The Role of Contrast in Limiting Vowel-to-vowel Coarticulation in Different Languages*

Sharon Y. Manuel†

Languages differ in their inventories of distinctive sounds and in their systems of contrast. Here we propose that this observation may have predictive value with respect to how extensively various phones are coarticulated in particular languages. This hypothesis is based on three assumptions: (1) there are "output constraints" on just how a given phone can be articulated; (2) output constraints are, at least in part, affected by language-particular systems of phonetic contrast; and (3) coarticulation is limited in a way that respects those output constraints. Together, these assumptions lead to the expectation that, in general, languages will tend to tolerate less coarticulation just where extensive coarticulation would lead to confusion of contrastive phones. This prediction was tested by comparing acoustic measures of anticipatory vowel-to-vowel coarticulation in languages which differ in how they divide up the vowel space into contrastive units. The acoustic measures were the first and second formant frequencies, measured in the middle and at the end of the target vowels /a/ and /e/, followed by /pV/, where /V/ was /i,e,a,o,u/. Two languages (Ndebele and Shona) with the phonemic vowels /i,e,a,o,u/ were found to have greater anticipatory coarticulation for the target vowel /a/ than does a language (Sotho) that has a more crowded mid- and low-vowel space, with the phonemic vowels /i,e,e,a,:l,o,u/. The data were based on recordings from three speakers of each of the languages.

INTRODUCTION

Languages differ in their inventories of distinctive sounds and in their systems of contrast. The premise of this paper is that this observation may have predictive value with respect to how extensively certain phones are coarticulated in particular languages. We provide an empirical test of our theoretical perspective by comparing vowel-to-vowel coarticulation in languages that differ in how they divide the vowel space into contrastive units.

* I would like to thank the faculty and students of the University of Zimbabwe for their generous help in the data collection. I am grateful to Suzanne Boyce, Harvey Gilbert, Marie Huffman, Rena Krakow, Harriet Magen, and Kenneth Stevens and three anonymous reviewers for valuable comments on an earlier draft of this paper, and to Louis Goldstein, Alvin Liberman, and Ignatius Mattingly for advice on my doctoral dissertation, on which this paper is based. I would also like to thank Tracy Sheppard, Linnea Bankey, and Yvonne Manning-Jones for help with manuscript preparation. This research was supported principally by NIH Grant HD-01994 to Haskins Laboratories and more recently by a NIH Postdoctoral Training Grant (DC-00005) to MIT.

Words are distinguished from one another by their phone composition, and since articulatory gestures (which ones, when they occur) and their acoustic consequences distinguish phones, we might expect speakers to exercise some effort to ensure that the acoustic consequences of articulatory gestures remain distinct (see, for example, Lindblom, 1983; Lindblom & Engstrand, 1989; Martinet, 1952, 1957; Stevens, 1989). That is, there may be output constraints (tolerances) on phones, definable in terms of articulatory or acoustic dimensions, that limit how much their articulatory gestures are allowed to stray from an ideally distinctive pattern. For example, a speaker who intends to say the word *tick* must make a tight tongue tip-alveolar closure—too weak a constriction will produce something that might be heard as *sick* or *thick*.

In speech, the articulatory requirements of one phone are often anticipated during the production of a preceding phone. This phenomenon, known as anticipatory coarticulation, results in contextually induced variability in the portion of the signal...
that we conventionally associate with a given phone. An example is the difference in how /t/ is produced in tea and tree. In the word tree the tongue may make a relatively more posterior contact with the roof of the mouth and may itself be retroflexed, in anticipation of the following retroflex consonant /r/. Similarly, in a vowel-nasal sequence such as in pan, the velum typically begins (and may complete) its lowering movement, associated with the nasal /n/, while the vocal tract is still open for the vowel /a/ and well before the oral occlusion for the /n/ is achieved.

Since coarticulation affects the very primitives of contrast between phones, i.e., articulatory gestures and their acoustic consequences, extreme coarticulation would possibly put speakers at risk of blurring or even obliterating phonetic contrasts. We might expect, then, that speakers generally limit coarticulation such that it does not destroy the distinctive attributes of gestures. That is, coarticulation might be limited so that the output constraints on distinctive gestures (i.e., gestures that have consequence for distinctiveness—analogous to distinctive features) are not violated (see also Engstrand, 1988; Manuel & Krakow, 1984; Manuel, 1987a, 1987b; Martinet, 1952, 1957; Schouten & Pols, 1979; Tatham, 1984; Keating, 1990).

But what counts as a distinctive gesture? Clearly, what counts varies from phone to phone and from language to language. For example, if the oral tract is completely closed, it matters quite a bit whether or not the velar port is appreciably open, since this is the major articulatory difference (and is responsible for the major acoustic difference) between voiced oral and nasal stops (e.g., /d/ vs. /n/). On the other hand, vowels in English are not distinguished by virtue of the fact that they are nasalized or not (though of course they normally are relatively nonnasal, the exceptions being when they occur in a sequence with nasal consonants). It is therefore not surprising that English tolerates the presence of a substantial velar lowering gesture during vowels in nasal contexts.

With respect to language-particular aspects of distinctiveness, we note that different languages essentially take a universally available articulatory/acoustic space and divide it up differently. For example, Swedish distinguishes [+round] from [-round] front vowels, whereas English does not. Similarly, vowels in French are distictively [+nasal] or [-nasal], but as noted above, English vowels are nasalized only in the context of a nasal consonant.

In some languages, very similar articulatory gestures may result in linguistically distinct phones. In order to maintain distinctiveness between those gestures, each must have a fairly narrow constraint on its production. On the other hand, if a language makes only a single phone in a particular region of some phonetic dimension, we might expect a fair amount of variability in production to be tolerated (see Keating and Huffman, 1984, for related discussion). For example, Malayalam makes two distinctive voiceless coronal stops—alveolar and dental, whereas English only makes one distinctive voiceless stop in that area of the vocal tract. English speakers have been shown to exhibit more variability in production of English voiceless coronal stops than do Malayalam speakers for Malayalam alveolar or dental stops (Jongman, Blumstein, & Lahiri, 1985). This observation suggests that the output constraints on gestures associated with particular phones, and consequently the degree to which those phones are susceptible to coarticulatory influence of neighboring phones, can be predicted to a large extent by examining the way a language uses a particular phonetic space.

The results of several previous studies can be indirectly or directly related to the role that contrast plays in constraining coarticulatory behavior in different languages (e.g., Chasaide, 1979; Lubker & Gay, 1982; Magen, 1984; Manuel & Krakow, 1984; Öhman, 1966). Öhman's early study of vowel-to-vowel coarticulation showed that for English and Swedish V1CV2 utterances, the articulators begin assuming configurations needed for V2 before the occlusion is reached for the medial consonant. However, similar effects were generally not found for Russian. As Öhman points out, Russian contrasts palatalized and unpalatalized consonants, and it is the position of the tongue body that distinguishes these two sets. Presumably in Russian the tongue body is not free to begin assuming the configuration for V2 during the V1 to C transition because the consonant itself makes (possibly conflicting) requirements on the tongue body.

Manuel and Krakow (1984) compared vowel-to-vowel coarticulation in English, a language with a relatively crowded vowel space (13 to 15 vowels, depending on dialect), and two Bantu languages (Shona and Swahili) for which the vowels are well spread-out (the phonemic vowels of Shona and Swahili are /i,e,a,o,u/). The prediction in Manuel and Krakow was that since the vowels of English are closer together in the articulatory/acoustic space, the range of productions for each English
The role of contrast in limiting vowel-to-vowel coarticulation in different languages

The vowel should be relatively small, while in contrast, since the vowels of Shona and Swahili are more spread apart, they could tolerate larger ranges of productions without danger of neutralizing phonemic differences. The main result of the experiment was as predicted. Vowel-to-vowel coarticulation (as reflected by values of the first and second formant frequencies measured in the middle of the vowels) was in fact stronger in Shona and Swahili than in English.

The results of the Manuel and Krakow study are consistent with the idea that proximity of contrastive phonetic units affects coarticulation. However, because the study was based on a single speaker of each language, we cannot conclude with certainty that Manuel and Krakow's results reflected language differences rather than simply speaker-to-speaker differences, given that speakers of a language may vary somewhat in amount and type of coarticulatory patterns (Lubker & Gay, 1982; Nolan, 1985). In addition, English vowels are often diphthongized, whereas Shona and Swahili vowels tend to be monophthongal. It is possible that the restricted degree of coarticulation in English was not due to constraints on the quality of vowel nuclei, but was instead due to demands of the diphthongal second part of the English vowels. Despite these confounding problems, the major finding was predicted by the assumption that distribution of vowels affects their output constraints, which in turn affect the amount of vowel-to-vowel coarticulation. We are unaware of any other hypothesis that has such predictive power for these results.

The current experiment is an expansion of the basic paradigm used by Manuel and Krakow. The three languages analyzed are all in the same family (Southern Bantu), and they are phonologically, morphologically, and syntactically similar to each other. Several speakers of each language were recorded, to allow comparison of coarticulation both between speakers of the same language, and across languages.

We included two languages (Shona and Ndebele) with the phonemic vowels /i,e,a,o,u/ and one language (Sotho) that has a more crowded vowel inventory /i,e,a,o,u/. It should be kept in mind that what is at issue here is not the number of vowels per se, but their distribution. As can be seen in Figure 1, the Sotho vowel space is more crowded in the low- and mid-vowel region than are the Shona and Ndebele vowel spaces. As a mnemonic for which languages have relatively less crowded (LC) and more crowded (MC) vowel spacing, we will use the shorthand Ndebele(LC), Shona(LC) and Sotho(MC).

![Figure 1](https://example.com/figure1.png)

*Figure 1. Examples of phonemic vowels of Ndebele, Shona and Sotho. Data are from one speaker of each of the languages.*
In Sotho(MC), the vowels /e, o/ intervene between phonemic /a/ and phonemic /e, o/. If this relative crowding affects output constraints and limits coarticulation in Sotho(MC), we would expect to find smaller vowel-to-vowel coarticulatory effects on /a/, /e/, and /o/ in Sotho(MC) than in Shona(LC) and Ndebele(LC). Specifically, we expect Sotho(MC) /a/ to show less movement into the /e, o/ space. For the vowel /e/, we expect less movement into the /e/ space, and for /o/, less movement toward the /o/ space. Here we examine anticipatory coarticulation for the target vowels /a/ and /e/. The acoustic measures of coarticulation were the first (F1) and second (F2) formant frequencies.

Method
A. Further specifics of languages and subjects
The languages we studied are characterized by open syllables, and the vowels are monophthongal (see Cole, 1955; Doke, 1954). Word stress is not distinctive, but rather it is fixed on the penultimate syllable. Tone is phonemic in all three of the languages. Phonetically, the vowels /i/ and /u/ are quite high in these languages. Similarly, the mid vowels /e/ and /o/ are also phonetically rather high, with /e/ approaching the quality of [i]. The low vowel /a/ is more fronted than its English counterpart. The Sotho(MC) vowel /e/ is phonetically slightly higher than the vowel in English bet and Sotho(MC) /o/ is similar to the vowel in English caught (for dialects which distinguish the vowels in caught and cot). Of possible relevance to the present study is the claim that Sotho(MC) has a phonological rule which raises its mid vowels /e, o, a/ when they are followed in the next syllable by a higher vowel or a syllabic nasal, as well as by some palatal consonants. This process has the effect of raising the target vowel by a half step. For example, raised /e/ is higher than nonraised /e/ but not as high as /i/. Similarly, raised /e/ is higher than nonraised /e/, but lower than nonraised /e/.

The data reported here are from three adult male native speakers of each of the languages studied. All subjects were fluent speakers of English as a second language. The subjects were paid for their participation.

B. Materials
The /a/ and /e/ vowels that we analyzed were a subset of a larger data base. This larger database was comprised of nonsense trisyllables of the form /pV1pV2pV3/, where the vowels could be any one of the five vowels /a, e, i, o, u/. The consonant /p/ was chosen because its articulation is relatively independent of articulations involving the tongue. These nonsense trisyllables were spoken in carrier phrases which were selected so that the last phoneme of the word preceding the target trisyllable was an /a/, and the first syllable of the word following the trisyllable was /pa/. The carrier phrases and their glosses are shown below:

Shona(LC): Taura pepapa pachena
"Say 'pepapa' clearly."

Ndebele(LC): Ngakhuluma phephapha pakati
"I speak 'pepapa' in the middle."

Sotho(MC): Kebua phephapha phakisa
"Say 'pepapa' quickly."

For each combination of target vowel and following vowel context (e.g., target /e/ followed by the context vowel /i/), we selected for analysis three of the 10 possible types of trisyllables which contained that particular combination. If the target vowel was in V₁, then V₂ was the context vowel, and V₃ was either /a/ (e.g., /pepapa/), or was identical to V₂ (e.g., /pepapi/). When the context vowel was /a/, these two types were not distinct. If the target vowel was itself V₂, then V₃ was the context vowel, and V₁ was always /a/ (e.g., /papepi/). We will pool data from these three types of trisyllables, as no relevant effects are found when the types are considered separately (see Manuel, 1987).

C. Recordings
Subjects were told that the purpose of the experiment was to find out how speakers of different languages produce speech sounds. They were instructed to read five randomized lists of the 125 trisyllables, inserting each trisyllable into the appropriate carrier phrase. Subjects were asked to produce all syllables on a low tone, and to speak at a normal rate. Despite this instruction, some subjects spoke very rapidly, particularly the Sotho(MC) speakers (this may have been due to the semantic content of the carrier phrase or because, according to some of the subjects, in their culture there is a popular game in which high value is put on speaking rapidly).

The recordings were made in a sound-treated recording studio at the University of Zimbabwe. The speech was recorded on audio tape at 3 3/4 ips on a NAGRA tape recorder. At the end of each list, subjects were asked to repeat trisyllables that had obviously been misread.

We had hoped to record those Sotho(MC) vowels which were phonetically closest to the Shona(LC) and Ndebele(LC) mid vowels, /e/ and /o/. Generally, in Sotho(MC) all of the allophones of
both /e/ and /o/ are represented orthographically by e, and the allophones of /i/ and /u/ as o; sometimes diacritics are employed to distinguish the phonemic vowels. Before reading the test materials, the Sotho(MC) speakers were presented with the five orthographic vowels a, e, i, o, u, and asked how they pronounced those vowels in isolation. For every speaker the quality of the isolated e was judged to be [e], and that of o to be [o]. We also monitored the speakers as they read the lists of trisyllables, and generally judged orthographic e and o to be produced as [e] and [o], respectively. However, subsequent listening to the audio recordings revealed that for two of the Sotho(MC) speakers there was a very large variability in the production of orthographic e, ranging from [i] to [e], and one Sotho(MC) speaker produced a lower mid vowel, more like English /l/. This observation indicates that, for the vowel /e/, we cannot be certain whether subjects were consistently producing a single phonemic vowel, and if not, which one in a given case. Though this results in some noise in the data, it does not crucially affect the ability to test the hypotheses under consideration. First, as shown in Figure 1, both Sotho(MC) /e/ and /e/ have closer neighboring vowels than does the mid vowel /l/ of Shona(LC) and Ndebele(LC). Second, as will be discussed below, when we look at the context effects on target vowel /a/, we concentrate on the effects of the /l/ and /u/, with less attention to the mid-vowel contexts.

D. Acoustic measurements

The speech was low-pass filtered with a cutoff frequency of 5 kHz and digitized at a 10 kHz sampling rate. The first (F1) and second (F2) formant frequencies were measured using linear predictive (LPC) analysis in the ILS package. A 20 ms analysis window was moved in 5 ms increments over the trisyllable. The formant peaks for each analysis frame were calculated using the root-solving procedure.

Here we will report on measures of F1 and F2 that were made at two points in the target vowels.4 The first measurement point was made in the middle of the vowel. This point was selected by examining the waveform and formant tracks and choosing a point, in the middle region of the vowel, which seemed to be most clearly associated with maximal F1 values. It was expected that this point approximated the time of maximal jaw opening. We expected the anticipatory effects of the following vowel to be seen more clearly at the end than in the middle of the target vowel. Therefore the second measurement point, the end point, was made as close to the /p/ closure (as observed in the waveform) as possible but with plausible F1 and F2 values.5 This procedure yielded four values for each token (two formant values (F1, F2) by two measurement points (middle, end)).

Occasionally this procedure failed to yield F1 or F2 at one of the selected measurement points; in these cases a point immediately preceding or following the desired point was selected. For some vowels the entire region of interest failed to yield formant values (this was particularly problematic for F1) and an attempt was made to determine the values by examining smooth spectra in that region. Finally, for each speaker, the mean and standard deviation were calculated separately for each of the four measures for both target vowels. Values which were more than two standard deviations above or below the mean were omitted from further analysis. This last procedure resulted in a loss of from 0 to 3% of the values from each measure for any one subject. Thus, while five tokens of each type were recorded, not all (occasionally as few as two) were ultimately usable; some trisyllables were misread, and for others it was impossible to determine the formant frequencies. Since we have pooled data from three types of utterances, for each speaker, and for each combination of target vowel and context vowel, there are from 9 to 15 (with a mean of 13.4) tokens. The exception is when the context vowel was /a/, for which there are only two types of utterances, and from 7 to 10 (with a mean of 9.1) tokens per target for each speaker.

Results

A. Non-Context effects: Distribution of /a/ and /e/

Each speaker's average F1 and F2 values in the middle of target vowels /a/ and /e/ are plotted in the F1/F2 space in Figure 2. For comparison, values for these speakers' /i, o, u/ are also shown; no further analysis was done on these reference vowels. For all subjects except Sotho(MC) speaker 1, /e/ has a relatively high F2 and low F1 value. The general trend of a high F2 and low F1 for most of the speakers' /e/ vowels is not surprising, given the auditory impression that this usually was a fairly high vowel. As noted earlier, some Sotho(MC) orthographies are ambiguous as to whether orthographic e represents the phoneme /e/ or /l/, and among those orthographic traditions which do distinguish the two, the conventions on use of diacritics are different. It seems not unlikely that Sotho(MC) speaker 1 produced the lower phoneme, while the other two speakers produced the higher one.
Figure 2. Average F1 and F2 values for the middle of target vowels /a/ and /e/ for the nine subjects. Values for the vowels /i/, /o/, and /u/ are from the medial vowel in /pipi/, /popo/, and /pupo/, respectively, and are based on from two to five tokens each. Values for /a/ and /e/ are based on more contexts, as indicated in the text.

B. Expected effects on F1 and F2 of fronting/backing and raising/lowering contexts

In Figures 3a-c we show the expected acoustic effects of variation in production of the vowel /a/. In these figures, the subscript assigned to the vowel /a/ identifies the following vowel context. The locations of the points indicate schematically how the formant frequencies for the vowel /a/ are expected to be influenced by the different following vowel contexts. Fronting and backing of /a/ should cause changes in F2, as shown in Figure 3b. Raising of /a/ should primarily affect F1, with little effect on F2, as shown in Figure 3a. A combination of raising and front/back movement should be reflected in movement of both F1 and F2, as shown in Figure 3c. For target vowel /a/, we are interested in how much it moves toward the phonetic /e/ and /o/ spaces. All of the vowel contexts we used here (other than /a/ itself) are contexts which would potentially move target /a/ into the phonetic spaces of interest.

The potential context effects for the target vowel /e/ are shown in Figures 4a-c. We looked at target /e/ as a function of two context vowels (/i,u/) which might raise the target /e/, three vowels which might back /e/ (/a,u,o/), one fronting context (/i/), and one lowering context /a/. As shown in Figure 4a, for a front vowel like /e/, variation along the front/back dimension should be reflected acoustically in F2: at a given height, the more front the vowel is, the higher its F2. Changes in
the height of /e/ have effects both in F1 and F2, since in general the higher a front vowel is, the lower its F1 and the higher its F2 (Fant, 1960). Thus if /e/ is sensitive to only the height of a following vowel, we might expect to see its contextually induced variants as shown in Figure 4b. Note that in this case a following /u/ actually results in a variety of /e/ which has a higher F2 than if the target /e/ is followed by /a/. Finally, if /e/ is affected by both the front/back and height character of the following vowel, we might expect a pattern somewhat like that of Figure 4c.

![Figure 3](image-url)

*Figure 3. Expected acoustic correlates of context-induced variation in the target vowel /a/: a) fronting-backing; b) raising; c) combination of raising and fronting-backing. The subscripts indicate the quality of the following vowel; for example, a₁ indicates the vowel /a/ when it is followed by the vowel /i/.

![Figure 4](image-url)

*Figure 4. Expected acoustic correlates of context-induced variation in the target vowel /e/: a) fronting-backing; b) raising-lowering; c) combination of raising and fronting-backing. Subscripts indicate the quality of the following context vowel (e.g., e₁ indicates the vowel /e/ when it is followed by the vowel /i/).
C. Obtained effects of vowel context: General

To get an overall impression of the anticipatory coarticulation effects, we averaged the data across the three speakers of each language. These averaged data are plotted in the F1/F2 space as a function of context and measurement point in Figures 5a-c (individual subject data are given in Table 1). In Figures 5a-c, for each language, there are two scatters of points, one scatter for the vowel /a/ (in the lower portion of the graphs), and one for the vowel /e/ (in the upper left portion of the graph). Each of these scatters is composed of two smaller sets of data, one for measurements made in the middle of the vowel (points enclosed by inner loops), and one for measurements made at the end of the vowel (points enclosed by outer loops). Each symbol represents the context in which the target vowel occurred (e.g., filled squares indicate values for the target vowels followed by the context vowel /i/).

D. Context effects on the vowel /a/

Beginning with the Ndebele(LC) data shown in Figure 5a, we see that all F1 values for target /a/ are lower at the end than in the middle of the vowel, presumably due to the labial closure. The consonant closure generally had a lowering effect on F2 as well, as can be seen by comparing the midpoint and endpoint F2 values for target vowel /a/ followed by context vowel /a/ (open circles).

In general, for Ndebele(LC), the following vowel context had a large effect on the target vowel /a/. Relative to the /a/ context, high vowels /i,u/ tend to lower F1 of target vowel /a/. The mid-vowel contexts /e,o/ also lower F1 of target vowel /a/, though to a lesser extent. Front vowels /i,e/ raise F2, and the back rounded vowels /o,u/ lower F2. Even in the middle of /a/ there is an effect of the following vowel. The effects of context on F2 increase as the measurement point moves closer to the end of the target vowel, and therefore to the contextual vowel itself. The vowel context effect on F1 does not appear to be much larger at the end than in the middle of the target vowel.

The vowel-context effects for Shona(LC) and Sotho(MC) are shown in Figures 5b and 5c, respectively. The influence of the context vowel is smaller in these languages, particularly in Sotho(MC), where there is very little effect of following vowel context at the steady-state point. In Sotho(MC) there is essentially no difference in F2 for target /a/ produced in the contexts of a following /a/, /o/, and /u/, whereas in both Ndebele(LC) and Shona(LC) there are some differences between these contexts, although they are small in the steady state. At later points in the target vowel, a front/back effect emerges for Sotho(MC), but it is much smaller than that seen in Ndebele(LC) or Shona(LC). Height effects in Sotho(MC) also appear to be relatively small. In general, Sotho(MC) /a/ appears to spread less into the F1/F2 space than does /a/ in Ndebele(LC) or in Shona(LC). In fact, most of the movement of Sotho(MC) /a/ is straight up the middle of the vowel space, and has the form expected if this movement were due mostly to the intervocalic /p/. In what follows, we will examine the F1 and F2 effects separately for the target vowel /a/.

Figure 5. Acoustic effects of following context vowel on the target vowels /a/ and /e/ in the three languages. Inner loops enclose data from the middle of the target vowels, outer loops enclose measurements made at the end of the target vowels. Data are averaged over the three speakers of each language.
The Role of Contrast in Limiting Vowel-to-vowel Coarticulation in Different Languages

Table 1. Individual subject data.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>MID</th>
<th>END</th>
<th>MID</th>
<th>END</th>
<th>MID</th>
<th>END</th>
<th>MID</th>
<th>END</th>
<th>MID</th>
<th>END</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sotho(MC) #1</td>
<td>1256</td>
<td>1266</td>
<td>1212</td>
<td>1199</td>
<td>1248</td>
<td>1136</td>
<td>1227</td>
<td>1130</td>
<td>1218</td>
<td>1131</td>
</tr>
<tr>
<td>Sotho(MC) #2</td>
<td>1169</td>
<td>1151</td>
<td>1160</td>
<td>1180</td>
<td>1118</td>
<td>1017</td>
<td>1108</td>
<td>1021</td>
<td>1142</td>
<td>1063</td>
</tr>
<tr>
<td>Sotho(MC) #3</td>
<td>1348</td>
<td>1281</td>
<td>1339</td>
<td>1272</td>
<td>1275</td>
<td>1169</td>
<td>1296</td>
<td>1174</td>
<td>1282</td>
<td>1111</td>
</tr>
<tr>
<td>Shona(LC) #1</td>
<td>1397</td>
<td>1384</td>
<td>1380</td>
<td>1353</td>
<td>1346</td>
<td>1227</td>
<td>1314</td>
<td>1183</td>
<td>1276</td>
<td>1157</td>
</tr>
<tr>
<td>Shona(LC) #2</td>
<td>1408</td>
<td>1348</td>
<td>1384</td>
<td>1342</td>
<td>1361</td>
<td>1230</td>
<td>1310</td>
<td>1192</td>
<td>1301</td>
<td>1196</td>
</tr>
<tr>
<td>Shona(LC) #3</td>
<td>1386</td>
<td>1313</td>
<td>1344</td>
<td>1298</td>
<td>1208</td>
<td>1128</td>
<td>1260</td>
<td>1141</td>
<td>1224</td>
<td>1124</td>
</tr>
<tr>
<td>Ndebele(LC) #1</td>
<td>1235</td>
<td>1295</td>
<td>1209</td>
<td>1234</td>
<td>1200</td>
<td>1070</td>
<td>1145</td>
<td>901</td>
<td>1129</td>
<td>892</td>
</tr>
<tr>
<td>Ndebele(LC) #2</td>
<td>1413</td>
<td>1458</td>
<td>1369</td>
<td>1385</td>
<td>1301</td>
<td>1134</td>
<td>1244</td>
<td>1029</td>
<td>1221</td>
<td>997</td>
</tr>
<tr>
<td>Ndebele(LC) #3</td>
<td>1476</td>
<td>1458</td>
<td>1470</td>
<td>1444</td>
<td>1322</td>
<td>1176</td>
<td>1350</td>
<td>1082</td>
<td>1353</td>
<td>1098</td>
</tr>
</tbody>
</table>

D.1 Front/back effects (F2) on target /a/. To simplify the analysis of front/back coarticulation, we concentrated on the two contexts that tended to yield the highest and lowest F2 values, that is /i/ and /u/, respectively. The F2 data for /a/ followed by /i/ and by /u/, averaged across speakers within each language, are shown in Figures 6a-c. In these figures, which can be thought of as highly schematic formant trajectory plots, F2 is shown on the vertical axis, and the two measurement points...
are shown on the horizontal axis. Individual subject data are shown in a similar fashion in Figures 6d-l.

Again, it is clear that on average, Ndebele(LC) speakers displayed a large amount of front/back coarticulation. The difference in the /i - u/ contexts was 141 Hz in the middle of the vowel, and 408 Hz at the end of the target vowel. These Ndebele(LC) data were submitted to an analysis of variance with repeated measure factors of the /i/ vs. /u/ contexts and the two measurement points. While the individual speaker data show that Ndebele(LC) speaker 2 contributes more to the average /i - u/ difference than do the other two Ndebele(LC) speakers, the front/back contrast is significant \[ F(1,2) = 107.6, p < 0.01 \]. The front/back contrast is much larger at the end of the vowel, and this was reflected in an interaction of the front/back effect with the measurement point \[ F(1,2) = 238.5, p < 0.01 \]. Simple main effects tests confirm that the front/back contrast is significant at the 0.05 level in the middle of the vowel, and at the 0.01 level at the end of the vowel.

For Shona(LC), the /i - u/ difference is about the same (130 Hz) as Ndebele(LC) in the middle of the vowel, but considerably smaller (189 Hz) than Ndebele(LC) at the end of the vowel. The Shona(LC) speakers all showed patterns similar to each other, and the front/back contrast was in fact significant \[ F(1,2) = 112, p < .01 \]. Although each Shona(LC) speaker had larger /i - u/ differences at the end of the vowel than in the middle, the interaction between measurement point and the /i - u/ difference was not statistically significant \[ F(1,2) = 6.16, p > 0.05 \].

On average, the Sotho(MC) speakers showed much less of an /i - u/ difference (44 Hz) in the middle of the target vowel than did speakers of either Shona(LC) or Ndebele(LC). While each speaker did show at least some tendency for the /i/ context to give higher F2 values than the /u/ context, no Sotho(MC) speaker showed as large a difference as even the smallest difference found among the Shona(LC) and Ndebele(LC) speakers. Furthermore, while there was a main effect of context \[ F(1,2) = 24.9, p < 0.05 \], simple main effects indicated that in the middle of the vowel, the /i - u/ contrast was not significant for Sotho(MC) speakers \[ F(1,2) = 14.1, p > 0.05 \]. At the end of the vowel, Sotho(MC) speakers did have a substantial /i - u/ difference (130 Hz), and this was statistically significant \[ F(1,2) = 30.4, p < 0.05 \]. Even so, this difference at the end of the vowel was only as large as the /i - u/ difference in the middle of the Shona(LC) /a/ vowel.

D.2 Context effects on F1 of target /a/. The averaged data shown in F1/F2 plots in Figures 5a-c indicate that there is some effect of context on the F1 value of target vowel /a/. For example, in the middle of target vowel /a/ for the Ndebele(LC) speakers, the high-vowel contexts (/i/ and /u/) lower F1 by about 65 Hz, relative to target /a/ followed by the low vowel /a/. The mid vowels /e/ and /o/ lower F1 by a smaller amount, about 30 Hz. A similar pattern can be seen at the end of the vowel, with additional lowering of F1 for the /o/ and /u/ contexts, compared to the /e/ and /i/ contexts. In the other languages, there is also a general tendency for F1 to lower when the target vowel is followed by high vowels. However, the details of the patterns are somewhat different. For example, for Shona(LC) speakers, at the end of target vowel /a/, the /u/ context has a higher F1 than does the /i/ context, whereas the opposite pattern was found in Ndebele(LC). In Sotho(MC), a following /a/ context actually gives a lower F1 value than does a following /u/. As can be seen in Table 1, there was a fair amount of subject-to-subject variability in the F1 data. This variability may have been due in part to general problems of obtaining F1 data, as discussed above and in Note 5.

As we did for F2, we simplified the F1 analysis by focusing on only some of the possible context contrasts. In this case we compared the average of the two high-vowel contexts (/i/ and /u/) to the data for the low-vowel context /a/. The results, averaged across the speakers within each language, are shown in Figures 7a-c.

For Ndebele(LC), the average difference between the high-vowel and low-vowel contexts was 66 Hz in the middle of the vowel, and 62 Hz at the end of the vowel. The average Shona(LC) data also show differences between the high-vowel and low-vowel contexts: 21 Hz in the middle of the vowel and 61 Hz at the end of the vowel. Despite the overall tendency for F1 to be lower in the high-vowel contexts, neither Ndebele(LC) nor Shona(LC) show a statistically significant F1 effect for high- vs. low-vowel contexts. However, as there are only three speakers for each language, the statistical tests are particularly susceptible to speaker-to-speaker variability, which was rather large for the F1 measure in Ndebele(LC) and Shona(LC). When we combine the data from all six speakers of these two languages, the high vs. low effect does reach significance \[ F(1,5) = 10.04, p < 0.05 \].
The Role of Contrast in Limiting Vowel-to-vowel Coarticulation in Different Languages

Figure 6. F2 of the target vowel /a/ followed by the front vowel /i/ (filled squares) and by the back vowel /u/ (open squares). Data are shown for measurements made in the middle and end of the target vowel. Plots a-c are averaged over the three speakers of each of the languages. Plots d-l are individual subject data.
Figure 7. Open circles are F1 values for target vowel /a/ in the context of a following low vowel (/a/), and filled squares show average of F1 value from the /i/ and /u/ contexts for target vowel /a/. Plots a-c are averaged over the three speakers of each of the languages. Plots d-l are individual subject data.
For the Sotho(MC) speakers, there was very little difference in the high and low contexts, either in the middle of the vowel (which was 7 Hz in the wrong direction) or at the end of the vowel, where the high-vowel contexts decreased F1 by only 13 Hz. In this case, the average data represent very well the individual subject data, as every Sotho(MC) speaker showed the same pattern (Figures 7f, i, l). Thus, the very small effects, which reversed in direction from the middle to the end of the vowel, gave an interaction of the context and measurement point factors \( F(1,2) = 226.7, p < 0.01 \), and significant effects of the high/low context at both the middle of the vowel \( F(1,2) = 23, p < 0.05 \) and end of the vowel \( F(1,2) = 56.3, p < 0.05 \).

E. Context effects on target vowel /ɛ/

Referring back to Figures 5a-c, we see that compared to target vowel /a/, target vowel /ɛ/ shows much less F1 fall as a function of movement toward the oral tract closure for /p/ than does target vowel /a/. This relatively small amount of F1 fall is presumably because /ɛ/ is already a high vowel, with a lower F1, than target vowel /a/.

Perhaps the most striking aspect of the /ɛ/ data, and of crucial relevance here, is that in none of the languages do the /ɛ/ values encroach very much on the phonetic /ɛ/ space. The vowel /ɛ/ has a relatively high F1 value even when followed by /a/ in Ndebele(LC) and Shona(LC). The production of flanking /p/ consonants may have encouraged the subjects to maintain a relatively high jaw position during the target vowel, thus limiting or canceling the lowering effects of a following low vowel.

To show more clearly the context effects that do exist, in Figures 8a-c we have expanded the upper left portion of the F1/F2 plots that were shown in Figures 5a-c. The points enclosed by loops are from measures made in the middle of target vowel /ɛ/, and the other points are from the measure at the end of the vowel. Again, beginning with Ndebele(LC), in the middle of target /ɛ/ there is very little effect of context, though /i/ has a tendency to both lower F1 and raise F2. A similar but smaller effect can be seen in the middle of Shona(LC) /ɛ/. In the middle of Sotho(MC) target /ɛ/, both a following /i/ and /u/ give lower F1 and higher F2 values than a following /ɛ/ context does. This pattern was found for Sotho(MC) speakers 1 and 2, and is consistent with what we would expect from a phonological rule that raises /ɛ/ when it is followed by a high vowel (see Figure 4b).

If we look at the measurements made at the end of target /ɛ/, a front/back effect emerges in all languages, such that for a given height of the context vowel, the back vowel contexts result in lower F2 values. This effect is strongest in Ndebele(LC). At the end of the vowel, on average each of the languages shows a height effect in that high front contexts yield lower F1 values than mid or low back vowels. However, at a given height specification, the back vowels give lower F1 values. This may be due to rounding for the /o/ and for /u/ contexts.

![Figure 8](image_url)

Figure 8. Expanded view of Figure 5, showing the acoustic effects of following-vowel context on the /ɛ/. Loops enclose data from the middle of the target vowels; remaining points show measurements made at the end of the target vowels. Data are averaged over the three speakers of each language.
We are primarily interested in seeing if the Ndebele(LC) and Shona(LC) /e/ show more movement into the /e/ phonetic space, a space that is used phonemically in Sotho(MC), but not in the other languages. The only vowel context which would be expected to cause both lowering and backing is /a/. Consequently, we focused our attention on the F1 and F2 values of /e/ followed by /a/ vs. /e/ followed by /e/.

The difference in the /e - a/ contexts for F1 was small, both in the middle and end of the target vowel, and did not differ substantially from language to language, especially when individual speaker differences were taken into account. In general, there was a high degree of speaker-to-speaker variability with respect to anticipatory coarticulation effects for target vowel /e/. The only language for which there was a significant effect of context on F1 was Shona(LC) [F(1,2) = 41.3, p < 0.05], but the /e - a/ context difference was actually smallest, and quite negligible, in this language (on average only 3 Hz in the middle of the vowel and 15 Hz at the end of the vowel for the /e/ vs. /a/ contexts). For F2, none of the languages had a statistically significant /e - a/ context effect, even though all Ndebele(LC) subjects had at least a 140-Hz higher F2 value for the /e/ than for the /a/ context, measured at the end of the vowel. On average, Shona(LC) had an 81-Hz, and Sotho(MC) a 50-Hz /e - a/ context difference at the end of the vowel. Possibly, with more speakers of each language, the context effects on F2, and even the small context effects on F1, would prove to be significant. When we pooled the data for the nine speakers together, there was a significant contrast of the /e/ and /a/ contexts at the end of the vowel for both F1 [F(1,8) = 14.5, p < 0.01] and F2 [F(1,8) = 9.4, p < 0.05].

Discussion

A. Inventories of contrast as predictors of output constraints

Our prediction was that Sotho(MC) speakers would exhibit smaller anticipatory coarticulation effects for /a/ and /e/ than would Shona(LC) or Ndebele(LC) speakers, because too much anticipation of an upcoming sound in Sotho(MC) would move the articulators (and therefore acoustics and perception) into competing distinctive vowel spaces. For example, we assumed that /a/ would have a tighter output constraint in Sotho(MC) than in Shona(LC) or Ndebele(LC), because of the proximity of Sotho(MC) /a/ to other phonemic Sotho(MC) vowels /e, a/.

When the target vowel was /a/, our prediction was borne out. Even when we take into account speaker-to-speaker variability, Shona(LC) and Ndebele(LC) show larger anticipatory coarticulation effects on /a/ than does Sotho(MC). Most of the Sotho(MC) /a/ movement is straight up the middle of the vowel space, and is mostly due to the oral tract closure associated with the intervocalic /p/. By limiting lateral movement, speakers of Sotho(MC) avoid excessive movement into the phonemic /e/ and /a/ spaces of this language.

We had expected to see similar effects for the target vowel /e/, with Shona(LC) and Ndebele(LC) exhibiting more movement of /e/ toward the phonetic /e/ space. However, none of the languages studied showed much movement into the /e/ space, and what effects there were did not seem to be less robust in Sotho(MC) than in the other languages. One reason for this result might be that the target vowel is flanked on both sides by the consonant /l/. This /l/ may itself impose constraints on the height of the jaw, and may limit the anticipatory lowering movement of the following /a/ during the target vowel /e/.

Thus far we have focused on the direct comparison of Sotho(MC), Shona(LC) and Ndebele(LC), which after all, do not differ very much in terms of their respective spacing of phonemic vowels. Actually, all of these languages might be expected to have relatively large coarticulatory effects as compared to a language like English (Very Crowded), which has many more vowels crowded into the vowel space. In the languages in the present study, the effects of V2 on V1 are clearly observable in the V1 C transitions, but they also extend into the middle portions of V1, and in Shona(LC) and Ndebele(LC) these effects are remarkably large. Examples of vowel-to-vowel effects in Ndebele(LC), Sotho(MC), and English(VC) can be seen Figure 9, which compares the medial /a/ in /papapa/ and /popapi/ for a single token for one speaker each of the three languages (see also Manuel and Krakow, 1984). The effects in English(VC) are very small, as we would expect given the proximity of English(VC) /a/ and /a/. Somewhat more coarticulation is seen in Sotho(MC), in which the nearest front vowel to /a/ is /e/. Finally, it is not surprising that Ndebele(LC) shows very extensive effects, since the next closest front vowel to /a/ is /e/ in Ndebele(LC).

Given these observations, we fully expect that anticipatory coarticulation in Shona(LC) and Ndebele(LC) may extend further back into the vowel, and perhaps is present even at the
beginning of V1. Interestingly, English schwa, a phoneme that is known to behave as if it had very lax output constraints, allows quite extensive anticipatory coarticulation. As Magen (1989) has recently shown for English trisyllables like /bababi/ and /bababa/, the effects of the third vowel are seen throughout a medial schwa and can extend into the first vowel.

Strictly speaking, since we have not tested a number of seven-vowel languages against a number of five-vowel languages, our data alone do not allow us to generalize to other five- vs. seven-vowel comparisons. It is possible, then, that our results have nothing to do with output constraints or phonemic contrasts, but are due to some other (perhaps arbitrary) fact about the languages studied. Further testing of other languages is necessary to determine whether or not the present data reflect a general tendency, or are due to a fortuitous sampling accident.

Having noted this qualification, we observe that similar results have been found in other studies that reflect on the effect of contrast in constraining coarticulation (e.g., Lubker & Gay, 1982; Magen, 1984; Manuel & Krakow, 1984; and Öhman, 1966). Note, however, Nartey's (1979) cross-linguistic analysis of fricatives found that variation in fricative production was not correlated with the number or distribution of contrastive fricatives in a language. It may be that the precision needed to make fricatives, or the more categorical nature of consonant perception or production, or perhaps simply arbitrary amounts of pickiness in some of the languages studied, is such that any added constraint from the system of contrasts is negligible.

Figure 9. First and second formants for the middle vowel /a/ in a single token of /papap/ and /papapi/ from a speaker of each of the languages indicated. The rising F2, in anticipation of following /u/, is particularly striking for the Ndebele speaker. The Ndebele and Sotho tokens shown here are part of the present general analysis. The English token was recorded specifically for this illustration.
B. Other factors affecting output constraints

Are output constraints determined solely by the distribution of contrastive phones in a language? It would seem not. In the present study there appear to be differences between the 2 five-vowel languages in amount of coarticulation. Ndebele(LC) seems to have more front/back coarticulation on /a/ than does Shona(LC), and this is not obviously predicted by anything else we know about the languages. One way of thinking about the various contributions to output constraints is that minimally phonemes must have some audible, distinctive output, and that languages are free to restrict the output of particular phonemes even further. Furthermore (1983) has pointed out that languages may choose to be more or less particular about how they make their phonemes. Our claim here, however, is that there are likely to be general constraints on the lack of fastidiousness.

There may be a number of other factors that contribute to acceptability, and the role of contrast may be to set maximal laxity for output constraints. One such factor is probably formal vs. informal styles of speech. In formal situations, people are generally more careful in their speech. But what is “careful” if not a more precise, narrowed range of production for particular phonemes? In situations where communication is difficult (e.g., noisy conditions, talking to children, nonnative speakers, or the hearing-impaired) people tend to “overarticulate” (Lindblom & Engstrand, 1989; Lindblom & MacNeilage, 1986). Overarticulation can be partially modeled as a slowing of articulation, but in all probability also involves a compression of the range of acceptable productions of phonemes, in terms of their spatial characteristics (Picheny, Durlach, & Braida, 1985; 1986). The very fact that people can vary their range of productions at will, and in formal or careful speech conditions tend to narrow those ranges toward clear exemplars of the phonemes, implies that at some level speakers have an awareness of the notion “best production” and the range of acceptable productions.

Are output constraints ever violated? Obviously they are, since phonological processes such as nasal place assimilation (e.g., input > [impol]) are common and neutralize the underlying difference between contrastive phonemes. However common these neutralizations are, though, it is equally obvious that they are not so rampant in any one language so as to lead to widespread or wholesale loss of phonemic contrast. Of course, another way of looking at these phenomena is to say that the output constraints themselves have been relaxed, not that the output constraints have been violated. It is not clear what, if any, data would distinguish between these two hypotheses. In any case, as pointed out by Martinet (1952, pp. 129), there are a variety of forces at play in speech, and sometimes the forces which push or pull one phoneme into the territory of another “may simply be more powerful than the functional factors working toward conservation.”

C. Incorporating output constraints into models of coarticulation

The concept of output constraints, and the role of contrast in setting output constraints, can potentially be of value in trying to understand particular instances of coarticulatory behavior, and can be used to try to restrict models of coarticulation. In this section we briefly explore how this concept might be incorporated into more formal models of coarticulation. Note that while the following discussion is limited to only a few types of models, the role of output constraints is (or should be) a concern for any coarticulation model.

We have suggested elsewhere (Manuel, 1987a; 1987b) that output constraints could be thought of as target spaces in some appropriate dimension(s) and that speakers generally try to move smoothly from one space to the next, crucially always trying to maximize efficiency (pick the easiest overall route, according to some measure of “ease” of articulation). The idea was that the movement from target to target is affected by the narrowness of the target spaces themselves: extremely narrow targets don’t allow for much variability in the movement from one target to the next, whereas large targets allow various trajectories through a given space. This concept is shown schematically in Figure 10, which demonstrates the effect of the narrowness of the medial target on a non-speech “connect the circles” drawing task. The task is to start in Circle A, move through Circle B, and end up in Circle C or D. When Circle B is quite small (analogous to a very restrictive output constraint), the trajectory from A to B is rather insensitive to whether the following target is C or D. In contrast, when Circle B is large, the trajectory from A to B is highly affected by the location of the following target, that is, whether it is C or D. A similar proposal, suggesting interpolation between and through targets of various sizes, has been made by Keating (1990). Working within a connectionist...
framework, Jordan (1990) has recently developed more explicit models for comparing the effect of loose output constraints (*don't care* conditions on outputs) with strict output constraints (*strongly care* conditions on outputs). Jordan's model learns a path through several target specifications, and Jordan demonstrates that varying these constraints affects just which path his model learns.

![Diagram](image)

*Figure 10.* When Target B is large (upper panel), the trajectory from A to B is affected by the location of the next Target (C vs. D). In contrast, when Target B is small (lower panel), the trajectory from A to B is more restricted, and is minimally influenced by the location of the next target.
The above accounts assume, at least implicitly, that given a starting point A and two following goals B and C, the actor/speaker somehow calculates an overall path through the two upcoming goals. But it may be that, for motor behaviors in general, the actual surface path is the result of combining an underlying invariant command to move from the starting point A to Goal B, and an underlying invariant command to move from Goal B to Goal C.

The idea that surface paths may reflect simultaneous input from distinct, invariant commands to move to different targets has been applied to speech, most notably in the coproduction models of, for example, Bell-Berti and Harris (1982), Brownman and Goldstein (1990), and Fowler (1981; 1986). In coproduction models, coarticulation is achieved by allowing a particular articulator (or articulatory system) to respond simultaneously to invariant commands associated with adjacent or neighboring phones. For example, the anterior /k/ closure in /aki/ (vs. the more posterior closure for /aka/) can be described by assuming that while the tongue dorsum is still responding to a command (associated with the /k/) to make a velar closure, it also begins responding to commands, associated with the goal of making the following /I/, to move anteriorly. The actual movement of the tongue dorsum reflects both of these inputs. While the basic concept of simultaneous response to two phones is not new, it has recently been explicitly modeled (e.g., Saltzman, Rubin, Goldstein, & Brownman, 1987).

In coproduction models, varying degrees of coarticulatory effects, such as those seen in different languages, can be achieved in two ways. First, the amount of temporal overlap of the commands can be increased. For example, given a sequence of phones XY, Y will have a greater effect on the production of X the more the commands for X and Y overlap. Second, the particular weights used to combine the various commands can be varied. Given simultaneous input from X and Y, the inputs could average, they could sum, or they could combine in some other linear or nonlinear fashion such that Y has a greater or smaller effect on the production of X. For example, in /akil/ there is a relatively lax input constraint on exactly where contact is made on the roof of the mouth for the /k/, and this constraint does not heavily suppress the forward movement for the following /I/ vowel. At the same time, since /k/ is a stop consonant, there is a strict constraint on the degree of oral tract constriction, and this constraint suppresses any contrary gestures from the following /I/ which would preclude making a complete oral closure.

The present results for Ndebele(LC), Shona(LC), and Sotho(MC) show that coarticulation can be both temporally and spatially very extensive. From the point of view of coproduction models, if anticipatory coarticulatory effects are found early in the first vowel of a V₁CV₂ utterance, we can conclude that the commands for the second vowel begin (at least) that early. In addition, since the first vowel clearly dominates production at least until the consonant is approached, those commands associated with the second vowel must have a lesser effect on the surface trajectory in that time period than do the commands associated with the first vowel itself. The fact that V₂ has a greater or lesser effect on V₁ in different languages could, in a mechanical sense, be modeled as being due to different tolerances or output constraints for particular phonetic gestures. These tolerances, which are predictable (at least in part) from certain general principles, lead to language-dependent amounts of suppression of the V₂ gestures in different languages. That is, the determination of the combinatorial algorithm for neighboring phones is affected by requirements for the maintenance of intelligible speech, i.e., maintaining distinctions between phones. Those requirements vary from one
speaking style or condition to the next, and from language to language, partially because languages have different systems of contrast. Contrast affects output constraints, and output constraints determine how gestures associated with neighboring phonetic units are combined.

**SUMMARY**

The data presented here show that the vowel /a/ is more susceptible to anticipatory coarticulation with a following transconsonantal vowel in Shona(LC) and Ndebele(LC), which have no near phonemic neighbors to /a/, than in Sotho(MC), which does have relatively near neighboring and contrasting phonemic vowels. This result is consistent with the idea that coarticulation is limited by output constraints on phones, and that these output constraints are determined, in part, by the need to maintain phonological distinctions in a language.

**REFERENCES**


FOOTNOTES

1Now at Wayne State University, Department of Communication Disorders and Sciences, Detroit, MI.
2As is well known, in a word like *can't* there may be no nasal murmur in the signal. That is, there may be no period of time where there is simultaneously (1) a wide velopharyngeal opening, (2) oral closure, and (3) substantial vocal fold vibration. In the absence of a nasal murmur, the nasal quality of the vowel itself can distinguish *cat* from *can't*.
3Ndebele and Sotho contrast voiceless ejective [p'] (orthographic *p*) with voiceless aspirated [ph] (orthographic *ph*). Consistent with the orthographic traditions of the three languages, the stimulus lists for these two languages used *ph*, while those for Shona used *p*, which also represents an aspirated [ph].
4In Manuel (1987a) we report data from an additional measurement point, made at 20 ms before the end of the target vowel. While there are differences in the values obtained at this point and the measurement point made at the end of the vowel, the conclusions reached are not affected by the omission of those data.
5At points very close to the end of the vowel, where there is a rapid change in frequency, and amplitude is dropping, the LPC technique sometimes picks spurious values for the formants. This was particularly true for F1, which tended to be "lost" earlier than F2. Occasionally the end point was actually made as much as 20 ms, though more often it was 5 or 10 ms, from closure. In addition, for some speakers F1 was hard to determine even in the middle of the vowel, particularly for the vowel /a/, in which F1 had a wide bandwidth, perhaps due to tracheal coupling. Note that the /p/ in these utterances was often heavily aspirated.

speakers of Spanish (a language with vowels which are well-spread apart in the acoustic and articulatory space) do not show more token-to-token variability for vowel height than do speakers of English, a language with a more crowded vowel space.

2It may well be that the range of normal token-to-token variability (vs. coarticulation induced variability) does not put sufficient pressure on output constraints to test the hypothesis that output constraints vary as a function of phoneme distribution. For vowels spoken in a single context, Flege (1989) found that