8788.00	-2.17	0.030	*
Block	60.83	2.89	8805.00	21.08	<2e-16	***
Task (First)	70.71	5.66	7876.00	12.50	<2e-16	***
Task (Last)	-30.43	6.55	7538.00	-4.64	3.5e-06	***
Trial	-5.47	2.76	8788.00	-1.98	0.048	*
Colour Harm. × Height Agr.	-7.09	2.80	8394.00	-2.54	0.011	*

Accuracy

	Estimate	Std. Err.	z value	Pr(> |z|)	
(Intercept)	1.8892	0.2027	9.32	<2e-16	***
Colour Harmony	-0.3798	0.0699	-5.43	5.5e-08	***
Height Agreement	0.0587	0.0308	1.90	0.0570	.
Locality	0.1707	0.0321	5.31	1.1e-07	***
Task (First)	-0.9495	0.0767	-12.38	<2e-16	***
Task (Last)	0.4918	0.0882	5.57	2.5e-08	***
Block	-0.2661	0.0351	-7.58	3.4e-14	***
Trial	0.0805	0.0295	2.73	0.0063	**
Colour Harm. × Height Agr.	0.2673	0.0308	8.67	<2e-16	***
Colour Harm. × Task (First)	-0.1409	0.0674	-2.09	0.0364	*
Colour Harm. × Task (Last)	-0.0981	0.0770	-1.27	0.2030	
Colour Harm. × Block	-0.0849	0.0344	-2.47	0.0135	*
Colour Harm. × Trial	-0.0534	0.0295	-1.81	0.0696	.

Table 7: Mixed effects models for accuracy and response time, block-matched trials, Experiment 4.

6.3 Discussion

This experiment again robustly replicated the main result of the previous three. This experiment also provides an opportunity to examine some of the effects of the task. First, we see that speed and accuracy are both greater in the *First Syllable* and *Last Syllable* conditions — this is not surprising, since the subjects' attention can be directed at a single point in the stimulus. In the *Last Syllable* condition, this resulted in a ceiling effect for accuracy — this is consistent with the performance advantage in Experiments 1-3 in conditions where the latency between the most recent informative cue and the isolated target vowel was reduced. In particular, because the length of the stimuli was consistent, subjects are free to disregard material preceding the attended syllable, and may adopt a strategy similar to that of an AX discrimination study. It is worth noting, however, that while accuracy was at ceiling across the board, there was still a significant main effect on response time, suggesting that preceding material may not have been entirely disregarded.

The *First Syllable* condition is the one which most closely resembles the real-world task of speech perception. In this condition, we see the main effect of Experiments 1–3 replicated: subjects were faster and more accurate on items with colour harmony. As in the previous studies, this effect was more pronounced for items with height agreement (when colour harmony means total identity) — but the effect is still present even when the colour-contrastive vowels disagree in height. In other words, perception of an initial [u] is facilitated by a subsequent back/round vowel, whether that is another [u] or a (non-identical) [o].

7 Error Analysis

The results in Experiments 1–4 raise questions about the types of errors subjects are making, and their distribution across conditions; this section provides a closer look at those errors, beginning with the block-mismatched data from Experiments 3 and 4 in Section 7.1 and continuing on to a comparison of error types in Section 7.2 and a brief discussion of block order in Section 7.3.
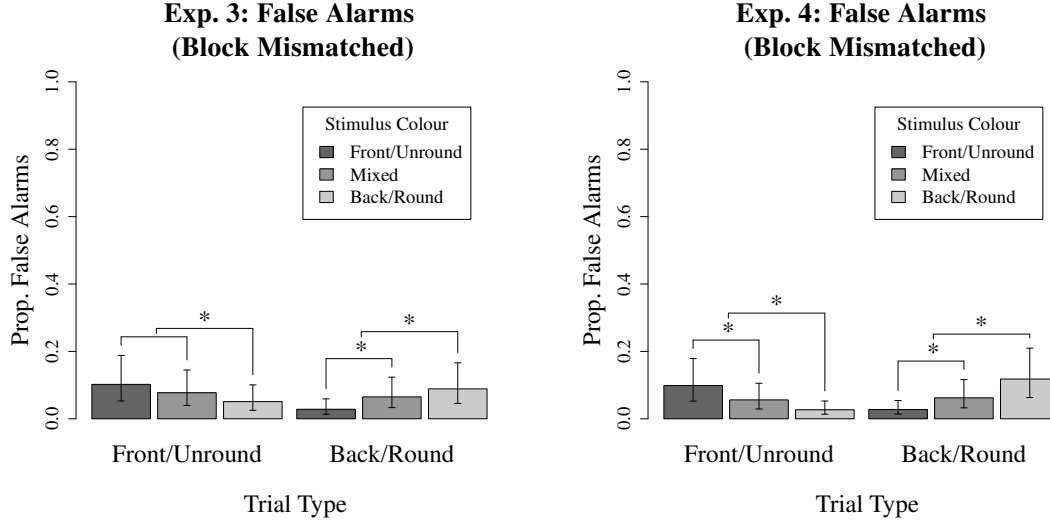
Figure 8: False alarms as a function of stimulus colour and trial type, block-mismatched items only, Experiments 3 and 4. Error bars represent 95% Confidence Intervals.

7.1 Block-Mismatched Trials

Recall that trials were classified as block-mismatched if they contained neither of the possible target vowels (Experiment 3 and the *Any Syllable* condition in Experiment 4) in the attended position (the *First Syllable* and *Last Syllable* conditions in Experiment 4). Mixed effects models were fitted on these data, with random intercept for subject and random slopes for subject; helmert contrast coding was used for the three levels of stimulus colour, and deviation coding was used elsewhere. Models for both Experiment 3 and Experiment 4 can be seen in Table 8.

Since neither target is present in these conditions, errors are necessarily *false alarms*. Figure 8 shows the false alarm rate for block-mismatched trials in Experiments 3 and 4. Across both experiments, there was a clear relationship between the colour of the vowels in the stimulus and the rate of false alarms — colour-harmonic items with all front/unround vowels were most likely to elicit false alarms for [i] and [e] trials, and least likely to elicit false alarms for [u] and [o] trials ($p < 0.001$ for both Experiments 3 and 4), and vice versa ($p < 0.001$ for both Experiments 3 and 4).

This pattern of false alarms is consistent with either feature-based identification or sen-

False Alarms (Mismatched Blocks), Experiment 3

	Estimate	Std. Err.	z value	Pr($>$ |z|)	
(Intercept)	-2.63970	0.34267	-7.703	1.33e-14	***
Trial Colour	0.15806	0.05259	3.005	0.002652	**
Colour (mixed)	0.12524	0.06380	1.963	0.049643	*
Colour (Back/Round)	0.02205	0.03833	0.575	0.565159	
Block Height	-0.20912	0.05307	-3.940	8.14e-05	***
Block	0.33867	0.05084	6.662	2.70e-11	***
Trial Colour $\times$ Colour (mixed)	-0.21796	0.06281	-3.470	0.000521	***
Trial Colour $\times$ Colour (Back/Round)	-0.22135	0.03830	-5.780	7.48e-09	***
Colour (mixed) $\times$ Block Height	-0.17637	0.06184	-2.852	0.004345	**
Colour (mixed) $\times$ Block Height	-0.03149	0.03772	-0.835	0.403811	

False Alarms (Mismatch Blocks), Experiment 4

	Estimate	Std. Err.	z value	Pr($>$ |z|)	
(Intercept)	-2.573013	0.343786	-7.484	7.19e-14	***
Trial	0.001352	0.035954	0.038	0.970012	
Colour (mixed)	0.007633	0.043644	0.175	0.861170	
Colour (Back/Round)	-0.017714	0.027538	-0.643	0.520043	
Task (First)	0.511697	0.104462	4.898	9.66e-07	***
Task (Last)	-0.065065	0.090585	-0.718	0.472589	
Block	0.740048	0.033589	22.032	<2e-16	***
Trial $\times$ Colour (mixed)	-0.360001	0.042091	-8.553	<2e-16	***
Trial $\times$ Colour (Back/Round)	-0.321130	0.026714	-12.021	<2e-16	***
Trial $\times$ Task (First)	0.169210	0.057121	2.962	0.003054	**
Trial $\times$ Task (Last)	-0.067407	0.050251	-1.341	0.179782	
Colour (mixed) $\times$ Task (First)	-0.203558	0.066502	-3.061	0.002207	**
Colour (Back/Round) $\times$ Task (First)	-0.092683	0.042201	-2.196	0.028075	*
Colour (mixed) $\times$ Task (Last)	0.118809	0.058941	2.016	0.043830	*
Colour (Back/Round) $\times$ Task (Last)	0.051313	0.036962	1.388	0.165053	
Colour (mixed) $\times$ Block	0.097722	0.038739	2.523	0.011650	*
Colour (Back/Round) $\times$ Block	-0.001602	0.023916	-0.067	0.946590	
Trial $\times$ Colour (mixed) $\times$ Task (First)	0.102736	0.066567	1.543	0.122749	
Trial $\times$ Colour (Back/Round) $\times$	0.164499	0.042334	3.886	0.000102	***
Trial $\times$ Colour (mixed) $\times$ Task (Last)	-0.108769	0.059079	-1.841	0.065609	.
Trial $\times$ Colour (Back/Round) $\times$ Task (Last) -	0.090290	0.037229	-2.425	0.015296	*

Table 8: Mixed effects models for false alarms, Experiments 3 and 4

sitivity to overall similarity between the whole nonsense word and the target vowel. If subjects are using sub-segmental features to identify the target vowels, the presence of [o] in e.g. *hogeka* will activate the [BACK/ROUND] component of the target vowel [u], leading to occasional erroneous perception or recall of [u] — an additional [o] in *hogoka* will strengthen that activation, and further increase the false alarm rate. If overall similarity is at play, subjects will judge *hogeka* to be more similar to [u] than *hegeka*, and will be more likely to erroneously report having heard [u].

Overall similarity is more likely to have an effect in the non-localised conditions (Experiment 3 and the *Any Syllable* condition of Experiment 4), where the whole word is the domain of attention. However, even in the *First Syllable* condition of Experiment 4, the effect remains significant ($p < 0.001$). This is driven in part by the presence of potential stimuli in non-initial syllables, which may serve as distractors — subjects were more likely to make false alarms with a potential target in a non-initial syllable ($p < 0.001$), but even when there was no potential target present anywhere in the word, the effect remained significant — the false alarm rate increased with the number of vowels in the nonsense word that shared colour features with the target vowel ($p < 0.01$).

This cannot, of course, rule out the possibility that subjects are still using overall similarity as part of their strategy; however, the high success rate of subjects in this condition indicates that they are indeed performing the task of identifying the vowel in the initial syllable, which is a realistic component of natural speech perception. The persistence of this effect in a context where overall similarity is less likely to be relevant or useful (or if it is, could plausibly be expected to be relevant and useful in natural speech perception as well) suggests that the effect is unlikely to be driven by an artificially task-induced reliance on whole-word similarity.

7.2   False Alarms vs. Misses

In block-matched data, where at least one of the possible targets is present in the attended location, errors can either be *false alarms* or *misses*. The data here is restricted to height-disharmonic items, where both error types are possible — for height-harmonic items (in Experiments 1–3 and the *Any Syllable* condition of Experiment 4), both error types are

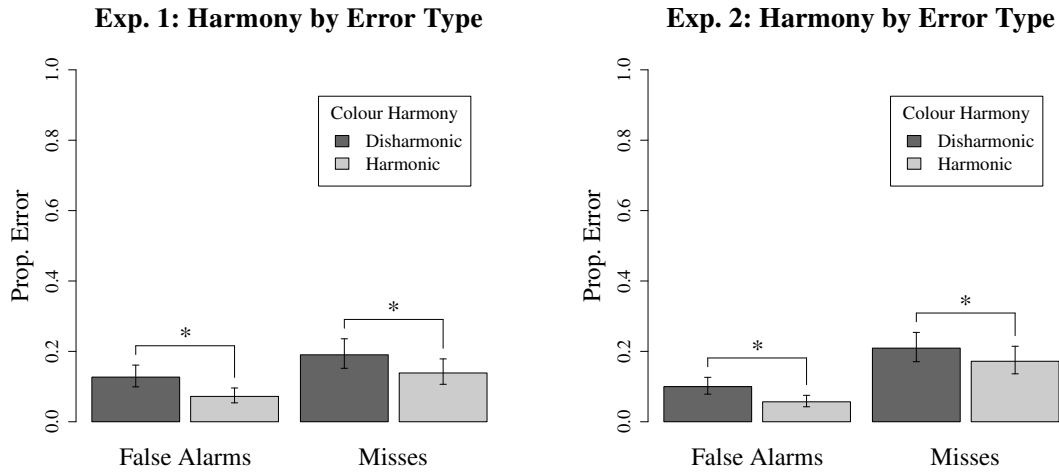**Exp. 1: Harmony by Error Type**   **Exp. 2: Harmony by Error Type**

Figure 9: Error rate a function of colour harmony and error type, height disharmonic items only, Experiments 1 and 2. Error bars represent 95% Confidence Intervals.

possible for items with colour harmony, but colour-disharmonic items contain both possible targets, so only misses are possible.[16]

As in previous analyses, mixed effects models were fitted on these data, with random intercept for subject and random slopes for subject × colour harmony; deviation coding was used throughout. Full models can be seen in Tables 9–12. The most relevant fixed effect (in addition to colour harmony) is correct response — errors on trials with a correct response of 'no' are false alarms, and errors on items with a correct response of 'yes' are misses.

Figures 9–10 show the effect of colour harmony on the error rate, across both false alarms and misses. In Experiment 1, there was a significant interaction between harmony and correct response ($p < 0.05$), the effect of harmony was significant in both false alarms ($p < 0.001$) and misses ($p < 0.001$). Likewise in Experiment 2; the interaction between harmony and correct response was significant ($p < 0.01$), but the effect of harmony was significant for both false alarms ($p < 0.001$) and misses ($p < 0.01$). Experiment 3 also showed a significant interaction between harmony and correct response ($p < 0.001$). In this case, however, the effect of harmony was significant for false alarms ($p < 0.01$), but interacted with block for

---

[16]The analysis presented here was also checked with height-harmonic items included, and it did not result in a qualitative difference to the results.
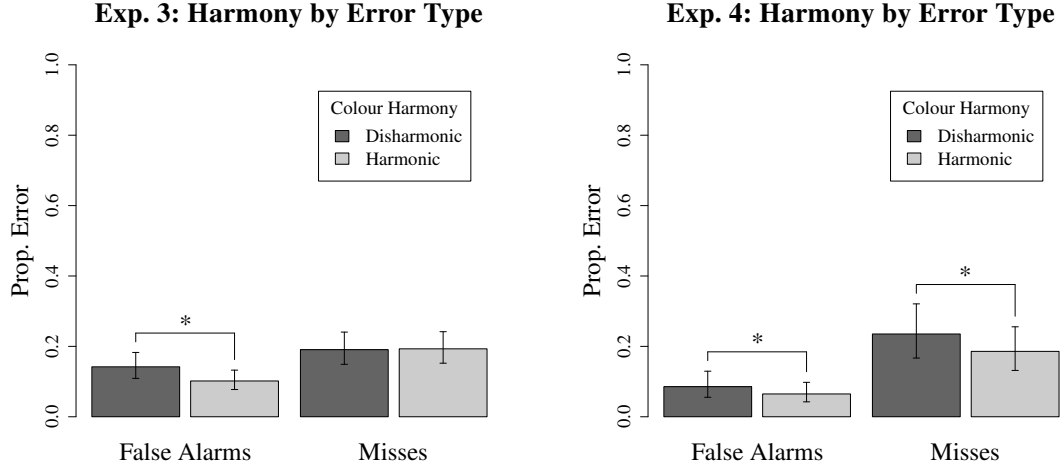
Figure 10: Error rate a function of colour harmony, error type, and Block order, height disharmonic items only, Experiments 3 and 4. Error bars represent 95% Confidence Intervals.

misses, where it was non-significant in Block 2 but marginal in Block 1 (p = 0.0577). In Experiment 4, the main effect of harmony was significant (p < 0.01) and did not interact with correct response.

False alarms and misses represent distinct categories of error — responding 'yes' on *hugoka ... i* in the case of the former, and responding 'no' on *hugoka ... u* in the case of the latter. For block-mismatched trials in Section 7.1, we saw that colour harmony was advantageous (compared to disharmony) when the nonsense word and target bore different feature values (leading to fewer false alarms), but disadvantageous when the nonsense word

Error Types (False Alarms vs. Misses), Experiment 1

	Estimate	Std. Err.	z value	Pr(> |z|)	
(Intercept)	2.24012	0.14192	15.784	<2e-16	***
Colour Harmony	-0.31358	0.05072	-6.183	6.30e-10	***
Locality	0.18353	0.02675	6.861	6.82e-12	***
Trial	0.07221	0.02673	2.702	0.0069	**
Correct Response	-0.60453	0.05424	-11.145	<2e-16	***
Block	-0.05546	0.02624	-2.113	0.0346	*
Colour Harm. × Loc.	-0.01643	0.02675	-0.614	0.5389	
Colour Harm. × Trial	0.04421	0.02672	1.654	0.0981	.
Loc. × Trial	0.01234	0.02672	0.462	0.6442	
Colour Harm. × Correct Resp.	0.12419	0.05424	2.290	0.0220	*
Colour Harm. × Loc. × Trial	-0.05347	0.02672	-2.001	0.0454	*

Table 9: Mixed effects model for error type, Experiment 1

Error Types (False Alarms vs. Misses)

	Estimate	Std. Err.	z value	Pr(> |z|)	
(Intercept)	2.505007	0.131890	18.99	<2e-16	***
Colour Harmony	-0.306846	0.059202	-5.18	2.2e-7	***
Distance (1)	-0.021081	0.060815	-0.35	0.7289	
Distance (2)	-0.014571	0.034505	-0.42	0.6728	
Correct Response	-1.054885	0.061715	-17.09	<2e-16	***
Trial	-0.000974	0.049871	-0.02	0.9844	
Colour Harm. × Dist. (1)	-0.141867	0.060815	-2.33	0.0197	*
Colour Harm. × Dist. (2)	0.009936	0.034505	0.29	0.7734	
Colour Harm. × Correct Resp.	0.185695	0.061704	3.01	0.0026	**
Dist. (1) × Correct Resp.	-0.037096	0.074018	-0.50	0.6162	
Dist. (2) × Correct Resp.	0.152681	0.043143	3.54	0.0004	***
Colour Harm. × Trial	0.161444	0.049871	3.24	0.0012	**
Colour Harm. × Trial	0.025854	0.061137	0.42	0.6724	
Colour Harm × Dist. (1) × Correct Resp.	0.220121	0.074017	2.97	0.0029	**
Colour Harm × Dist. (2) × Correct Resp.	-0.010224	0.043144	-0.24	0.8127	
Colour Harm × Correct Resp. × Trial	-0.199652	0.061137	-3.27	0.0011	**

Table 10: Mixed effects model for error type, Experiment 2

Error Type (False Alarms vs. Misses)

	Estimate	Std. Err.	z value	Pr(> |z|)	
(Intercept)	1.98868	0.14501	13.715	<2e-16	***
Colour Harmony	-0.19042	0.04983	-3.822	0.000133	***
Correct Response	-0.55054	0.06056	-9.091	<2e-16	***
Trial Colour	-0.15578	0.04589	-3.395	0.000687	***
Block Height	0.19568	0.04597	4.256	0.0000207946	***
Block	-0.02502	0.02974	-0.841	0.400130	
Colour Harm. × Correct Resp.	0.19745	0.06026	3.277	0.001050	**
Colour Harm. × Trial Colour	-0.02244	0.02976	-0.754	0.450816	
Correct Resp. × Block Height	-0.15919	0.06033	-2.638	0.008328	**
Correct Resp. × Trial Colour	0.25713	0.06031	4.264	0.0000201072	***
Trial Colour × Block Height	0.16475	0.04580	3.597	0.000322	***
Colour Harm. × Block Height	-0.04411	0.02976	-1.482	0.138247	
Colour Harm. × Block	-0.07073	0.02968	-2.383	0.017156	*
Block Height × Block	0.35275	0.14035	2.513	0.011958	*
Trial Colour × Block	-0.08914	0.02967	-3.004	0.002665	**
Correct Resp. × Trial Colour × Block Height	-0.34059	0.06025	-5.653	0.0000000157	***
Colour Harm. × Block Height × Block	0.07639	0.03317	2.303	0.021266	*
Trial Colour × Block Height × Block	0.12363	0.02967	4.167	0.0000309207	***

Table 11: Mixed effects model for error type, Experiment 3

Error Types (False Alarms vs. Misses)

	Estimate	Std. Err.	z value	Pr(> |z|)	
(Intercept)	2.6807	0.2412	11.11	<2e-16	***
Colour Harmony	-0.1492	0.0513	-2.91	0.00363	**
Correct Response	-1.1121	0.1460	-7.62	2.6e-14	***
Task (First)	-0.8106	0.1657	-4.89	9.9e-07	***
Task (Last)	-0.1374	0.1862	-0.74	0.46071	
Block	-0.7242	0.1327	-5.46	4.9e-08	***
Trial	0.0778	0.0391	1.99	0.04626	*
Locality	0.2310	0.0424	5.45	4.9e-08	***
Correct Resp. × Task (First)	-0.2319	0.1574	-1.47	0.14073	
Correct Resp. × Task (Last)	0.7459	0.1787	4.17	3.0e-05	***
Correct Resp × Block	0.6121	0.1459	4.20	2.7e-05	***
Task (First) × Block	0.6304	0.1480	4.26	2.0e-05	***
Task (Last) × Block	0.2658	0.1591	1.67	0.09479	.
Correct Resp. × Task (First) × Block	-0.5290	0.1573	-3.36	0.00077	***
Correct Resp. × Task (Last) × Block	-0.4860	0.1786	-2.72	0.00651	**

Table 12: Mixed effects model for error type, Experiment 4

and the target bore the same feature value (leading to more false alarms). Here we see that — even without the total identity offered by concurrent height agreement — harmony results overall in fewer false alarms than disharmony. And with the exception of Experiment 3, harmony results in fewer misses than disharmony. In other words, harmony serves to both dissuade an erroneous detection and reinforce a successful one.

This effect is still consistent with a listener strategy based on overall similarity between nonsense word and target, and once again the persistence of the effect even in the *First Syllable* condition in Experiment 4 (recall that there was no interaction between harmony and either task or correct response in the error analysis) is suggestive but not conclusive.

7.3   Block Order

By examining the effects of block order, it's possible to find some indication of the effect of the time-course of the experiment on the error rate. For experiments where block order showed a significant effect, subjects made fewer errors in the second block than in the first ($p < 0.05$ for Experiment 1, $p < 0.001$ in Experiment 4). This suggests that errors are not induced by participant fatigue (which would result in greater errors in Block 2) — rather, increasing experience with the task likely allowed subjects to develop knowledge of the task (relevant alternatives, etc.) that streamlined their decision process.

8 General Discussion

The primary objective in this series of experiments has been to investigate whether vowel harmony is perceptually advantageous, and the extent to which that advantage persists among non-adjacent vowels.

A summary of the main findings of the four experiments can be seen in Table 13. The most robust finding, replicated across all experiments, is the advantage of harmony — subjects were consistently faster and more accurate at identifying a colour contrast in words with harmony. Even in conditions where the effect of harmony was diminished — among non-adjacent vowels and among vowels which disagreed in height — it remained consistently statistically significant. There was also a consistently replicated interaction between this effect and height agreement — harmony was more advantageous among vowels of the same height than among vowels of different heights.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Main effect of Harmony	✓	✓	✓	✓
...even non-locally	✓	✓	✓	✓
...even without height agreement	✓	✓	✓	✓
...in both false alarms and misses	✓	✓	✗	✓
...when only initial syllable attended	n/a	n/a	n/a	✓
Interaction with Height Agreement	✓	✓	✓	✓
Interaction with Locality	✓	(✓)	✗	✗
Interaction with Trial Colour	✓	✓	✗	(✗)
Interaction with Trial Height	n/a	n/a	✓	n/a

Table 13: Summary of results, experiments 1–4.

Less stable effects included interactions with locality and with Trial Colour. For locality, Experiment 1 found a small but significant interaction between locality and harmony, with the advantage of harmony diminishing at a distance of one syllable. For Experiments 2–4, this interaction appeared as a non-significant trend. In Experiment 2, the effect of harmony was further diminished at a distance of two syllables, but only reached statistical significance in height agreement conditions. For Trial Colour, Experiments 1 and 2 found that the effect

of harmony was greater for [u] trials than for [i] trials, though this was not replicated in Experiment 3–4.

In Experiment 3, the design was expanded to include both high-target and mid-target blocks, and it was found that the advantage of harmony was greater for high targets than for mid targets. In Experiment 4, the design was expanded to include tasks which asked subjects to attend to specific syllables —initial syllables were of particular interest, and the effect of harmony was found to be significant in this task as well.

The remainder of this section discusses the implications that these findings have for a theory of phonetically grounded vowel harmony. Section 8.1 discusses the connections between these results and asymmetries found in the typology of vowel harmony, while Section 8.2 briefly locates these findings within existing proposals about the role of phonetic substance in phonological grammar.

8.1   Connections to Typology

The findings summarised above provide evidence that harmony confers a perceptual advantage, and that this advantage includes non-local instances. In local cases of harmony, this perceptual advantage can be seen as adding to previously well-established articulatory sources of grounding (rather than seeking to replace them). However, in the case of transparent harmony, the two sources of grounding make disparate predictions. A theory of phonetic grounding based solely on articulatory factors predicts that harmony should only ever be local — such theories struggle to account for the empirical reality of long-distance harmony. The finding that the perceptual advantage of harmony persists even among non-adjacent vowels provides phonetic grounding for the existence of long-distance harmony processes. In addition, this conflict between articulatory and perceptual sources of grounding helps to explain why long-distance harmony, while attested, is typologically marked — every vowel harmony process which applies across a distance also applies locally, but not vice versa. Local harmony has both sources of phonetic grounding supporting it, while long-distance harmony relies only on perceptual factors.

Additionally, there is an attested tendency for the likelihood of harmony to decrease across a distance (as in Hungarian, discussed above). This is somewhat difficult to explain

articulatorily, since increasing distance between harmony triggers and targets does not alter the number of gestural transitions required to complete the sequence.[17] However, the finding in Experiment 2 that the advantage of harmony is significantly decreased across a distance of two vowels can provide a source of phonetic grounding for this distance effect.

The finding that height agreement has an effect on the advantage of harmony corresponds with cross-linguistic patterns of *parasitic harmony*, where harmony is only applicable among segments which agree along some other feature dimension (see Cole and Trigo 1988; Cole and Kisseberth 1995; Kaun 1995 for further discussion). Like local harmony, this also has an articulatory explanation, at least for rounding harmony — on the basis of Linker's (1982) articulatory data, Kaun (1995) argues that the rounding gestures involved in executing high and non-high round vowels are qualitatively different, meaning that executing a height-disagreeing sequence of round vowels would require two separate rounding gestures. This would therefore not accomplish the articulatory goals of harmony as well as executing a height-agreeing sequence of round vowels. In this case, again, articulatory and perceptual grounding appear to be in agreement — harmony is predicted to occur even among height-disagreeing vowels, but should be additionally licensed among height-agreeing vowels.

The finding that subjects perform better at identifying a colour contrast in high vowels than in mid vowels lends support for Kaun's (1995) perceptually-driven explanation of non-high vowels' status as preferential triggers in rounding harmony. This, like the existence of long-distance harmony and distinctions among degrees of distance, finds no explanation from articulatory sources of grounding — in fact, the articulatory extremes of high vowels should result in *stronger* vowel-to-vowel coarticulation, the precursor to (local) harmony. The study here advances the chain of reasoning where this asymmetry is concerned: Linker's (1982) findings show that non-high vowels are less articulatorily extreme than their high counterparts, and Terbeek (1977) has demonstrated that listeners rate the latter as sounding more extremely round. The results here show that this difference in extremity does in fact have an impact on listeners' success in identifying the relevant feature contrast, and the main

---

[17](Benus, 2005) does predict that, since the magnitude of the vowel-to-vowel coarticulation used to induce backness across a transparent [i] diminishes with distance, this should also diminish the probability of transparent harmony. As discussed in Section 2.1, Szeredi (2012) provides evidence that the effect does not surpass the Just Noticeable Difference threshold, so diminution of the effect with distance should be even less accessible.

effect seen across all three experiments shows that harmony is indeed a way of achieving improvement in feature recognition.

The interaction between harmony and colour may provide a source of grounding for the generalisation that [+ROUND] is phonologically active, while [–ROUND] does not ever appear to be. However, this finding should be taken with an appropriately-sized grain of salt: first, the effect did not consistently replicate, and second, if the phonological activity of feature values is an integral part of the structure of feature systems, the effects seen here may be a *result* of that structure rather than a precursor for it. Indeed, this raises more questions than it provides answers — to what extent can privativity vs. equipollence be predicted from the salience of phonetic cues, and to what extent does the specification of features impact perception? Caution is indicated in interpreting the results discussed above, but at the very least these findings suggest that this could be an interesting avenue for further inquiry.

8.2   Possible Mechanisms

The connection between the results of these experiments and the typology of vowel harmony raises a pressing question: how might perceptual effects like these shape phonological systems? The role of phonetic grounding in phonological grammar is a long-standing and unresolved issue in phonological theory, and resolution of the question is far beyond the scope of this paper. However, in this section I will briefly attempt to locate these results within current proposals in this area.

In discussion of phonetic grounding or phonetic precursors, most approaches rely on the notion of *bias*, which can be divided into several broad categories: *channel bias* is the systematic distortion in the speech signal itself in the course of articulation and perception, while *analytic bias* represents a systematic tendency on the part of the learner to preferentially acquire particular types of phonotactic generalisations or phonological rules (Moreton, 2008). Moreton and Pater (2012) further divide analytic bias into *structural* bias, the preferential learning of computationally simpler patterns, and *substantive* bias, the preferential learning of patterns which correspond with sources of phonetic grounding. Setting aside structural bias,[18] most of the results discussed above are compatible with either channel

---

[18]Because the issue here is one of phonetic grounding, structural complexity is of limited relevance —

biases in perception or substantive analytic bias in learning.

In a theory based on channel bias, such as Evolutionary Phonology (Blevins, 2004; Ohala, 1994), these systematic effects in perception shape the listener's mental representations of surface phonological forms. The traditional explanation of the precursors of harmony in channel bias has been primarily articulatory — gradient vowel-to-vowel coarticulation results in vowels whose feature values are misperceived, and sequences of vowels with a shared feature value are posited by the listener. The main contribution of this study is to show that misperception can be biased in this direction (towards harmonic sequences) *even in the absence of articulatory asymmetries in the speech signal*. If systematic misperception is possible even without detectable coarticulation, the traditional explanation for the evolution of vowel harmony is consistent with the existence of long-distance as well as local harmony.

In a theory based on substantive bias, a learner has some sort of knowledge about the phonetic properties of speech — including the relationship between those properties and articulatory ease, perceptual salience, etc. — and that knowledge influences which potential generalisations over the learning data are given priority in learning. The form that this knowledge takes and the mechanisms by which it exerts influence over learning differ among proposals that make use of substantive bias — it can be innate or it can arise through a learner's experience with his or her own articulatory and perceptual organs; it can be a strict limit on possible generalisations or it can contribute to differences in the prior likelihoods of possible generalisation (see e.g. Collins 2013; Bermúdez-Otero and Kersti Börjars 2006; Wilson 2006; Hayes and White 2013 for further discussion). The perceptual asymmetries found in the experiments above, then, would reside in a learner's phonetic awareness — among the patterns given priority by the relevant learning mechanisms would be those which render colour contrasts among vowels more salient, including long-distance harmony.

Another possible mechanism would introduce harmony processes via lexical competition — if words exhibiting vowel harmony are recognised and recalled more effectively than words which do not, those lexical items should accrue over time, and this accrual in the lexicon may lead learners to posit a categorical phonological rule instantiating the generalisation.

---

additionally, the complexity of long-distance harmony processes is highly theory-dependent and therefore difficult to assess.

See e.g. (Martin, 2007) for a proposal for how lexical competition can introduce changes to a phonological grammar.

There is very little here to differentiate between channel bias and substantive analytic bias as a mechanism for connecting this possible source of phonetic grounding to the phonological grammar; as we have just seen, the main result can easily be accommodated in either approach. One potential divergence, though, concerns the status of non-high rounded vowels as preferential triggers. The results here lend support to an explanation of this typological pattern in terms of perceptual salience in a way that is not straightforwardly accounted for by Evolutionary Phonology.[19] In a theory based on substantive bias, knowledge that non-high vowels are perceptually impoverished means that a phonological process which increases their salience should have some learning advantage over a process which boosts the salience of high vowels. Because this depends on the learner prioritising patterns which benefit the listener — rather than following from the result of mistakes that the learner makes as a listener — it would seem to lend at least some degree of support to theories of substantive bias.

9   Conclusion

This paper has presented experimental evidence that vowel harmony is perceptually advantageous — and, crucially, that this advantage obtains even among non-adjacent vowels. Across all three experiments, subjects consistently performed significantly better at a feature-based identification task with colour-harmonic nonsense words than with their disharmonic counterparts. The advantage of harmony was diminished across increasing distance, among vowels which disagreed in height, and for trials with non-round targets (mirroring various cross-linguistic tendencies in the typology of vowel harmony) — but the effect of harmony was robustly detectable across all conditions.

This provides a potential source of phonetic grounding for long-distance (or 'transparent') vowel harmony, which has been described as an 'unnatural' phonological process. This lends

---

[19]This is not to say that this explanation cannot be incorporated into this theory with further analysis or development.

support to theories of vowel harmony in phonology which make use of explicitly non-local representations; while locality still has an important role to play in phonological theory, it should not be understood as a strict, inviolable restriction. While theories of phonetic grounding based solely in articulation must hold transparent harmony as an aberration, including perceptual factors allows for both direct and explicit analysis of long-distance harmony and preservation of phonetic grounding.

References

Archangeli, Diana, and Douglas Pulleyblank. 1994. *Grounded phonology*. MIT Press.

Astheimer, L. B., and L. Sanders. 2011. Predictability affects early perceptual processing of words onsets in continuous speech. *Neuropsychologia* 49:3512–3516.

Astheimer, Lori B, and Lisa D Sanders. 2009. Listeners modulate temporally selective attention during natural speech processing. *Biological psychology* 80:23–34.

Bach, Emmon, and R.T. Harms. 1972. How do languages get crazy rules? In *Linguistic change and generative theory*, ed. Robert Stockwell and Ronald Macaulay, 1–21. Bloomington: Indiana University Press.

Bakovic, Eric. 2000. Harmony, dominance, and control. Doctoral Dissertation, University of California, Santa Cruz. ROA-360.

Bates, Douglas, Martin Maechler, Benjamin M. Bolker, and Steven Walker. 2014. lme4: Linear mixed-effects models using eigen and s4. ArXiv e-print; submitted to *Journal of Statistical Software*.

Beckman, Jill. 1998. Positional faithfulness. Doctoral Dissertation, University of Massachusetts, Amherst.

Benus, Stefan. 2005. Dynamics and transparency in vowel harmony. Doctoral Dissertation, New York University.

Benus, Stefan, and Adamantios I. Gafos. 2007. Articulatory characteristics of Hungarian transparent vowels. *Journal of Phonetics* 35:271–300.

Bermúdez-Otero, Ricardo, and Kersti Börjars. 2006. Markedness in phonology and in syntax: The problem of grounding. *Lingua* 116:710–756.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1962. *The Menomini language*. Yale University Press.

Blust, Robert A. 1996a. Low vowel dissimilation in Ere. *Oceanic Linguistics* 35:96–112.

Blust, Robert A. 1996b. Low vowel dissimilation in Oceanic languages: An addendum. *Oceanic Linguistics* 35:305–309.

Boersma, Paul. 1998. Spreading in functional phonology. Ms, University of Amsterdam.

Clements, George N. 1977. The autosegmental treatment of vowel harmony. In *Phonologica 1976*, ed. W. Dressler and O. Pfeiffer.

Cole, Jennifer, and Charles W. Kisseberth. 1995. Nasal harmony in Optimal Domains Theory. Ms., University of Illinois. ROA-49.

Cole, Jennifer, and L. Trigo. 1988. Parasitic harmony. In *Features, segmental structure, and harmony processes II*, ed. Harry van der Hulst and Norval Smith, 19–38. Foris.

Collins, Joe. 2013. Modal-Dependence and Naturalness in Phonology: Confronting the Ontogenetic Question. MA thesis, University of Tromso.

Finley, Sara. 2015. Learning non-adjacent dependencies in phonology: Transparent vowels in vowel harmony. *Language* 91:48–72.

Flemming, E. 1995. Auditory representations in phonology. Doctoral Dissertation, University of California, Los Angeles.

Flemming, E. 2006. The role of distinctiveness constraints in phonology. Ms, MIT.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In *Phonetically-based phonology*, ed. B. Hayes, R. Kirchner, and D. Steriade, 232–276. Cambridge University Press.

Fox, John. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* 8:1–27. URL `http://www.jstatsoft.org/v08/i15/`.

Gafos, Adamantios I. 1998. Eliminating long-distance consonantal spreading. *Natural Language and Linguistc Theory* 223–278.

Gafos, Adamantios I. 1999. The articulatory basis of locality in phonology. Doctoral Dissertation, Johns Hopkins University.

Gallagher, Gillian. 2010. The perceptual basis of long-distance laryngeal restrictions. Doctoral Dissertation, MIT.

Gerrits, E, and M E H Schouten. 2004. Categorical perception depends on the discrimination task. *Perception & psychophysics* 66:363–376.

Gick, Bryan, Douglas Pulleyblank, Fiona Campbell, and Ngessimo Mutaka. 2006. Low vowels and transparency in Kinande vowel harmony. *Phonology* 23:1–20.

Goldsmith, J. 1976. Autosegmental phonology. Doctoral Dissertation, MIT. Published 1979, Garland, New York.

Gordon, Matthew K. 1999. The "neutral" vowels of Finnish: How neutral are they? *Linguistica Uralica* 35:17–21.

Hansson, Gunnar. 2001. Theoretical and typological issues in consonant harmony. Doctoral Dissertation, University of California, Berkeley.

Hayes, Bruce, and Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23:59–104.

Hayes, Bruce, and James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44.

Hayes, Bruce, Kie Zuraw, Péter Siptár, and Zsuzsa Cziráky Londe. 2009. Natural and unnatural constraints in Hungarian. *Language* 85:822–863.

Itô, Junko. 1984. Melodic dissimilation in Ainu. *Linguistic Inquiry* 15:505–513.

Kaun, Abigail. 1995. The typology of rounding harmony: An Optimality Theoretic approach. Doctoral Dissertation, University of California, Los Angeles.

Kimper, Wendell. 2011. Competing triggers: Transparency and opacity in vowel harmony. Doctoral Dissertation, University of Massachusetts Amherst.

Kimper, Wendell. 2013. Perceptual motivations for parasitic restrictions in vowel harmony. In *Proceedings of the 41st Annual Meeting of the North East Linguistic Society*, ed. Yelena Fainleib, Nicholas LaCara, and Yangsook Park, volume 1, 259–268. Amherst, MA: GLSA.

Kiparsky, Paul. 1981. Vowel harmony. Ms., MIT.

Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2014. *lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)..* URL `http://CRAN.R-project.org/package=lmerTest`.

Linker, Wendy. 1982. Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study. Doctoral Dissertation, UCLA.

Macmillan, Neil A., and Douglas Creelman. 2004. *Detection theory: A user's guide*. Psychology Press, 2nd edition.

Martin, Andrew. 2007. The evolving lexicon. Doctoral Dissertation, University of California, Los Angeles.

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

Moreton, Elliott. 2012. Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language* 67:165–183.

Moreton, Elliott, and J Pater. 2012. Structure and Substance in Artificial Phonology Learning, Part I: Structure. *Language and Linguistics Compass* 686–701.

Ni Chiosain, Maire, and Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In *Segmental phonology in optimality theory: Constraints and representations*, ed. Linda Lombardi. New York: Cambridge University Press.

Ohala, John J. 1994. Towards a universal, phonetically-based theory of vowel harmony. In *ICSLP3*, 491–494.

Pisoni, David B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics* 13:253–260.

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality theory: Constraint interaction in generative phonology*. Blackwell Press.

Pulleyblank, Douglas. 1996. Neutral vowels in Optimality Theory: A comparison of Yoruba and Wolof. *Canadian Journal of Linguistics* 41:295–347.

R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Rhodes, Russell. 2012. Vowel harmony as agreement by correspondence. UC Berkeley Phonology Lab Annual Report.

Riggle, Jason. 1999. Relational markedness in Bantu vowel height harmony. Master's thesis, University of California, Santa Cruz.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80:475–531.

Schlindwein, Deborah. 1987. P-bearing units: A study of Kinande vowel harmony. In *NELS 17*, 551–567.

Schneider, W., A. Eschman, and A. Zuccolotto. 2002. *E-prime user's guide*. Psychology Software Tools Inc.

Steriade, Donca. 1995a. Positional neutralization. In *NELS 24*. University of Massachusetts, Amherst.

Steriade, Donca. 1995b. Underspecification and markedness. In *The handbook of phonological theory*, ed. John Goldsmith, 114–174. Oxford: Blackwell.

Suomi, Kari. 1983. Palatal harmony: A perceptually motivated phenomenon? *Nordic Journal of Linguistics* 6:1–35.

Szeredi, Daniel. 2012. Acceptability of harmonic mismatch for neutral vowel stems in hungarian. Ms, New York University.

Terbeek, Dale. 1977. A cross-language multi-dimensional scaling study of vowel perception. Doctoral Dissertation, UCLA.

Vago, R. M. 1976. Theoretical implications of Hungarian vowel harmony. *Linguistic Inquiry* 7:243–263.

Walker, Rachel. 1998. Nasalization, neutral segments, and opacity effects. Doctoral Dissertation, University of California, Santa Cruz.

Walker, Rachel. 2005. Weak triggers in vowel harmony. *Natural Language and Linguistc Theory* 23:917–989.

Wilson, Colin. 2006. Unbounded spreading is myopic. Bloomington, IN: PhonologyFest 2006.