Journal of Phonetics (2002) **30,** 591–627
doi:10.1006/jpho.2002.0177
Available online at http://www.idealibrary.com on IDEAL®

Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates

Patrice Speeter Beddor, James D. Harnsberger* and Stephanie Lindemann†

Department of Linguistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.

Received 24th July 2001, and accepted 6th February 2002

Three experiments tested the hypothesis that V-to-V coarticulatory organization differs in Shona and English, and that Shona- and English-speaking listeners' sensitivity to V-to-V coarticulatory effects is correspondingly language-specific. An acoustic study of Shona and English CV'CVCV trisyllables (Experiment 1) showed that the two languages differ in the carryover *vs.* anticipatory influences of stressed and unstressed vowels on each other. In 4IAX discrimination tests in which both Shona and English coarticulatory effects were spliced into different coarticulatory contexts (Experiment 2), Shona and English listeners perceptually compensated more (i.e., attributed more of a vowel's acoustic properties to its coarticulatory context in targeted test trials) for effects that were consistent with their linguistic experience. Similarly, when these listeners identified synthetic target vowels embedded into different vowel contexts (Experiment 3), Shona listeners compensated more (i.e., showed larger category boundary shifts) for the vowel contexts that triggered larger acoustic influences in the production study. English listeners' boundary shifts were more complicated but, when these data were combined with those from a follow-up identification study, they showed the perceptual shifts expected on the basis of the English coarticulatory findings. Overall, the relation between the production and perception data suggests that listeners are attuned to native-language coarticulatory patterns.

1. Introduction

The impetus for this cross-language investigation rests in two broad experimental findings, one from speech production and the other from speech perception, both

Address correspondence to: P. S. Beddor. E-mail: beddor@umich.edu.
*Current affiliation: Department of Communication Sciences and Disorders, University of Florida, Gainesville, FL, U.S.A.
†Current affiliation: Department of Applied Linguistics and ESL, Georgia State University, Atlanta, GA, U.S.A.

figuring prominently in the coarticulation literature. The findings from speech production research reveal systematic differences between languages in the spatiotemporal characteristics of coarticulation (see Manuel, 1999, for a review). Our emphasis is on language-specific patterns of vowel-to-vowel (V-to-V) coarticulation. Such patterns have been attributed in part to a language's consonant and vowel inventories. In his classic spectrographic study of VCV sequences in three languages, Öhman (1966) found that the *F*2 values of target vowels varied more due to vowel context in English and Swedish than in Russian, a finding later replicated on a considerably larger scale for English and Russian (Choi & Keating, 1991; Evans-Romaine, 1998). Öhman attributed the coarticulatory differences to the languages' consonant systems, arguing that the requirements on the tongue body imposed by contrastive palatalization in Russian, but not in English or Swedish, restricted transconsonantal coarticulation in Russian. Consonant restrictions on V-to-V coarticulation have also been reported by Recasens, who found less V-to-V coarticulation (i.e., spatially or spectrally smaller and temporally shorter effects) across the velarized lateral of Catalan than across the "clear" lateral of Spanish (Recasens, 1987) and German (Recasens, Pallarès & Fontdevila, 1998). Recasens and colleagues ascribed the coarticulatory differences to different lingual constraints for these laterals.

A language's system of vowel contrasts also influences V-to-V effects. A general pattern in the data reported in the literature is that languages with larger vowel systems tend to exhibit weaker V-to-V coarticulatory effects than those with smaller systems. Weaker effects have been shown for English compared to the much smaller five-vowel systems of Shona, Swahili (Manuel & Krakow, 1984), and Japanese (Magen, 1984); weaker effects also hold for Sotho, with seven vowels, compared to Shona and Ndebele, with five vowels (Manuel, 1990). Such coarticulatory differences are particularly apparent in the most (least) populated regions of these languages' vowel systems, in keeping with the view that contrast preservation restricts variability in dense regions of articulatory or acoustic space (e.g., Manuel, 1990; Marchal & Hardcastle, 1993) and more generally consistent with the notion that articulatory variation is constrained by listener-oriented factors (e.g., Lindblom, 1990; Lindblom, Guion, Hura, Moon & Willerman, 1995). Beyond segmental inventory constraints, other factors operable in VCV coarticulatory organization include language-specific phonological processes such as consonant or vowel harmony (e.g., Boyce, 1990) and prosodic properties, including stress (Fowler, 1981; Farnetani, 1990; Majors, 1998) and timing patterns (Smith, 1995).

The second basic finding motivating the current set of experiments, this time from speech perception research, is that listeners are sensitive to coarticulatory variability. This sensitivity has been clearly demonstrated for the influences of adjacent vowels on consonants, and *vice versa* (see Rosner & Pickering, 1994, for a review). The smaller set of studies investigating perception of long-distance V-to-V effects provides more mixed results, but the overall picture that emerges is that listeners use information from a context vowel in making judgments about a target vowel. A strong test of listeners' use of coarticulatory information is whether listeners can identify a deleted target vowel on the basis of a coarticulated vowel. Lehiste & Shockey (1972) presented excised portions of original V_1CV_2 tokens (where V = /i æ a u/) to American English listeners and found that they were not able to identify deleted V_1 based on V_2 nor deleted V_2 from V_1. However, when V_1 was /ə/, English listeners (Alfonso & Baer, 1982) and Dutch listeners (van Heuven & Dupuis, 1991) identified V_2 at better than chance accuracy on the basis of the anticipatory coarticulatory information in /ə/. V-to-V coarticulatory information also speeds listeners' reaction time: English listeners (Martin & Bunnell, 1981, 1982) and Shona listeners (Manuel, 1987) identified V_2 faster and with fewer errors when V_1 provided coarticulatorily appropriate information (e.g., [$u_a$ba], where [u_a] = [u] from [uba]) than when V_1 had been cross-spliced to provide inappropriate coarticulatory information (e.g., [$u_i$ba], where [u_i] = [u] from [ubi]).

Other perceptual findings suggest that listeners not only use coarticulatory influences from a context vowel to facilitate target vowel identification, but they accommodate or "compensate" for such influences on the target. Fowler (1981; Fowler & Smith, 1986) used cross-spliced stimuli similar to those for the reaction time studies described above, and paired the stimuli in a 4IAX discrimination paradigm. Test trials compared acoustically different target /ə/s in coarticulatorily appropriate contexts (e.g., [i$b_i$$ə_i$bi]–[a$b_a$$ə_i$bi]) with acoustically identical /ə/s, one of which was spliced into a coarticulatorily inappropriate context ([i$b_i$$ə_i$bi]–[a$b_i$$ə_i$bi]). In judging the similarity of the medial /ə/s, American English listeners often based their judgments on coarticulatory appropriateness rather than on acoustic identity, tending to hear different but coarticulatorily appropriate vowels as more similar than identical vowels. Apparently, listeners perceptually reduced acoustic differences between target vowels when these differences could be attributed to coarticulatory context (see also the results of Clark & Sharf, 1973).[1]

Pulling together these two lines of experimental inquiry—that there are systematic cross-language differences in coarticulation and that listeners use coarticulatory information in making perceptual decisions—we designed a series of experiments to investigate possible links between language-specific patterns of coarticulation and patterns of perceptual compensation. Our approach focuses on the view that compensatory perceptual responses reflect listener compensation for the acoustic consequences of coarticulation (e.g., Ohala, 1981, 1986; Fowler, 1986, 1996). However, the production–perception links in our results also address issues relevant to noncoarticulatory, auditory accounts that attribute compensatory(-like) responses to more general auditory processes (e.g., Lotto, Kluender & Holt, 1997; Lotto & Kluender, 1998).

Our working hypothesis is that language-specific coarticulatory patterns give rise to listener expectations concerning the acoustic consequences of coarticulation, which in turn lead to language-specific patterns of perceptual sensitivity to coarticulation. For example, speakers of a language with large spatiotemporal effects of V-to-V coarticulation are expected to attribute more of a vowel's acoustic structure to coarticulatory context than would speakers of a language with smaller effects; compensatory differences are also expected for speakers of a language with predominantly anticipatory V-to-V effects compared to speakers whose native language has greater carry-over effects. Importantly, this need not be the case:

[1]The most often-cited evidence of compensatory responses to coarticulatory influences comes from identification results that show a shift in listeners' category boundaries as a function of phonetic context (e.g., Mann, 1980; Mann & Repp, 1980; Whalen, 1981; Petersen, 1986; Krakow, Beddor, Goldstein & Fowler, 1988; Fowler, Best & McRoberts, 1990; Fowler, Brown & Mann, 2000). These studies are not reviewed here because they fall outside our focus on V-to-V coarticulation; for recent reviews, see Lotto & Kluender (1998) and Beddor & Krakow (1999).

compensatory context effects may be language-general, with speaker-listeners compensating for coarticulatory influences to the extent they are specified in the signal (e.g., Mann, 1986; Fowler, Best & McRoberts, 1990). Beddor & Krakow (1999) examined this question for coarticulatory vowel nasalization as perceived by native speakers of English and Thai. Because Thai has less anticipatory nasalization in /VN/ sequences than does American English, Thai listeners should, according to our approach, compensate less than English listeners for coarticulatory nasality. Listeners discriminated and rated the relative nasality of English vowels cross-spliced into oral and nasal contexts (e.g., [bõd] *vs.* [mõn]; [bod] *vs.* [mon]). Although the language-group differences were small, they were in the predicted direction of less compensation by the listeners whose native language provided less experience with nasalization.

The present study extends this general approach to V-to-V coarticulation, where the relevant articulators are the tongue, jaw and lips. It differs from our earlier work in including a detailed acoustic analysis of coarticulatory effects in the target languages, allowing comparison of specific coarticulatory patterns with specific perceptual outcomes. Additionally, in this study, we examined each language group's identification as well as discrimination of both native and non-native coarticulatory patterns, providing extensive cross-language testing of our hypothesis.

Guided by previous experimental findings, we chose Shona (a Bantu language spoken primarily in Zimbabwe) and English as languages expected to differ in the acoustic consequences of V-to-V coarticulation. Manuel's (1987) study of anticipatory V-to-V coarticulation in Shona measured the acoustic effects of context vowels /i e a o u/ on target /e/ and /a/ produced by three native Shona speakers. /pV_1ˈpV_2pV_3/ trisyllables yielded measures of the effects of stressed V_2 on unstressed target V_1, and of unstressed V_3 on stressed target V_2. $F2$ values showed systematic anticipatory effects, regardless of whether the target vowel was unstressed or stressed. As expected, $F2$ values for target /e/ and /a/ were highest in the context of /i/ and lowest in the context of /u/, with $F2$ differences between the /i/ and /u/ contexts being on the order of 200–300 Hz at target vowel offset and 40–100 Hz at target midpoint. Moreover, in a separate study of one Shona speaker's productions, anticipatory effects exceeded carry-over (Manuel & Krakow, 1984).

In contrast, most studies that compared anticipatory and carry-over V-to-V effects for English reported greater carry-over effects, particularly for unstressed vowels (Bell-Berti & Harris, 1976; Fowler, 1981; Parush, Ostry & Munhall, 1983; but compare Magen, 1997). For example, Fowler's (1981) acoustic measures of native English speakers' unstressed /ə/ in /ˈVpəˈpV/ sequences (where V = stressed /i a u/) showed that midpoint $F2$ values differed for the /i/ *vs.* /a u/ contexts by 115 Hz for carry-over effects but only 55 Hz for anticipatory effects. For stressed /ə/ (/VˈpəpV/), coarticulatory effects were smaller (on the order of 35 Hz) and nearly the same for both coarticulatory directions. In a small-scale comparison of English and Shona V-to-V coarticulation (one speaker per language), Manuel & Krakow (1984) found that Shona anticipatory effects in the $F1/F2$ space exceeded English carry-over effects.[2] Our goal in the current production study was to pursue these possible

[2]Manuel & Krakow's (1984) finding of greater Shona anticipatory than English carry-over coarticulation might appear to conflict with the Shona (Manuel, 1987) and English (Fowler, 1981) midpoint $F2$ differences we have cited in the text. However, stimulus differences between these studies (e.g., Fowler but not Manuel measured coarticulatory effects on /ə/) make direct comparison difficult.

language differences, providing a fuller characterization of V-to-V coarticulatory effects in Shona and English.

Several aspects of the phonology of Shona are potentially relevant to Shona patterns of V-to-V coarticulation. Shona has open syllable structure, and a five-vowel system /i e a o u/. As described by Beckman (1997) (based on Fortune, 1955), Shona verbs are subject to a process of vowel height harmony that restricts the distribution of nonlow vowels in verb roots: if V_1 of the root is high or low, then V_2 cannot be mid; if V_1 is mid, then V_2 cannot be high (except that /CeCu/ is allowed). Additionally, Shona is a lexical tone language with contrastive high and low tones. Stress in Shona, which falls on the penultimate syllable, is marked more by duration and amplitude than pitch (U.S. Foreign Service Institute, 1965; Myers, 1999, pers. comm.). While previous studies of V-to-V coarticulation have shown larger coarticulatory effects on unstressed than stressed vowels (e.g., Farnetani, 1990), these findings are based on nontonal languages in which pitch is an important indicator of stress.

A three-step approach tested the hypothesis that Shona and English differ in their patterns of V-to-V coarticulation, and that these coarticulatory differences parallel differences in Shona and English listeners' sensitivity to coarticulatory information. Experiment 1 was a production study to determine the detailed nature of the influences of vowels on each other in Shona and English. The coarticulatory patterns that emerged were used to formulate language-specific perceptual predictions. Experiments 2 and 3 tested the predictions by asking Shona and English listeners to discriminate naturally produced vowels spliced into different coarticulatory contexts (Experiment 2) and to identify synthetic vowels embedded in different vowel contexts (Experiment 3).

2. Experiment 1: acoustic analysis

Previous acoustic findings summarized in the Introduction lead us to expect the largest V-to-V coarticulatory effects in Shona to be anticipatory, with similar effects on stressed and unstressed vowels. English should show especially large carry-over influences on unstressed vowels. To explore these and other coarticulatory influences, Experiment 1 studied anticipatory and carry-over coarticulation in Shona and English stressed and unstressed vowels.

2.1. *Methods*

2.1.1. *Stimulus materials*

The materials for the acoustic analysis were trisyllables that constitute phonetically appropriate sequences in Shona and English, but are in most cases nonsense words in both languages. The trisyllables were $CV_1^{'}CV_2CV_3$ sequences, where V = /i e a o u/, and $V_1 = V_2$ and/or $V_2 = V_3$. Within these constraints, all possible vowel combinations were included, for a total of 45 different trisyllables: five in which $V_1 = V_2 = V_3$ (e.g., Ci'CiCi), 20 $V_1 = V_2 \neq V_3$ (e.g., Ci'CiCa), and 20 $V_1 \neq V_2 = V_3$ (e.g., Ca'CiCi). Thus, each of /i e a o u/ occurred in stressed and unstressed positions in each of five vowel contexts, under possible anticipatory or carry-over

TABLE I. Experiment 1: Stimulus materials for stressed and unstressed target vowel /i/ (in bold) grouped according to context vowel and direction of potential influences of the context on the target. (Consonants are omitted; e.g., i ˈi e = CiˈCiCe.)

		Context vowel				
		i	e	a	o	u
Stressed target	Anticipatory	i **i** i	i ˈ**i** e	i ˈ**i** a	i ˈ**i** o	i ˈ**i** u
	Carryover	i ˈ**i** i	e ˈ**i** i	a ˈ**i** i	o ˈ**i** i	u ˈ**i** i
Unstressed target	Anticipatory	**i** ˈi i	**i** ˈe e	**i** ˈa a	**i** ˈo o	**i** ˈu u
	Carryover	i ˈi **i**	e ˈe **i**	a ˈa **i**	o ˈo **i**	u ˈu **i**

Note: All examples in the left-most column (context vowel /i/) are the same stimulus.

influences. Table I illustrates the design for target vowel /i/. All Shona trisyllables were produced with a low tone.

The five vowels in the stimulus materials were selected as the closest counterparts in the two languages: phonetically, [i e̞ ɑ̟ o̞ u] in Shona and [i e ɑ o̞ u̟] in English. Anticipating our acoustic results, Fig. 1 gives the $F1/F2$ space for these Shona and English vowels. (Formant measures were taken at vowel midpoint, and are averaged across speakers and all vowel contexts.) Shona mid vowels are slightly higher, and Shona rounded vowels are further back, than their English counterparts. When stressed, the English mid vowels were sometimes diphthongal ([eɪ], [oʊ]).

The C in the CV_1ˈCV_2CV_3 stimuli was /p/ in Shona and /b/ in English, chosen as the phonetically most similar bilabial stops in these languages. More nearly identical consonants were not possible because the voiced bilabial stop in Shona is an implosive ([ɓ]) and the voiceless bilabial stop in English is typically heavily aspirated in stressed position. Shona /p/, as produced by the speakers we recorded, was more moderately aspirated (see Section 2.1.3).

Speakers were recorded reading multiple randomized lists of the 45 trisyllables embedded in a sentence context, Shona [ˈtaura] __ [paˈtʃena] ("Say __ clearly") and English "Give me a __ tomorrow".[3] (Nearly all speakers read nine randomizations; due to time constraints, two Shona speakers read only six.) The first randomization was treated as practice and was not analyzed. The remaining 5–8 randomizations per speaker were digitized at 11.025 kHz.

2.1.2. *Speakers*

Seven native speakers of Shona (three female and four male) and five native speakers of English (two female and three male) were recorded. The Shona speakers were born in Zimbabwe and raised in Shona-speaking homes, and were University of Michigan or Michigan State University students at the time of recording. Their length of stay in the U.S. ranged from 4 months to 4 years. Although fluent in

[3]The /p/ of Shona "pachena" *vs.* /t/ of English "tomorrow" may have imposed different vocal tract constraints on V_3, but we expect that these constraints would have primarily influenced V_3 offset, which was not analyzed in this study.

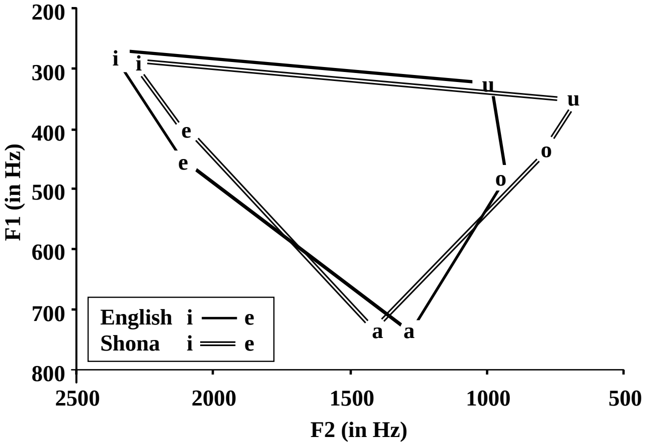

Figure 1. $F1/F2$ space for English (single line) and Shona (double line) /i e a o u/, summing across vowel contexts and speakers. Formant measures were taken at vowel midpoint.

English (as is the case for all formally educated Shona speakers), all but one continued to speak Shona in the home while in the U.S. The native English speakers were students, plus one faculty member, in linguistics at the University of Michigan. Phonetically trained native English speakers were chosen to avoid both difficulties with English orthography and excessive vowel reduction.[4] Speakers were recorded in a sound-attenuated room either in the Phonetics Laboratory at the University of Michigan or in the Audiology Laboratory at Michigan State University.

2.1.3. *Acoustic measures*

Using SoundScope/16 (GW Instruments) speech analysis software, measures of the first three formants, $F1$–$F3$, were taken at three temporal locations in each vowel in a trisyllable: vowel onset, midpoint, and offset. Temporal locations were identified from the waveform display. Formant measures were taken from superimposed FFT and LPC (14-coefficient autoregressive analysis) displays, with a 25 ms analysis window. The window was centered over the initial and final 25 ms of vocalic voicing for onset and offset measures, respectively; the midpoint window was centered around the point halfway between vowel onset and offset. For English, where C = /b/, vocalic voicing was identified as the point of substantial change—increase for onset and decrease for offset—in periodicity. For Shona, where C = /p/, periodic pulsing was taken to be vocalic voicing. Formant frequency decisions were based primarily on LPC peaks.

[4]Shona unstressed vowels exhibit little vowel reduction and, coarticulatory influences aside, we wanted the qualities of the Shona and English vowels to be as similar as possible. Because our sequences did not violate English phonotactics, phonetically naive English speakers could have been used if provided with appropriate real English words as models. Phonetically trained speakers provided a simpler alternative, as their training should not influence V-to-V coarticulatory patterns.

Although our focus was spectral patterns, we also verified that the temporal locations of the spectral measures were comparable in the two languages. Vowel duration varied across vowel height and speakers, but did not systematically vary across languages. The duration of most unstressed vowels fell in the 85–135 ms range; stressed vowels averaged about 10% longer than unstressed in both languages. The duration of the interval between stop release and onset of periodic voicing for the following vowel was also comparable (10–30 ms) for the English speakers' and four of the Shona speakers' productions. However, three Shona speakers showed longer voice onset times (30–50 ms for most tokens). Regarding stop closure duration, both languages had longer medial closures at the onset of stressed syllables (C_2 in $C_1V'C_2VC_3V$) than at the onset of unstressed syllables (C_3). However, closure duration was longer overall for Shona /p/ than for English /b/, and stress-onset lengthening was more pronounced in Shona (85 ms for C_2 *vs.* 68 ms for C_3) than in English (65 ms for C_2 *vs.* 57 ms for C_3).

2.1.4. *Statistical design*

Statistical analyses were restricted to *F*1 and *F*2 measures. We conducted separate repeated measures ANOVAs for *F*1 and *F*2, plus separate tests on midpoint *vs.* onset/offset values. Tests on onset/offset values (or vowel "edge") compared target vowel measures that fell temporally closest to the context vowel. For example, tests of the anticipatory *vs.* carry-over effects of context /a/ on stressed target /i/ compared the offset measure of V_2 in Ci'<u>Ci</u>Ca to the onset measure of V_2 in Ca'<u>Ci</u>Ci; study of the same effects on unstressed target /i/ compared the offset measure of V_1 in <u>Ci</u>'CaCa to the onset measure of V_3 in Ca'Ca<u>Ci</u>. (Thus, onset V_1 and offset V_3 measures were not included in the analyses.) Midpoint values were always compared to other midpoint values.

Of central concern is whether Shona and English differ in the acoustic effects of V-to-V coarticulation, with Shona showing predominantly anticipatory effects, and English showing predominantly carry-over effects especially on unstressed vowels. As a measure of the extent of coarticulation, we calculated "coarticulation scores" for each speaker representing that speaker's context-induced deviations in *F*1 or *F*2. The calculations involved four steps. Step (1) $\text{mean}_{\text{Target V}}$: across context vowels, calculate each speaker's mean *F*1 (or *F*2) value for each stressed and unstressed target vowel. Step (2) context-specific deviations from $\text{mean}_{\text{Target V}}$: subtract each speaker's *F*1 (or *F*2) of a target vowel *in a given context* from that speaker's $\text{mean}_{\text{Target V}}$. [For example, for English speaker D, $\text{mean}_{\text{Target V}}$ *F*2 of stressed /e/ was 2229 Hz, and *F*2 of stressed /e/ preceded by /o/ was 1994 Hz at vowel onset (averaging across the eight tokens of [bo'bebe]), yielding an *F*2 deviation of 235 Hz for the carry-over effects of context /o/ on stressed target /e/.] Step (3) signed (+/−) deviations: to each deviation, assign + or − based on well-attested findings about the acoustic effects of coarticulation. (Signs differentiate deviations which conform to expected coarticulatory effects, i.e., + values, from small, sporadically occurring deviations in the opposite direction. Negative values were infrequent, and were largely restricted to vowel midpoint, by which time the V-to-V coarticulatory influences were diminishing.) Relative to $\text{mean}_{\text{Target V}}$ (which fell near the center of that vowel's acoustic space), coarticulation should lower *F*2 in back (/ɑ o u/) vowel contexts, raise *F*2 in front (/i e/) contexts, lower *F*1 in high (/i u/) contexts, and raise

$F1$ in nonhigh (/e ɑ o/) contexts.[5] (For example, the 235 Hz decrease in $F2$ in the example above, which conformed to coarticulation expectations, was coded as +235 Hz.) $F1/F2$ deviations were not calculated for vowels in the context of themselves ([ba'baba], [bi'bibi], etc.), because prior work on V-to-V coarticulation does not lead to clear predictions about such cases. Moreover, such vowels are generally regarded as being relatively canonical, that is, close to the configuration of noncoarticulated vowels. Step (4) coarticulation scores: calculate average +/− deviations across target vowels and contexts, yielding 4 "coarticulation scores" per speaker (per formant and temporal location): anticipatory stressed and unstressed, and carry-over stressed and unstressed. Four repeated measures ANOVAs (2 formants × 2 temporal locations) each tested for the effects of Language (Shona, English), Direction (anticipatory, carry-over), and target vowel Stress (stressed, unstressed) on speakers' coarticulation scores.

2.2. *Results*

The mean anticipatory and carry-over effects of the context vowels, averaging across target vowels and speakers, are shown in Fig. 2 for English and Fig. 3 for Shona. Underived Hz values were used in these plots for easy interpretability; plots of the deviation Hz values used in calculating the coarticulation scores result in very comparable vowel spaces. The vowel spaces formed by these coarticulatory effects show the usual variation in formant frequencies: overall, $F2$ is lowered in back and raised in front vowel contexts, while $F1$ is lowered in high and raised in low vowel contexts. As expected, the acoustic effects diminished as the formant measures moved further from the coarticulatory source, as seen by substantially greater context effects at vowel onset/offset—henceforth, vowel edge (top panels)—than at midpoint (bottom panels).

Important aspects of the expected language-specific patterns are borne out. For English, context effects are greater for unstressed than for stressed vowels (Fig. 2 left *vs*. right), and carry-over effects are at least as large as anticipatory (Fig. 2 dashed *vs*. solid lines). For Shona, context effects are more nearly comparable for unstressed and stressed vowels (Fig. 3 left *vs*. right). Also, Shona anticipatory effects are greater than carry-over, at least at vowel edge (Fig. 3(a) and (b)). However, unexpectedly, V-to-V coarticulatory effects were not greater in Shona than in English; indeed, the opposite is the case at vowel edge, especially in $F1$ (Fig. 3(a), (b) *vs*. Fig. 2(a), (b)). The statistical analyses, conducted on $F1$ and $F2$ measures at vowel edge and midpoint, directly compared the coarticulatory effects in the two languages and supported these observations.

2.2.1. *F2 dimension*

For $F2$ at vowel edge, a three-way repeated measures ANOVA with factors Language (Shona, English), Direction (anticipatory, carry-over), and Stress (stressed,

[5] The assumption that mid vowels should raise $F1$ relative to $\text{mean}_{\text{Target V}}$ was based on the fact that $\text{mean}_{\text{Target V}}$ fell slightly higher than the center of $F1$ space (i.e., closer to the high vowels). This outcome was due to our inclusion of two high vowels, as opposed to only one low vowel, in the stimulus set. Not surprisingly, however, the $F1$ deviations in mid vowel contexts relative to $\text{mean}_{\text{Target V}}$ were generally quite small.

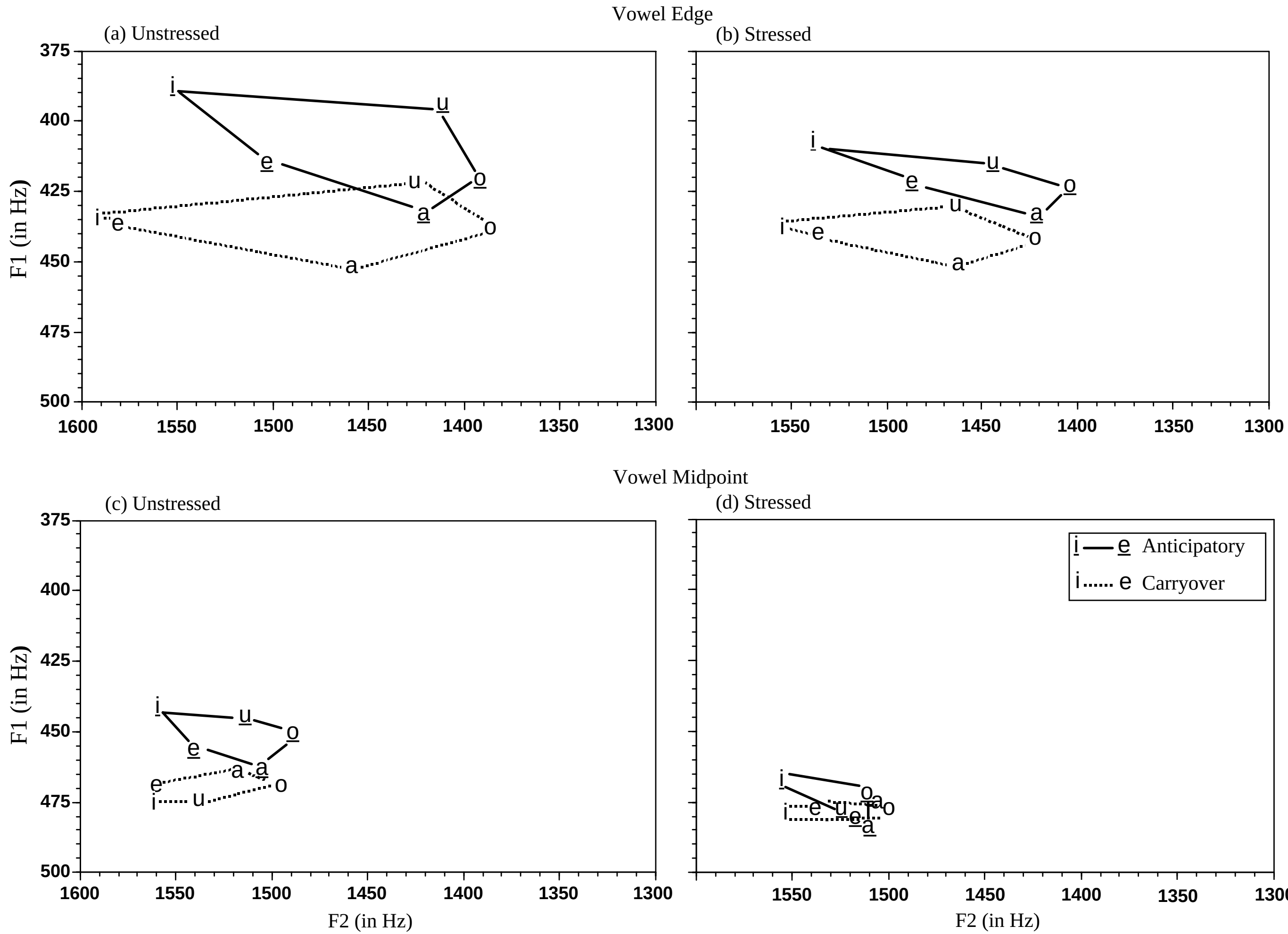

Figure 2. English V-to-V coarticulation: Anticipatory (solid line) and carry-over (dashed line) effects of the context vowels /i e a o u/ on $F1$ and $F2$ of unstressed (left panels) and stressed (right) target vowels, averaging across five target vowels and five speakers. Upper panel measures were taken at target vowel edge (offset for anticipatory and onset for carry-over); lower panel measures were taken at target vowel midpoint.

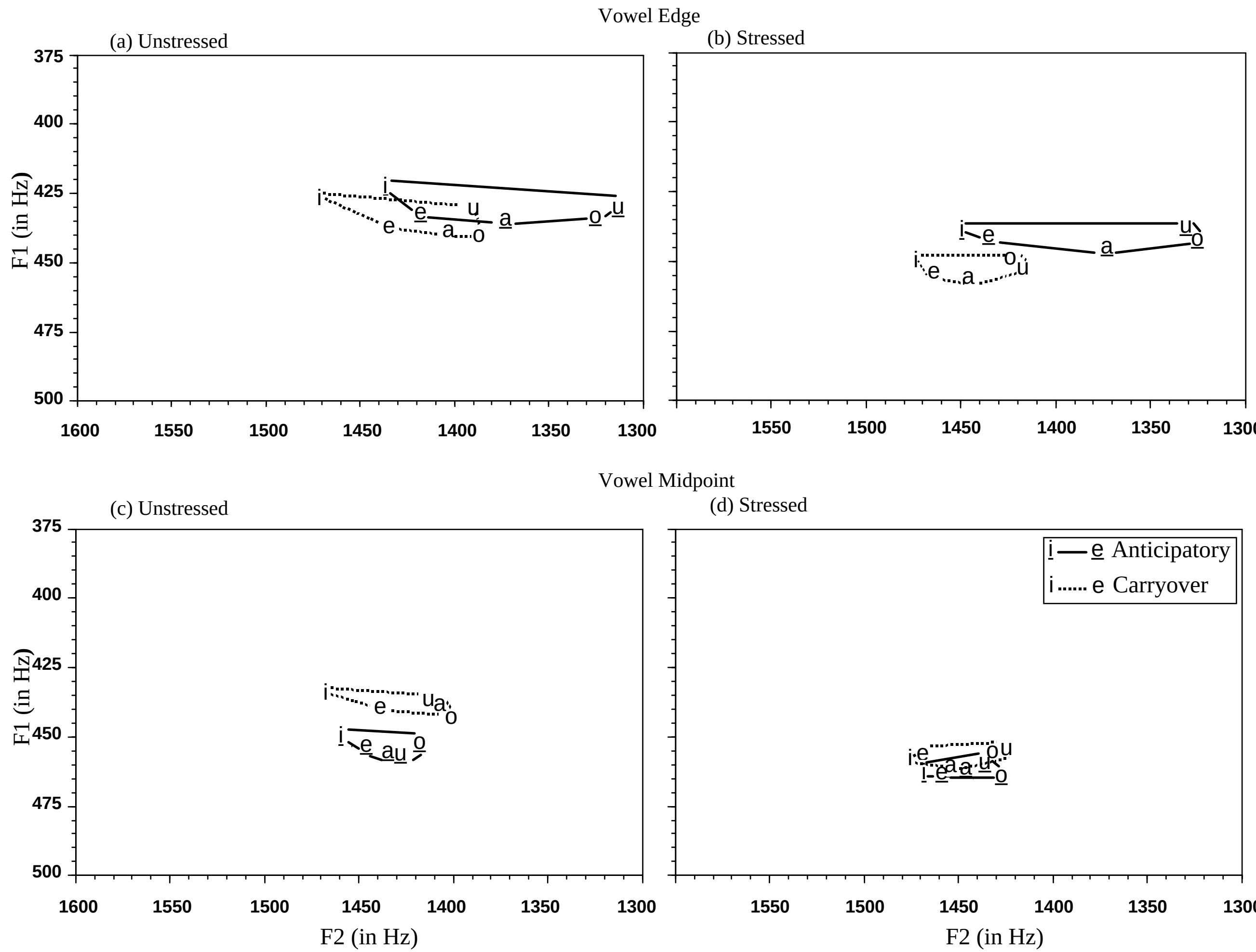

Figure 3. Shona V-to-V coarticulation: Anticipatory and carry-over effects of the context vowels /i e a o u/ on $F1$ and $F2$ of unstressed (left panels) and stressed (right) target vowels, averaging across five target vowels and seven speakers. Upper panel measures were taken at vowel edge; lower panel measures were taken at vowel midpoint.

unstressed) was conducted on the coarticulation scores for the five English and seven Shona speakers' productions. All main effects and two-way interactions were significant. The main effects of Language, Direction, and Stress [$F(1,10) = 7.44$, $p < 0.05$; 5.45, $p > 0.05$; 13.25, $p < 0.005$; respectively] are best understood in terms of the two-way interactions involving Language. *Post hoc* Tukey pairwise comparisons investigating the Language by Direction interaction [$F(1,10) = 12.08$, $p < 0.01$] showed $F2$ carry-over effects to be significantly larger for English than for Shona [$t(10) = 4.12$, $p < 0.005$], with mean English carry-over effects being 2.4 times the size of those of Shona (compare the dashed-line $F2$ dimensions in Figs. 2(a) and 3(a), and 2(b) and 3(b)). However, anticipatory effects did not differ for the two languages (the mean difference being less than 5 Hz; compare the solid-line $F2$ dimensions in Figs. 2(a) and 3(a), and 2(b) and 3(b)). The Language by Stress interaction [$F(1,10) = 12.03$, $p < 0.01$] was due to significantly larger coarticulatory effects on unstressed vowels in English than in Shona [1.5 times larger; $t(10) = 3.84$, $p < 0.005$]; in contrast, no language difference was found for stressed vowels. Thus, at vowel edge, $F2$ carry-over effects are greater in English than in Shona, especially for unstressed vowels. While Shona anticipatory effects are larger than Shona carry-over effects [on average, 2.0 times larger; $t(10) = 4.50$, $p < 0.005$], they are not larger (or smaller) than English anticipatory effects ($p > 0.5$).

By vowel midpoint, the cross-language $F2$ differences have statistically disappeared. A three-way ANOVA (Language, Direction, Stress) conducted on coarticulation scores showed a marginally significant effect of Stress [$F(1,10) = 4.86$, $p < 0.052$], with coarticulatory effects on unstressed vowels being 1.6 times the size of the effects on stressed vowels. The unstressed–stressed difference was larger for English than for Shona, but this difference was not significant, nor was any other language or directional effect significant. (This means, for example, that the slightly greater carry-over than anticipatory $F2$ effects seen for Shona unstressed vowels in Fig. 3(c) were not significant.)

2.2.2. $F1$ *dimension*

Figs. 2 and 3 exhibit greater coarticulatory effects on $F1$ for English than for Shona speakers' productions. This language difference is especially robust at vowel edge. A three-way ANOVA run on $F1$ coarticulation scores at vowel edge showed a significant main effect of Language [$F(1,10) = 33.18$, $p < 0.001$], with English scores being four times larger than Shona scores. Language interacted with Direction [$F(1,10) = 7.20$, $p < 0.05$], and *post hoc* Tukey comparisons showed that the interaction was due to larger anticipatory than carry-over effects in English [$t(10) = 3.04$, $p < 0.05$] but comparable anticipatory and carry-over effects in Shona. For both directions, however, $F1$ coarticulatory effects at vowel edge were greater for English than for Shona. The ANOVA also showed a significant main effect of Stress [$F(1,10) = 13.06$, $p < 0.005$], due to larger coarticulatory effects on unstressed vowels. The unstressed–stressed difference was again greater for English than for Shona, with only the English difference reaching significance [$t(10) = 3.04$, $p < 0.05$]. No other main effect or interaction was significant.

By vowel midpoint, the statistically larger coarticulatory effects for English are restricted to unstressed vowels in the anticipatory direction. The three-way ANOVA on midpoint $F1$ coarticulation scores showed a significant main effect only of

Direction [$F(1,10) = 12.24$, $p<0.01$], which significantly interacted with Language [$F(1,10) = 24.41$, $p<0.001$]. These two factors also participated in a marginally significant three-way interaction with Stress [$F(1,10) = 4.10$, $p<0.071$]. Tukey comparisons revealed that while anticipatory effects on unstressed vowels were significantly larger for English than for Shona [$t(10) = 4.02$, $p<0.005$; compare the solid-line $F1$ dimensions in Figs. 2(c) and 3(c)], the remaining cross-language comparisons were not significant.

2.3. *Discussion*

Shona and English differ in the acoustic effects of V-to-V coarticulation in two primary respects. First, both languages have considerable anticipatory coarticulation, but English exhibits more extensive carry-over V-to-V coarticulation than does Shona. This difference diminishes by vowel midpoint, but is present in both the $F1$ and $F2$ dimensions near vowel onset. Second, unstressed vowels in English, but not Shona, coarticulate more than do stressed vowels. These language patterns hold for individual speakers' productions as well. For example, other than a single exception, there was no overlap in the carry-over coarticulation scores for $F1$ or $F2$ at vowel edge for the five English *vs.* seven Shona speakers. In contrast, the anticipatory $F2$ coarticulation scores at vowel edge overlapped completely for the language groups, with English speakers' scores falling within the Shona range. Regarding stress patterns, all English speakers' productions showed robust unstressed–stressed differences in coarticulation, while only one Shona speaker exhibited a similar tendency (restricted, in her case, to $F2$). From a perceptual standpoint, these findings mean that English listeners should have considerably more experience than Shona listeners with large carry-over coarticulatory effects on unstressed vowels.

These overall coarticulatory patterns also hold for most individual target vowels. For example, the majority of English target vowels showed the unstressed–stressed asymmetry in $F1$ and $F2$ at midpoint and edge, while no Shona target vowel showed a consistent stress pattern. In addition, each of the five Shona target vowels showed larger anticipatory than carry-over effects in $F2$ at vowel edge, while no English target vowel showed this pattern.[6] In both languages, target /i/ showed the weakest coarticulatory effects, a finding consistent with previous investigations of these and other languages (Farnetani, Vagges & Magno-Caldognetto, 1985; Recasens, 1987, 1999; Beddor & Yavuz, 1995; Kondo & Arai, 1998).

What factors might be responsible for these cross-language coarticulatory differences? The anticipatory effects on $F2$ are comparable in the two languages, but larger carry-over effects in English than in Shona hold for both $F1$ and $F2$. The English pattern of strong carry-over effects especially on unstressed vowels—i.e., strong effects of V_2 on V_3—is consistent with left-headed prosodic feet in English

[6]These summary statements about the behavior of individual target vowels are based on the results of a series of four-way Repeated Measures ANOVAs that tested for the effects of Target Vowel (/i e a o u/), Context Vowel (/i e a o u/), Direction (anticipatory, carry-over), and target vowel Stress (stressed, unstressed). Due to the large number of factors, these analyses were conducted separately for the two languages (Shona, English), as well as formants ($F1$, $F2$), and temporal locations (midpoint, edge). In the interest of space, and in view of the only indirect comparison of the language groups, the detailed results are not reported here. However, as noted in the text, the outcomes were broadly consistent with the cross-language statistical comparisons reported in 2.2, which averaged across target vowels.

(Vayra, Fowler & Avesani, 1987). (However, robust anticipatory effects of stressed V_2 on preceding unstressed V_1 indicate that the foot is not the exclusive domain of V-to-V coarticulation; see also Magen, 1997.) Whether Shona prosodic structure differs from that of English in ways that might explain Shona's more limited carry-over patterns is not clear. Systematic penultimate stress in Shona is in keeping with left-headed prosodic feet, as in English. At the same time, the foot head may be less prominent in Shona, which has lexical tone and does not use pitch (tone) to mark the head. This possibility meshes with Shona's weaker carry-over effects of stressed V_2 on V_3, and the overall similar coarticulatory behavior of stressed and unstressed Shona vowels. Weak carry-over effects of Shona V_1 on V_2 might also stem from this possibility: Keating, Cho, Fougeron & Hsu (in press) have speculated that languages that do not use pitch to mark domain heads might have especially strong domain edges—here, C_2 (in $C_1V_1{}^{\prime}C_2V_2C_3V_3$), whose closures were particularly long in Shona (see Section 2.1.3)—thereby limiting coarticulation across such edges. At this point, although aspects of coarticulation are clearly governed by prosodic structure (e.g., de Jong, Beckman & Edwards, 1993; Smith, 1995; Byrd, Kaun, Narayanan & Saltzman, 2000), not enough is known about the detailed effects of prosody on coarticulation to offer more than speculative considerations about this possible source of cross-language V-to-V differences.

An unexpected result was that not only the magnitude of carry-over effects, but also the magnitude of anticipatory effects, in Shona did not exceed those in English. This outcome conflicts with the view that vowels in small vowel systems are more free to vary than those in large systems (see also Bradlow, 1995, for findings that conflict with inventory-size predictions). Despite the conflict, our measures for Shona and English largely conform to previously reported findings for these languages. Our absolute formant values for Shona coarticulation compare favorably with Manuel's (1987) values at both vowel edge and midpoint, from similar stimulus materials. Our values for English are compatible with those reported by Magen (1997), who found coarticulatory influences of English /i/ and /a/ on each other as large as 50 Hz at *F*2 midpoint—i.e., about the same size as our effects—but in her case the effects were across an intervening /ə/. The failure of our data to support inventory-size predictions could be due to a number of factors. As Manuel (1999) has pointed out, freedom to vary (in terms of phonological contrasts) does not necessarily entail extensive coarticulation in a language; languages may exhibit more or less variation for other reasons, including sociolinguistic ones (e.g., affiliation with a linguistic community through use of certain perceptible but noncontrastive properties; Ladefoged, 1984, 1989). In this experiment, it may be that coarticulatory overlap in English was especially apparent due to shorter intervocalic stop closure durations in English than Shona. Consistent with this possibility, the English speaker whose productions had the longest stop closures also had the smallest V-to-V coarticulatory effects. Perceptually, of importance here is that English V-to-V coarticulation provides English listeners with substantial experience with anticipatory and carry-over patterns, and that Shona V-to-V coarticulation provides Shona listeners with substantial experience with anticipatory, but not carry-over, patterns.

A factor that did not play any obvious role in Shona V-to-V coarticulation was phonological vowel height harmony in Shona. Although only certain sequences of nonlow vowels may occur in Shona verb roots (see the Introduction), the trisyllables in our data that formed permissible verb sequences did not behave differently from those that formed impossible verb sequences, presumably because all sequences conformed to possible non-verbs in Shona. Moreover, despite height harmony, V-to-V coarticulatory effects in $F1$ were not particularly robust in the Shona data.

3. Experiment 2: discrimination of V-to-V coarticulatory effects

The perceptual experiments tested the hypothesis that the language-specific patterns of coarticulation that emerged in the production study are paralleled by language-specific perceptual responses to coarticulatory information. Previous research has shown that Shona (Manuel, 1987) and English (e.g., Alfonso & Baer, 1982; Martin & Bunnell, 1982; Fowler & Smith, 1986) listeners are sensitive to V-to-V coarticulatory information; we explored the possibility that listeners' use of this information is determined in part by native-language coarticulatory patterns.

Experiment 2 investigated listeners' perception of V-to-V coarticulatory effects using a 4IAX discrimination paradigm similar to that used by Fowler (Fowler, 1981; Fowler & Smith, 1986; see also Beddor & Krakow, 1999). In this approach, each trial consisted of two pair types, one pair with acoustically distinct target vowels both embedded in coarticulatorily appropriate contexts, and the other pair with acoustically identical target vowels, one of these being embedded in an inappropriate coarticulatory context. Perceptual compensation for coarticulation was measured by whether listeners judged vowel similarity based on the vowels' acoustic identity (a noncompensatory response) or on the appropriateness of the vowels' coarticulatory context (a compensatory response). We selected for investigation the V-to-V coarticulatory patterns in each language that showed large—so presumably perceptible—effects. For Shona, anticipatory effects were stronger than carry-over, and we chose to investigate anticipatory coarticulation on stressed vowels. For English, we chose the coarticulatory pattern that was maximally distinct from Shona: carry-over coarticulation on unstressed vowels.

3.1. *Methods*

3.1.1. *Stimulus materials*

Two discrimination tests were created, a "Shona" test of anticipatory effects created from stimuli produced by a native Shona speaker, and an "English" test of carry-over effects created from stimuli produced by a native American English speaker. Both were female speakers whose coarticulatory patterns were analyzed in the acoustic portion of the study and were representative of the overall patterns for each language.

The original stimuli were ['CV_1CV_2] disyllables, where the vowels involved all possible combinations of /i e a o u/ (e.g., ['CiCa], ['CuCe], ['CaCa]) and consonants were English /b/ or Shona /p/. Multiple randomizations of the 25 disyllables, each embedded in the sentence context used for the acoustic study, were recorded. Disyllables, rather than the trisyllables from the acoustic study, were used for

TABLE II. Experiment 2: Average $F1$ and $F2$ differences (in Hz) at vowel onset and midpoint for English vowel pairings

Context vowel	Target vowel	Vowel pairing	Onset		Midpoint	
			$F1$	$F2$	$F1$	$F2$
/i/	/a/	$_{i}a–_{a}a$	−114	302	−186	136
	/o/	$_{i}o–_{o}o$	−83	361	36	172
/e/	/a/	$_{e}a–_{a}a$	−57	281	−107	101
	/o/	$_{e}o–_{o}o$	−69	270	22	126
/a/	/o/	$_{a}o–_{o}o$	32	232	0	28
/o/	/a/	$_{o}a–_{a}a$	−50	−107	−56	65
	/e/	$_{o}e–_{e}e$	−7	−395	43	−227
/u/	/a/	$_{u}a–_{a}a$	−148	29	−179	79
	/e/	$_{u}e–_{e}e$	14	−342	−14	−172

TABLE III. Experiment 2: Average $F1$ and $F2$ differences (in Hz) at vowel midpoint and offset for Shona vowel pairings

Context vowel	Target vowel	Vowel pairing	Midpoint		Offset	
			$F1$	$F2$	$F1$	$F2$
/i/	/a/	$a_{i}–a_{a}$	−57	79	−79	381
	/o/	$o_{i}–o_{o}$	−7	57	14	270
/e/	/a/	$a_{e}–a_{a}$	8	22	7	432
	/o/	$o_{e}–o_{o}$	43	114	−7	256
/a/	/e/	$e_{a}–e_{e}$	43	−86	15	−201
/u/	/a/	$a_{u}–a_{a}$	43	0	−54	−430
	/e/	$e_{u}–e_{e}$	−11	−43	21	−366

perceptual testing to simplify the listeners' task, but the same acoustic patterns were found for the disyllabic tokens.

In creating the experimental stimuli, the ˈCV_1CV_2 disyllables were excised from the carrier sentence, and target ˈCV_1s (Shona) or CV_2s (English) were in turn excised from the disyllables. Excised single syllables included the entire vocalic portion and the preceding stop (stop closure and burst plus any aspiration). These excised syllables were spliced into new contexts ([ˈ_CV_2] for Shona and [ˈCV_1_] for English), and the resulting cross-spliced disyllables were paired in a 4IAX design for presentation to listeners. Target vowels in paired disyllables were phonologically identical, with target vowel differences being V-to-V coarticulatory effects: paired disyllables always had one target vowel originally produced in the context of itself (e.g., /a/ in the context of /a/) and one target originally produced in the context of another vowel (e.g., /a/ in the context of /i/). We selectively paired target vowels

TABLE IV. Experiment 2: Trial types for the English 4IAX discrimination task, where target vowels are V_2

V_2s	Control	Test: Canonical	Test: Noncanonical
$_i$a-$_a$a	bib$_i$a-bib$_i$a / bib$_i$a-bib$_a$a	bab$_a$a-bib$_a$a / bab$_a$a-bib$_i$a	bab$_i$a-bib$_i$a / bab$_a$a-bib$_i$a
$_i$o-$_o$o	bib$_i$o-bib$_i$o / bib$_i$o-bib$_o$o	bob$_o$o-bib$_o$o / bob$_o$o-bib$_i$o	bob$_i$o-bib$_i$o / bob$_o$o-bib$_i$o
$_e$a-$_a$a	beb$_e$a-beb$_e$a / beb$_e$a-beb$_a$a	bab$_a$a-beb$_a$a / bab$_a$a-beb$_e$a	bab$_e$a-beb$_e$a / bab$_a$a-beb$_e$a
$_e$o-$_o$o	beb$_e$o-beb$_e$o / beb$_e$o-beb$_o$o	bob$_o$o-beb$_o$o / bob$_o$o-beb$_e$o	bob$_e$o-beb$_e$o / bob$_o$o-beb$_e$o
$_a$o-$_o$o	bab$_a$o-bab$_a$o / bab$_a$o-bab$_o$o	bob$_o$o-bab$_o$o / bob$_o$o-bab$_a$o	bob$_a$o-bab$_a$o / bob$_o$o-bab$_a$o
$_o$a-$_a$a	bob$_o$a-bob$_o$a / bob$_o$a-bob$_a$a	bab$_a$a-bob$_a$a / bab$_a$a-bob$_o$a	bab$_o$a-bob$_o$a / bab$_a$a-bob$_o$a
$_o$e-$_e$e	bob$_o$e-bob$_o$e / bob$_o$e-bob$_e$e	beb$_e$e-bob$_e$e / beb$_e$e-bob$_o$e	beb$_o$e-bob$_o$e / beb$_e$e-bob$_o$e
$_u$a-$_a$a	bub$_u$a-bub$_u$a / bub$_u$a-bub$_a$a	bab$_a$a-bub$_a$a / bab$_a$a-bub$_u$a	bab$_u$a-bub$_u$a / bab$_a$a-bub$_u$a
$_u$e-$_e$e	bub$_u$e-bub$_u$e / bub$_u$e-bub$_e$e	beb$_e$e-bub$_e$e / beb$_e$e-bub$_u$e	beb$_u$e-bub$_u$e / beb$_e$e-bub$_u$e

Note: Trials are grouped according to context vowel (V_1). Control trials hold context constant. For each test trial, the left pair member has acoustically identical V_2s, one being in an inappropriate coarticulatory context; the right pair member has acoustically distinct V_2s, both in appropriate coarticulatory contexts. Canonical and noncanonical trials differ only in the coarticulatory effect in the identical V_2s (see text for explanation).

TABLE V. Experiment 2: Trial types for Shona 4IAX discrimination task, where target vowels are V_1

V_1s	Control	Test: Canonical	Test: Noncanonical
a$_i$-a$_a$	pa$_i$pi-pa$_i$pi / pa$_i$pi-pa$_a$pi	pa$_a$pa-pa$_a$pi / pa$_a$pa-pa$_i$pi	pa$_i$pa-pa$_i$pi / pa$_a$pa-pa$_i$pi
o$_i$-o$_o$	po$_i$pi-po$_i$pi / po$_i$pi- po$_o$pi	po$_o$po-po$_o$pi / po$_o$po-po$_i$pi	po$_i$po-po$_i$pi / po$_o$po-po$_i$pi
a$_e$-a$_a$	pa$_e$pe-pa$_e$pe / pa$_e$pe-pa$_a$pe	pa$_a$pa-pa$_a$pe / pa$_a$pa-pa$_e$pe	pa$_e$pa-pa$_e$pe / pa$_a$pa-pa$_e$pe
o$_e$-o$_o$	po$_e$pe-po$_e$pe / po$_e$pe-po$_o$pe	po$_o$po-po$_o$pe / po$_o$po-po$_e$pe	po$_e$po-po$_e$pe / po$_o$po-po$_e$pe
e$_a$-e$_e$	pe$_a$pa-pe$_a$pa / pe$_a$pa-pe$_e$pa	pe$_e$pe-pe$_e$pa / pe$_e$pe-pe$_a$pa	pe$_a$pe-pe$_a$pa / pe$_e$pe-pe$_a$pa
a$_u$-a$_a$	pa$_u$pu-pa$_u$pu / pa$_u$pu-pa$_a$pu	pa$_a$pa-pa$_a$pu / pa$_a$pa-pa$_u$pu	pa$_u$pa-pa$_u$pu / pa$_a$pa-pa$_u$pu
e$_u$-e$_e$	pe$_u$pu-pe$_u$pu / pe$_u$pu-pe$_e$pu	pe$_e$pe-pe$_e$pu / pe$_e$pe-pe$_u$pu	pe$_u$pe-pe$_u$pu / pe$_e$pe-pe$_u$pu

Note: Trials are grouped according to context vowel (V_2). (See Table IV for explanation of trial types, substituting V_1 for V_2.)

that had perceptibly distinct coarticulatory effects as determined by pilot testing with phonetically naive native Shona and English speakers. Pairings were excluded if native listener accuracy fell below 67% (2/3) correct on control trials (see below). Pilot results yielded nine English pairings of coarticulatory effects and seven Shona pairings, listed in Tables II and III, that could be reliably differentiated. Subscripts in the tables denote the original vowel context of an excised syllable (e.g., [a$_i$] has /a/ excised from original Shona [papi]; [$_i$a] has /a/ excised from original English [biba]). Pilot testing was also used to select three different tokens of each type of pairing, except that for two Shona pairings (a_a–a_u, u_e–u_u) only two reliably discriminable pair tokens were identified. Tables II and III give the average $F1$ and $F2$ differences for the pairings.

The nine English and seven Shona types of coarticulatory pairings were arranged into three different types of 4IAX trials, given in Tables IV and V. We describe here the pairings for the English tests, but the same manipulations—except for V_1 rather than V_2—hold for Shona. In all English trials, the V_2s of one pair were acoustically identical (first pair in each trial in Table IV) and the V_2s of the other pair were acoustically different; all V_2s in a trial belonged to the same phonological vowel category. The control trials held context constant and varied only V_2 coarticulatory differences.[7] Control trials were included to verify that the coarticulatory effects were discriminable using this paradigm, and listeners were expected to perform well on these trials. The test trials investigated whether discrimination of the V_2 differences depended on vowel context. For each test trial, the V_2s of the pair with acoustically different vowels were both in coarticulatorily appropriate contexts (e.g., [bab$_a$a-bib$_i$a]), while one of the V_2s of the pair with acoustically identical vowels was in a coarticulatorily inappropriate context. Each vowel pairing had two types of test trials for each pairing because, for every type of pairing, there are two possible inappropriate contexts (e.g., for pair [$_i$a]–[$_a$a], [$_i$a] in the context [bab_] or [$_a$a] in the context [bib_]). In the "canonical" trials, the contextually inappropriate V_2 was originally produced in the context of a phonologically identical V_1 (that is, in the context of itself; e.g., [$_a$a]). In the "noncanonical" trials, the inappropriate V_2 originally followed a phonologically distinct V_1 (e.g., [$_i$a]). This terminology reflects our acoustic finding that vowels produced in the context of themselves tend to be more prototypical or canonical versions of these vowels (closer, for example, in formant values to the less coarticulated V_1 versions; see also Manuel, 1987, p. 136). So that all stimuli underwent similar waveform manipulations, all stimuli were spliced into different utterances (e.g., the first and second syllables of [bib$_i$a] are from two different tokens of original [biba]); however, the splicing process introduced no detectable discontinuities in the signal.

3.1.2. *Listeners*

All listeners were recruited from introductory linguistics classes and were paid for their participation. Participants were 16 native English-speaking students (ages 18–21) at the University of Michigan and 23 native Shona-speaking students (ages 19–23) at the University of Zimbabwe in Harare. Several of the Shona speakers had difficulty with the task; listeners whose accuracy fell below 70% on either the Shona or English control task were excluded from the analysis. Analyses are therefore based on the responses of 16 native English speakers and 17 native Shona speakers. All Shona speakers were bilingual, being fluent speakers of English.

3.1.3. *Procedure*

Three randomized test sequences—control, canonical and noncanonical—were constructed for each language. Each English test sequence contained 108 trials (9 vowel pairings × 3 tokens × 4 repetitions), while each Shona sequence had 76 trials

[7]In the control trials listed in the tables, $V_1 \neq V_2$. For each type of vowel pairing, we also included a set of control trials in which $V_1 = V_2$. For example, for pairing [$_i$a]–[$_a$a], the additional control trial was [bab$_a$a–bab$_a$a] / [bab$_a$a–bab$_i$a]. Because the two sets of controls gave very similar results (less than 2% difference across all conditions), we include here—and in the results presentation—only the slightly more difficult of the two.

(4 repetitions of [(5 vowel pairings $\times$ 3 tokens) + (2 vowel pairings $\times$ 2 tokens)]). For all trials, the within-pair interval was 0.5 s, the inter-pair interval was 1 s, and the inter-trial interval was 3 s.

All listeners heard all test sequences, regardless of their native language. Stimuli were presented to groups of 1–4 listeners over Sennheiser HD 265 headphones via a digital Sony (TCD-D8) recorder. English listeners were tested in a sound-attenuated room in the Phonetics Laboratory at the University of Michigan. Shona listeners were tested in a quiet room in the Department of Linguistics at the University of Zimbabwe. The instructions to the Shona listeners were presented by a native Shona-speaking experimenter, although instructions were given in English (chosen by the Shona experimenter because English, as the national language of Zimbabwe, is the language of academic discourse). The same testing equipment was used for both groups of listeners, and the English-speaking experimenter who conducted the tests with the English listeners was also present for the tests with the Shona listeners.

Listeners completed the tests in two sessions over 2 days, with the English tests on one day and the Shona tests on the other; the order of test language was counterbalanced. For the English trials, listeners were asked to choose the pair in which the *second* vowels sounded more different; for the Shona trials, listeners were asked to choose the pair in which the *first* vowels sounded more different. (Therefore, the correct responses in Tables IV and V would always be the pair on the right, although the order of pair presentation was balanced in the actual test.) The 4IAX control sequence always preceded the test sequences, but the canonical and noncanonical sequences were counterbalanced. Listeners heard a 20-item control practice before completing the control condition and a 20-item test practice before completing the first of the two test conditions. Listeners were given the option to hear the practice more than once to make sure they understood the task and could detect some vowel differences.

3.1.4. *Predictions*

To facilitate interpretation of the results, we first consider the predictions of a theory of language-specific perception of coarticulated speech. Acoustically, English has more extensive carry-over V-to-V coarticulation than does Shona, especially on unstressed vowels. Shona and English have similar patterns of anticipatory coarticulation on stressed vowels (except for larger $F1$ effects at vowel offset in English). Thus, Shona listeners should compensate less than English listeners for carry-over effects (i.e., on the English test) and English listeners should compensate at least as much as Shona listeners for anticipatory effects (i.e., on the Shona test).

The 4IAX test offers two ways to test for language differences in perceptual compensation. One method is direct comparison of the percent correct responses on test trials by Shona vs. English listeners. In this test, a compensatory response would be selection of the *incorrect* pair member of a test trial: compensating listeners hear the acoustically identical V_1 (Shona) or V_2 (English) pairing as more different than the acoustically distinct pairing. This is because acoustically distinct vowels should sound similar if their differences can be attributed to coarticulatory context (e.g., [$_{i}$a] *vs*. [$_{a}$a] in [bib$_{i}$a]–[bab$_{a}$a]), while acoustically identical vowels should sound different if some acoustic properties can be assigned to context in one, but not the other, vowel (e.g., [$_{i}$a] in [bib$_{i}$a] but not [bab$_{i}$a]). Thus, Shona listeners should offer more correct (noncompensatory) responses than English listeners on the carry-over (English) test, while the two language groups should respond similarly to the anticipatory (Shona) test. This comparison method assumes similar baseline (control trials) performance across the two listener groups.

A second method of measuring cross-language differences, not based on this assumption, is to compare the listener groups' *pattern* of performance across the different trial types. If listeners especially compensate for *familiar* coarticulatory effects, then acoustically identical vowels should sound more different from each other in trials with noncanonical effects than in those with canonical effects. Canonical effects (e.g., [$_a$a], [$_i$i]) outside of their coarticulatory context ([$_a$a] in [bib$_a$a]) are roughly comparable to lack of coarticulation and are arguably similar to careful productions that might occur in natural speech. Thus, in a canonical trial such as [bab$_a$a-bib$_i$a]/[bab$_a$a-bib$_a$a], listeners who compensate might nevertheless hear [$_a$a] in [bab$_a$a] as similar to [$_a$a] in [bib$_a$a] because a production close to [bib$_a$a] might actually occur in citation-form speech. On the other hand, noncanonical effects (e.g., [$_i$a], [$_a$i]) do not normally occur outside of their coarticulatory context ([$_i$a] in [bab$_i$a]); they should sound peculiar out of that context and thus noticeably different than when in an appropriate context ([bib$_i$a]). For listeners familiar with the relevant coarticulatory patterns, noncanonical trials such as [bab$_a$a-bib$_i$a]/[bab$_i$a-bib$_i$a] should be especially likely to elicit compensatory responses. Consequently, for English listeners on the English test, and Shona and English listeners on the Shona test, the ranking of experimental conditions in decreasing order of accuracy should be control, then canonical, then noncanonical trials. Shona listeners responding to the English test should perform more similarly on the canonical and noncanonical trials. That is, because the large carry-over effects of English are unfamiliar to Shona listeners *even under coarticulatorily appropriate conditions*, noncanonical [$_i$a] should sound somewhat unusual in [bib$_i$a] as well as [bab$_i$a], and so would not be especially likely to elicit incorrect responses. We consider both comparison methods in evaluating the discrimination results.

3.2. *Results*

3.2.1. *Discrimination of English carry-over coarticulation*

The English tests investigated discrimination of vowels differing in the carry-over effects of stressed V_1 on unstressed V_2 for nine types of vowel pairings. Fig. 4 gives native English listeners' pooled responses to the nine pairings for the control, canonical, and noncanonical conditions. All vowel pairings show a similar pattern of responses across the experimental conditions. English listeners accurately selected the correct response—the "different" pairing—in the control trials (on average, 95% correct). As expected, English listeners were less accurate on the test conditions: correct responses dropped by an average of 17% from the control trial to the corresponding canonical trial and by another 9% to the corresponding noncanonical trial.

The Shona listeners' responses to the same English tests in Fig. 5 exhibit some similarities to the native English listeners' responses, but there are important differences as well. Although less accurate than the English listeners, Shona listeners reliably discriminated the control trials (on average, 84% correct). Like the English

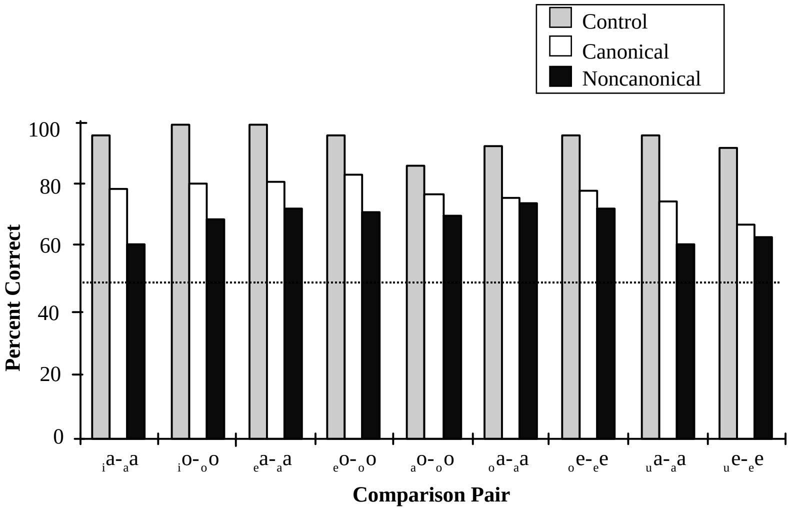

Figure 4. Experiment 2: Pooled percent correct responses, according to type of vowel pairing, of 16 native English listeners to the English 4IAX tests of discrimination of carryover coarticulatory effects. Bar type indicates the trial type for each pairing. Each pairing corresponds to a line in Table IV; e.g., for the left-most $_{i}$a-$_{a}$a pairing: ▨ = bib$_{i}$a-bib$_{i}$a/bib$_{i}$a-bib$_{a}$a; □ = bab$_{a}$a-bib$_{a}$a/bab$_{a}$a-bib$_{i}$a; ■ = bab$_{i}$a-bib$_{i}$a/bab$_{a}$a-bib$_{i}$a. The dashed horizontal line indicates chance (50%) performance.

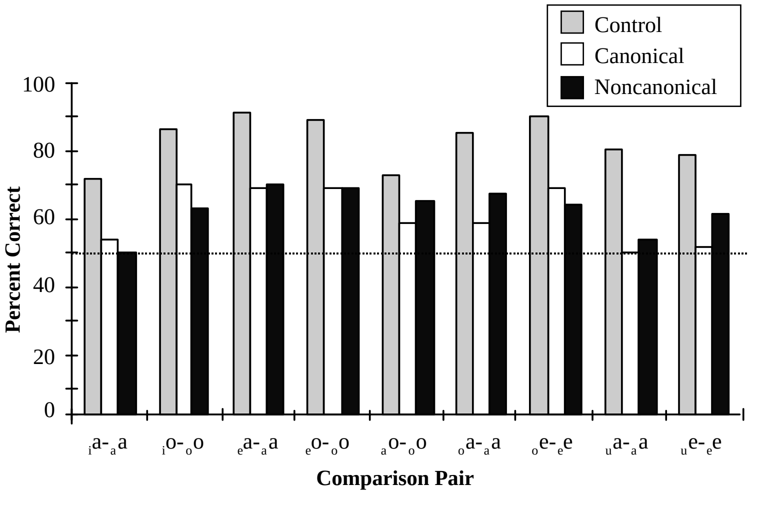

Figure 5. Experiment 2: Pooled percent correct responses, according to type of vowel pairing, of 17 native Shona listeners to the English 4IAX tests of discrimination of carryover coarticulatory effects. (Vowel pairings are as in Fig. 4.)

listeners, Shona listeners showed a systematic decrease in accuracy from the control to the test conditions, with Shona test discrimination being—unexpectedly—no better (and, indeed, somewhat poorer) than English test discrimination. However, unlike English listeners, trials with noncanonical effects were not consistently more difficult than those with canonical effects for Shona listeners: noncanonical trial performance was less accurate than canonical in only three (${}_{i}a$–${}_{a}a$, ${}_{i}o$–${}_{o}o$, ${}_{o}e$–${}_{e}e$) of the nine pairings. This outcome is not due to floor effects, as Shona responses to noncanonical trials were above 60% correct for all but two of the vowel pairings.

A two-way repeated measures ANOVA tested for Listener Language (English, Shona) and Condition (control, canonical, noncanonical), averaging across the nine vowel pairings. The ANOVA results showed a significant main effect of Listener Language [$F(1,31) = 14.21$, $p<0.001$], due to better overall performance of the native English listeners, and Condition [$F(2,62) = 109.90$, $p<0.0001$]. The interaction between these factors was also significant [$F(2,62) = 5.82$, $p<0.005$], which we explored through one-way ANOVAs for each language, testing for an effect of Condition. For English listeners, there was a main effect of Condition [$F(2,30) = 71.04$, $p<0.0001$], with pairwise comparisons showing significant differences for all condition types: control performance was better than test [$t(30) = 7.57$ for control–canonical and 11.76 for control–noncanonical, $p<0.0001$], and canonical performance was better than noncanonical [$t(30) = 4.19$, $p<0.001$]. For Shona listeners, there was again a main effect of Condition [$F(2,32) = 48.84$, $p<0.0001$], with control performance better than test [$t(32) = 8.95$ for control–canonical and 8.11 for control–noncanonical, $p<0.0001$]. However, Shona responses to the two test conditions were not significantly different (test means differed by only 2%; $p>0.4$). This result conforms to the prediction that, for listeners familiar with the target coarticulatory patterns (here, the English listeners), noncanonical effects are most likely to lead to compensatory responses—i.e., responses based on vowels' coarticulatory appropriateness rather than their acoustic identity.

The relation between the perceptual results and specific acoustic differences between paired target vowels is also revealing. For some vowel pairings, English responses showed less accurate discrimination of *control* pairings of vowels with relatively *small* acoustic differences, but less accurate discrimination of *test* (especially noncanonical) pairings of vowels with *large* acoustic differences. For example, although English listeners performed well on all control pairings, they were least accurate on control ${}_{a}o$–${}_{o}o$, with smaller acoustic differences (see especially midpoint measures in Table II). But among their most poorly discriminated noncanonical trials—i.e., trials with among the most compensatory responses—was the ${}_{i}a$–${}_{a}a$ pairing, with nearly perfect control performance and large $F1$ and $F2$ differences. In contrast, Shona listeners' performance on these non-native comparisons show a potential link between control and test condition detectability: with the exception of the ${}_{a}o$–${}_{o}o$ pairing, the coarticulatory differences that were hardest for Shona listeners to detect on the controls were similarly poorly discriminated on the corresponding tests.

3.2.2. *Discrimination of Shona anticipatory coarticulation*

The Shona tests investigated native Shona- and English-speaking listeners' discrimination of the anticipatory effects of unstressed V_2 on stressed V_1 for seven types of vowel pairings. The pooled responses to these pairings are given in Figs. 6 (English) and 7 (Shona). Responses to the control pairings show that, compared with these same listeners' responses to the English controls (Figs. 4 and 5), overall

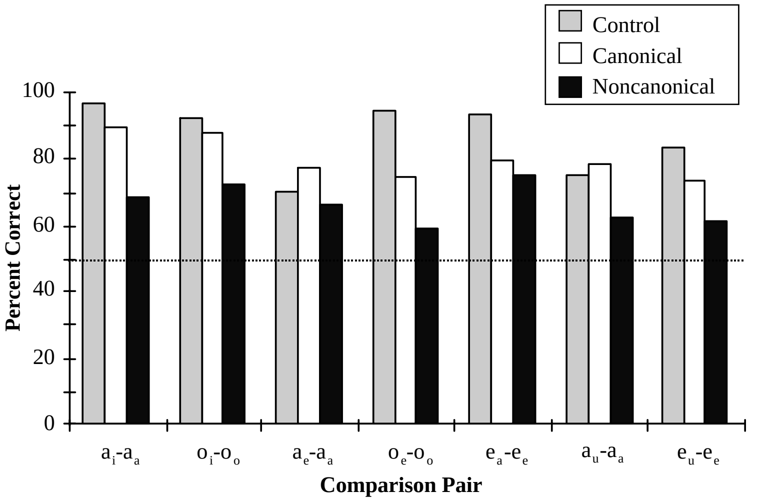

Figure 6. Experiment 2: Pooled percent correct responses, according to type of vowel pairing, of 16 native English listeners to the Shona 4IAX tests of discrimination of anticipatory coarticulatory effects. Bar type indicates the trial type for each pairing. Each pairing corresponds to a line in Table V; e.g., for the left-most a_i-a_a pairing: ▨ = pa_ipi-pa_ipi/pa_ipi-pa_api; □ = pa_apa-pa_api/pa_apa-pa_ipi; ■ = pa_ipa-pa_ipi/pa_apa-pa_ipi

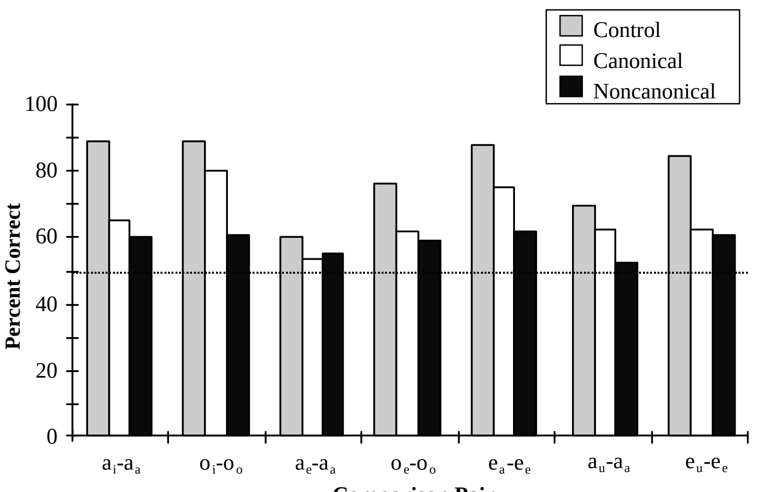

Figure 7. Experiment 2: Pooled percent correct responses, according to type of vowel pairing, of 17 native Shona listeners to the Shona 4IAX tests of discrimination of anticipatory coarticulatory effects. (Vowel pairings are as in Fig. 6.)

accuracy dropped, indicating that the tests with the Shona data were generally more difficult. However, the non-native (English) listeners' control performance dropped more than the native (Shona) listeners' performance (8% *vs*. less than 3%). An unexpected outcome was the difficulty of the a_e–a_a control pairing for both listener groups, where native-listener (Shona) performance fell below the criterion level established in pilot testing. Due to below-criterion control results, this pairing was omitted in the statistical analyses for the Shona 4IAX tests.

Recall that, because English and Shona listeners are both familiar with anticipatory coarticulatory effects of the magnitude used here, the two groups are predicted to show similar response patterns. This is what we find when we consider the relations among control, canonical, and noncanonical pairings. Across the vowel types, control pairings were discriminated more accurately than test pairings (85 *vs*. 68% correct), and test pairings with canonical effects were discriminated more accurately than those with noncanonical effects (74 *vs*. 63% correct). Excluding the below-criterion a_e–a_a pairing, Shona listeners showed this pattern for all six vowel pairings (Fig. 7) and English listeners for five of the six (Fig. 6). The pattern is further supported by the results of a two-way repeated measures ANOVA on listeners' responses to the Shona tests with factors Listener Language (English, Shona) and Condition (control, canonical, noncanonical). Listener Language [$F(1,31) = 8.56$, $p < 0.01$] was again significant due to better overall performance of the native English listeners, albeit by a smaller margin than for the English tests. Condition was also significant [$F(2,62) = 77.72$, $p < 0.0001$]. However, unlike the corresponding analysis for the English tests, Condition did not interact with Listener Language ($p > 0.4$), indicating that the listener groups did not differ significantly in how they responded to the experimental conditions. Pairwise comparisons exploring the effect of Condition showed that control performance was significantly better than test [$t(62) = 5.82$ for control–canonical and 11.50 for control–noncanonical, $p < 0.0001$], and canonical test performance was better than noncanonical [$t(62) = 5.68$, $p < 0.0001$].

Inspection of the acoustic differences for the paired anticipatory effects (Table III), compared to their discrimination accuracy, shows results similar to those observed for the English listeners for some of the English pairings. For example, the Shona pairing a_i–a_a elicited highly accurate control performance and showed large acoustic differences in $F1$ and $F2$ extending into vowel midpoint, while the pairings a_e–a_a and a_u–a_a had the poorest control performance and small acoustic differences at vowel midpoint. In contrast, test pairings generally do not show this link—exactly what we would expect if test responses were compensatory, as discussed below.

3.3. *Discussion*

We predicted that, if perceptual sensitivity to coarticulation is language-specific, two types of findings might emerge from the discrimination results. First, listeners whose native language provides systematic exposure to the targeted coarticulatory patterns should offer more compensatory responses than those whose language does not. That is, the listener group familiar with the coarticulatory patterns should make more mistakes consistent with judging vowel similarity on the basis of coarticulatory appropriateness rather than acoustic identity. Thus, English listeners should compensate more than Shona listeners for carry-over coarticulation due to Shona listeners' more limited experience with these effects. However, more compensation for familiar than unfamiliar coarticulatory patterns was not found, a finding we argue was largely due to native Shona listeners' overall greater difficulty with the task, including control conditions. We tentatively attribute this greater difficulty to their inexperience with comparable experimental test situations. (We expect that native English-speaking listeners, recruited from large introductory courses, were more likely to have participated in previous experiments.) Although testing conditions were held as constant as possible, testing in a quiet room for Shona listeners compared to a sound-attenuated room for English listeners may also have played a small role. Previous cross-language studies have also reported overall less accurate perception by one language group than another for reasons not obviously linked to linguistic experience (e.g., Polka, 1992; Pruitt, 1995; Harnsberger, 2001).

A second predicted difference was that, for listeners familiar with the targeted coarticulatory patterns, noncanonical trials—which placed coarticulatory effects in particularly unexpected contexts—should elicit more compensatory responses than canonical trials. This pattern of performance, which has the advantage of being independent of overall performance level, was systemically upheld. Shona and English listeners have considerable experience with anticipatory coarticulation and, correspondingly, both listener groups compensated more for noncanonical than canonical effects on the Shona test of anticipatory coarticulation. On the English test of carry-over coarticulation, native English—but not native Shona—listeners compensated more for noncanonical than canonical effects, consistent with the acoustic evidence that English listeners have more extensive native experience with carry-over V-to-V coarticulation. Importantly, the lack of performance drop from canonical to noncanonical trials for Shona listeners on the English carry-over test cannot be attributed to their overall less accurate performance. Overall accuracy was independent of whether or not individual Shona listeners showed the performance drop for the English stimuli (only five of the 17 did, in contrast to 16 out of 17 for the Shona stimuli). For example, the Shona listener who was most accurate on the English canonical trials (85% correct) was yet slightly more accurate (88%) on the noncanonical trials.

The results showed an intriguing relation between the acoustic properties of the stimuli and the perceptual responses they evoked in that, for listeners discriminating familiar coarticulatory patterns, vowel pairings with larger acoustic differences were among the most accurately discriminated in control trials and among the most poorly discriminated in noncanonical test trials. We attribute this difference to compensation on the latter trials. On control trials, where context is held constant and compensation is thus not at issue, listeners should be most accurate on vowel pairs with large—and presumably salient—acoustic differences. But these same salient differences on test trials, especially those involving noncanonical effects, should be particularly likely to mislead listeners if they reflect systematic coarticulatory effects in the listeners' native language.

For both listener groups, our interpretation that compensation for V-to-V coarticulation led to *mistakes* in perception is, of course, restricted to the experimental setting. We would argue that, in natural communicative settings, perception is *facilitated* when listeners attribute coarticulatory effects to their source.

4. Experiment 3: vowel identification in different coarticulatory contexts

To determine whether the language-specific response patterns obtained with the 4IAX paradigm also hold for Shona and English listeners' categorial (labeling) responses, Experiment 3 tested these listeners' identification of vowels in different phonetic contexts. We created synthetic disyllables in which members of a target vowel continuum occurred in either V_1 (anticipatory) or V_2 (carry-over) position flanked by either a front or a nonfront context vowel. Previous identification studies report context-induced shifts in the location of listeners' category boundary, with boundary shifts being consistent with compensation for the acoustic effects of the coarticulatory context (e.g., Mann & Repp, 1980; Krakow *et al*., 1988; Fowler *et al*., 1990; see footnote 1). Our hypothesis is that shifts in target vowel boundaries due to context vowels should reflect native-language coarticulatory patterns.

4.1. *Methods*

4.1.1. *Stimulus materials*

Four series of /ˈpVpV/ disyllables were generated on the Klatt synthesizer (KLSYN88) in its cascade configuration. All series involved a 10-step /a-e/ continuum varying in *F*1–*F*3. Two series tested carry-over effects of V_1 on V_2 identification and two series tested anticipatory effects of V_2 on V_1. In the "carry-over" series, V_1 was held constant at either /i/ or /a/ and target V_2 varied incrementally from /a/ to /e/: /pipa-pipe/ and /papa-pape/; in the "anticipatory" series, V_2 was either /i/ or /a/ and target V_1 varied: /papi-pepi/ and /papa-pepa/.

Spectral and temporal parameters of the stimuli were based on values for two female native speakers (analyzed in the acoustic study) of each language. The values chosen were rough averages and generally fell within the ranges of all four speakers. (As shown in Fig. 1, Shona and English counterparts for /i e a/ fall in similar regions of the *F*1/*F*2 space.) The synthesized endpoints were pilot tested with native speakers of both languages who were asked to identify the disyllables.

Table VI gives *F*1–*F*3 frequencies for context /i/ and /a/, and endpoint members of the target /a-e/ continuum. (Formant frequencies for context /a/ and continuum endpoint /a/ were the same.) *F*4 was held constant at 3250 Hz for all vowels except /i/, whose *F*4 was 3500 Hz. Eight intermediate stimuli between the endpoint target vowels were calculated by linear interpolation; *F*1–*F*3 step sizes were 43, 105 and 36 Hz, respectively. Corresponding target stressed (V_1) and unstressed (V_2) continuum members had the same formant frequencies; stress was primarily encoded by amplitude (level for stressed V_1 *vs*. 1 dB lower at onset and falling

TABLE VI. Experiment 3: *F*1–*F*3 frequencies (in Hz) for synthesized context and endpoint target vowels

	*F*1			*F*2			*F*3		
	ON	SS	OFF	ON	SS	OFF	ON	SS	OFF
i	180	350	180	2165	2650	2165	3150	3250	3150
e	180	460	275	1965	2450	1965	2925	3025	2895
a	565	850	665	1015	1500	1015	2600	2700	2570

Note: ON = onset, SS = steady state, OFF = offset.

for unstressed V_2) and $f0$ (12 Hz higher in stressed V_1 than in unstressed V_2 at steady state).

Each context and target V_1 and V_2 began with a 35-ms onset to steady-state noise-excited (aspirated) transition in $F1$–$F3$; V_1 ended with a 35-ms voiced $F1$–$F3$ transition followed by a 40-ms closure for /p/. Total disyllable duration was 450 ms, except for the /papi-pepi/ series with 410 ms stimuli due to relatively short, unstressed context /i/.

4.1.2. *Procedure and listeners*

Randomized test sequences for each of the four continua—/pipa-pipe/, /papa-pape/, /papi-pepi/, and /papa-pepa/—were constructed, each with 100 identification trials (10 stimuli $\times$ 10 repetitions). Within each test, stimuli were grouped into blocks of 10 trials; blocks were separated by 6 s, otherwise the inter-trial interval was 3 s. To familiarize listeners with the stimuli, each test was preceded by a 10-item practice series consisting of one token of each stimulus in random order.

Participants were the same 17 native speakers of Shona and 16 native speakers of English who participated in the discrimination task. The same general testing procedures were used. The identification task was administered immediately after the corresponding 4IAX task: the two anticipatory tests of V_1 identification followed discrimination of Shona V_1s, and the carry-over tests of V_2 identification followed discrimination of English V_2s. (All tests, discrimination and identification, were therefore completed by listeners over 2 days.) Listeners were asked to decide whether the continuum vowel sounded like /e/ or /a/ by circling the appropriate vowel on their answer sheet. For the Shona listeners, these sounds were represented by *e* and *a*, consistent with Shona orthography; for English listeners, the representations were *ay* and *ah*, respectively.

4.1.3. *Predictions*

Compensation predicts that listeners will hear fewer /e/s when the context vowel is /i/ than when it is /a/. That is, for phonetically ambiguous stimuli, listeners should accept lower $F1$ and higher $F2$ frequencies as /a/ in a high front vowel context—a context that would, in natural coarticulated speech, be expected to have these acoustic effects on the target. (An auditory theory of frequency contrast would yield the same general prediction.)

Language differences should emerge in the magnitude of the perceptual effects of context. If language-specific coarticulatory patterns influence category judgments, then carry-over contexts (/papa-pape/ *vs*. /pipa-pipe/) should trigger a larger /a-e/ boundary shift for English listeners than for Shona listeners. In anticipatory contexts (/papa-pepa/ *vs*. /papi-pepi/), boundary shifts should be more comparable for the two language groups. A within-language prediction, based on acoustic evidence of greater anticipatory than carry-over coarticulation on Shona vowels (including, specifically, target /a/ and /e/; see footnote 6), is that Shona listeners should show a larger boundary shift due to anticipatory than carry-over contexts. For English, a clear within-language prediction does not emerge from the overall acoustic findings, which offer a mixed picture regarding carry-over *vs*. anticipatory effects when $F1$ and $F2$ are considered.

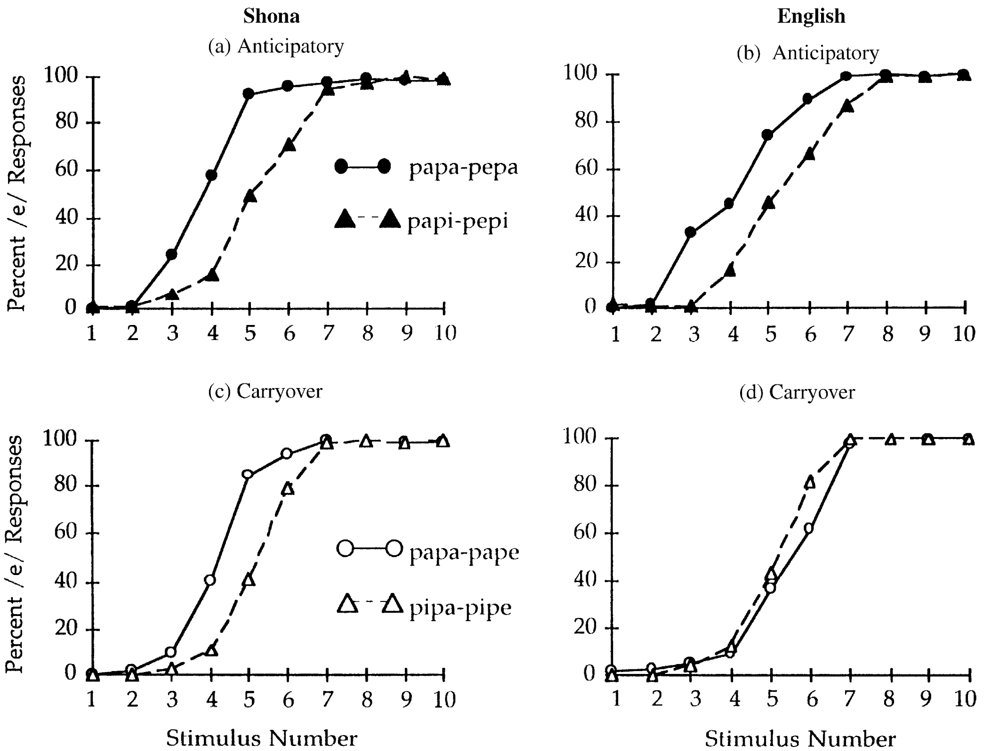

Figure 8. Experiment 3: Pooled percent /e/ responses of the 17 Shona (left panels) and 16 English (right) listeners to the anticipatory (top) and carry-over (bottom) series for the /a-e/ continuum. Front /i/ context is indicated by dashed lines and back /a/ context by solid lines.[8]

4.2. *Results*

The identification functions in Fig. 8 show the percent /e/ responses of the Shona (left panels) and English (right) listeners for the anticipatory (top) and carry-over (bottom) series.[8] Responses to the corresponding /a/ and /i/ contexts (i.e., each panel in Fig. 8) show that, in three of the four comparisons (all but Fig. 8(d)), listeners heard more /e/s in the /a/ context. These three cases fit the predicted effect of vowel context on target vowel identification. We attribute the generally more gradual /a–e/ crossovers for the English than Shona functions to differences in the vowel inventories: for native English—but not Shona—listeners, the /a-e/ continuum passed through phonemic /ɛ/. The unexpected pattern of English listeners' responses to the carry-over /papa-pape/ series is considered in detail in the Discussion.

[8]In constructing the test for the carry-over /papa–pape/ series, we erroneously omitted stimulus 7 from the series. Therefore, in Fig. 8(c) and (d), the data point (circle) for stimulus 7 is omitted, as is the connecting solid line. In the statistical analyses, each listener's value for stimulus 7 for this series was interpolated from that listener's percent /e/ responses for stimuli 6 and 8. (To justify the interpolation, we tested several additional English listeners on the corrected test. Their responses to stimulus 7 were as expected.) For all Shona and English listeners, stimulus 7 occurred *after* the 50% crossover from /a/ to /e/.

Table VII. Experiment 3: Mean 50% crossover values (in stimulus numbers) for the Probit identification functions

Coarticulatory direction	Context vowel	Listener language	
		Shona	English
Anticipatory	/a/	3.84	4.12
	/i/	5.15	5.35
Carryover	/a/	4.21	5.35
	/i/	5.20	5.10

Note: Lower values indicate more /e/ responses. (Anticipatory /a/ and /i/ = /papa–pepa/ and /papi–pepi/, respectively; carry-over /a/ and /i/ = /papa–pape/ and /pipa–pipe/, respectively.)

Statistical analyses were performed on the 50% /a-e/ crossover values calculated for each listener's judgments of the vowels in each of the four series using Probit Analysis (Finney, 1971). Mean crossover values are given in Table VII; values correspond to stimulus numbers, with lower values indicating more /e/ responses. Crossover values show that responses to vowels in the /i/, but not the /a/, context remained relatively stable across languages and directions (the crossover range for the /a/ context being six times that of the /i/ context).

A three-way repeated measures ANOVA testing for Listener Language (Shona, English), Direction (anticipatory, carry-over) and Context Vowel (/i/, /a/) was conducted on the crossover values. The significant main effects were Context Vowel [$F(1,31) = 31.63$, $p < 0.0001$], due to earlier crossovers for the /a/ context, and Direction [$F(1,31) = 9.28$, $p < 0.01$], attributable to overall earlier crossovers for the anticipatory series. These two factors significantly interacted with each other [$F(1,31) = 22.19$, $p < 0.0001$], and participated in a three-way interaction with Listener Language [$F(1,31) = 9.03$, $p < 0.01$]. In assessing the three-way interaction, we tested, in each direction, for Listener Language and Context Vowel. For the anticipatory series, only Context Vowel was significant [$F(1,31) = 36.07$, $p < 0.0001$], indicating that English and Shona listeners compensated similarly—as expected—for anticipatory effects (top panels of Fig. 8). In contrast, the carry-over series showed significant main effects of Listener Language [$F(1,31) = 4.19$, $p < 0.05$], Context Vowel [$F(1,31) = 7.95$, $p < 0.01$], and their interaction [$F(1,31) = 22.41$, $p < 0.0001$]. Although the language groups were expected to differ in their identification of the carry-over series, the nature of the difference was not as predicted: for carry-over, Shona listeners showed a larger boundary shift due to Context Vowel [$F(1,16) = 28.48$, $p < 0.0001$] than did English listeners (bottom panels). English boundaries for the carry-over series did not significantly differ for the /a/ *vs.* /i/ contexts.

A within-language prediction based on the acoustic results was that Shona listeners should compensate more (show larger boundary shifts) for anticipatory (Fig. 8(a)) than carry-over (Fig. 8(c)) effects. To test this prediction, we calculated, for each Shona listener, the *difference* between the /a/- and /i/-context crossovers for the anticipatory series and compared that value with the crossover difference for the carry-over series. A one-tailed paired *t*-test conducted on the boundary-shift differences supported the expected pattern [$t(16) = 1.87$, $p < 0.04$].

4.3. *Discussion*

The findings demonstrate long-distance, transconsonantal influences on vowel identification. Shona listeners offered more target /e/ responses when the flanking vowel was /a/ than when it was /i/, regardless of the V_1 or V_2 position of the target; English listeners did the same when the target vowel was V_1. This overall pattern is expected if listeners attribute spectral characteristics of the target to a context vowel, accepting as /e/ ambiguous vowels that sound relatively low and back in a context expected to have such coarticulatory effects. It is also the pattern expected on the basis of a noncoarticulatory, auditory account such as frequency contrast (e.g., Lotto & Kluender, 1998): the target vowel's $F1$ should sound even lower than it actually is in the context of /a/'s higher $F1$, while the target vowel's $F2$ should sound even higher than it actually is in the context of /a/'s lower $F2$. Moreover, this same perceptual pattern is consistent with a phonetic or phonological contrast effect: ambiguous target vowels near the /a-e/ boundary are unlike context /a/, leading to fewer /a/ responses in an /a/ context.

Although the identification results do not tease apart these accounts, a particularly strong test of a link between identification responses and language-specific coarticulatory behavior would be to show that the size of the shift in listeners' identification boundary compares favorably with the size of speakers' coarticulatory effects on these vowels as measured in the acoustic study. A close link emerges for the Shona data. Fig. 9 plots the acoustic effects of /i e a o u/ contexts on Shona target /a/, and shows that the $F2$ difference between the /i/ *vs.* /a/ contexts is 132 Hz for anticipatory effects on stressed vowels and 102 Hz for carry-over effects on unstressed vowels. These values are paralleled by the perceptual shifts: the Shona identification boundary shift of 1.31 stimuli for the /i/ *vs.* /a/ anticipatory contexts (Table VII) corresponds to 140 Hz in $F2$, while the Shona boundary shift of 0.98 stimuli for the carry-over contexts is a 103 Hz change in $F2$.

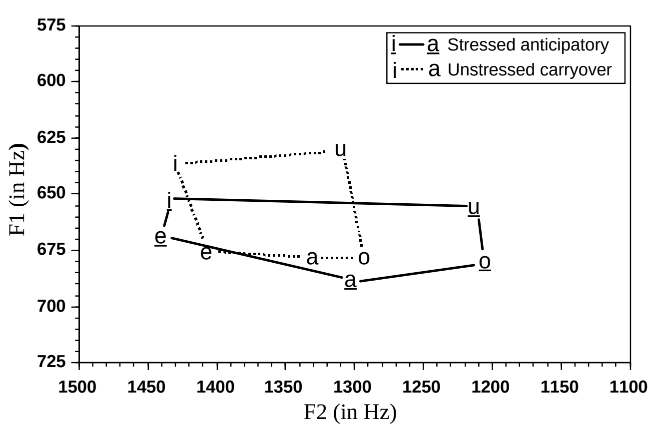

Figure 9. Shona V-to-V coarticulation: anticipatory effects on $F1$ and $F2$ of stressed target /a/ (solid line) and carry-over effects on unstressed target /a/ (dashed line) of the context vowels /i e a o u/, averaging across seven speakers.

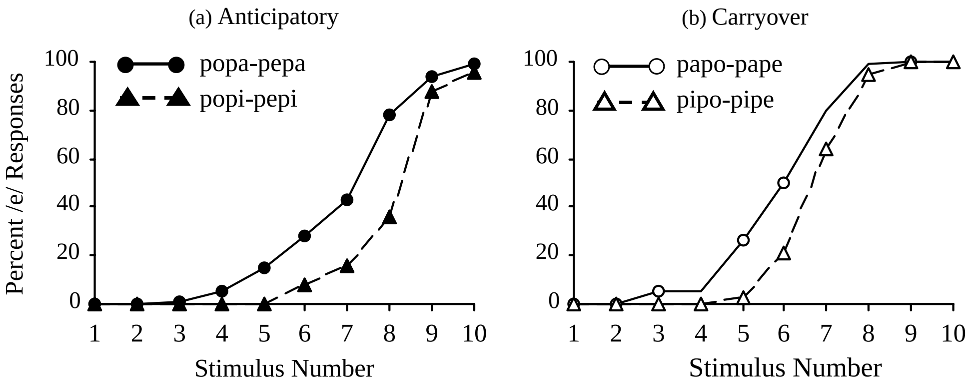

Figure 10. Experiment 3 follow-up: Pooled percent /e/ responses of eight English listeners to the anticipatory (left) and carry-over (right) series for the /o-e/ continuum. Front /i/ context is indicated by dashed lines and back /a/ context by solid lines.

However, one finding was not expected: English listeners failed to show a carry-over effect of vowel context on the target /a-e/ boundary. Our account based on acoustic findings for English V-to-V coarticulation, as well as contrast accounts, would predict a boundary shift for the carry-over series comparable to that for the anticipatory series. We attribute the lack of shift to factors specific to English phonology. Although English unstressed vowels are often reduced, final unstressed /a/ is particularly likely to be reduced to [ə]. We speculate that listeners were willing to label ambiguous unstressed /a-e/ continuum vowels with intermediate $F1$ and $F2$ values as /a/ because reduced /a/ ([ə]) is higher than [a] and lower than [e]. Arguably, this phonological influence was sufficiently robust that it overrode any compensatory (or contrast) effects of coarticulatory context.

To test our phonological explanation of the English listeners' responses to the carry-over series, we generated a new target vowel continuum that did not pass through the reduced /a/ region: /o-e/. The /e/ endpoint was that of the /a-e/ continuum (see Table VI). Across the 10-step /o-e/ continuum, $F1$ was held constant; $F2$ increased from /o/ to /e/ in 157 Hz increments (beginning at 1040 Hz), and $F3$ increased in 18 Hz increments (beginning at 2860 Hz). Continuum vowels were embedded in the same contexts as before, yielding two anticipatory series, /popi-pepi/ and /popa-pepa/, and two carry-over series, /pipo-pipe/ and /papo-pape/. The pooled identification responses by a new group of eight English listeners to these series are given in Fig. 10. Both the anticipatory and carry-over series elicited more target /e/ responses in the /a/ context. That the /o-e/ series showed a carry-over effect of vowel context is consistent with our suggestion that the absence of a comparable effect for the /a-e/ series was due to the phonological behavior of English /a/. Although the carry-over effect is smaller than the anticipatory effect, which is contrary to expectations based on overall English $F2$ coarticulatory patterns, it is in fact consistent with expectations based on /e/ in particular. Like the Shona results for the /a-e/ series, the size of the English listeners' boundary shift for the /o-e/ series compares favorably to the coarticulatory effects found in Experiment 1. As seen in Fig. 11, which plots the acoustic effects of the context vowels on English target /e/, the $F2$ effect of the /i/ *vs.* /a/ contexts is 70 Hz larger in the

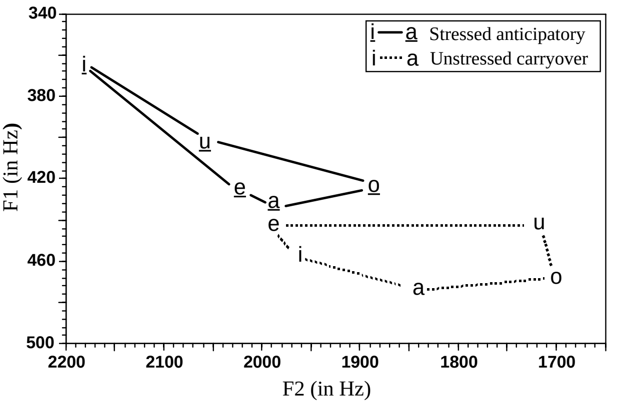

Figure 11. English V-to-V coarticulation: Anticipatory effects on $F1$ and $F2$ of stressed target /e/ (solid line) and carry-over effects on unstressed target /e/ (dashed line) of the context vowels /i e a o u/, averaging across five speakers.

stressed anticipatory condition than in the unstressed carry-over condition. This acoustic difference is close to the 65 Hz difference in the corresponding identification data (Fig. 10), where the /o-e/ boundary shift due to anticipatory contexts corresponds to 170 Hz in $F2$ while that due to carry-over contexts is 105 Hz.

To summarize, the predicted cross-language difference in listener identification of the /a-e/ carry-over series did not emerge, apparently due to phonological influences of English unstressed /a/. However, the within-language perceptual patterns that emerge with the combined results of the /a-e/ and /o-e/ series, when compared with the corresponding acoustic data, point toward listener sensitivity to native-language coarticulatory patterns.

5. General discussion

The results of Experiment 1 detail the nature of V-to-V coarticulation in English and Shona, and demonstrate that these languages systematically differ in the ways in which vowels in flanking syllables influence each other. One difference is that, while English stressed vowels have especially large coarticulatory influences on unstressed vowels, Shona showed comparable influences on stressed and unstressed vowels (see also Manuel, 1987). A second coarticulatory difference between the two languages pertains to direction of influences: carry-over effects were larger, both in $F1$ and $F2$, in English than in Shona. We have speculated that these coarticulatory differences may be linked to prosodic differences between English and Shona, but at this stage our account only raises intriguing possibilities, rather than compelling evidence. Shona and English were more similar in anticipatory effects, particularly in $F2$: although Shona anticipatory effects were larger than Shona carry-over, they were not statistically larger than English anticipatory effects.

Experiments 2 and 3 explored the influences of native coarticulatory patterns on Shona and English listeners' discrimination and identification of coarticulated vowels. We hypothesized that listeners' coarticulatory experience gives rise to expectations about the acoustics of coarticulation, and consequently to language-specific patterns of perceptual compensation for coarticulation. Although not all of the hypothesized cross-language perceptual differences were found, both listener groups were particularly sensitive to native(-like) coarticulatory effects. In Experiment 2, this sensitivity emerged in listeners' responses to vowels spliced into inappropriate coarticulatory contexts. A systematic outcome was that, the more inappropriate the context (noncanonical as compared to canonical trials), the more listeners compensated (i.e., based their responses on coarticulatory appropriateness rather than acoustic identity)—but only as long as the paired *appropriate* coarticulatory effects were *familiar* to the listeners. That is, Shona listeners showed this compensation pattern for native anticipatory effects, but not non-native (English) carry-over effects. English listeners showed this pattern for both English carry-over and Shona anticipatory conditions, consistent with English listeners' general familiarity with both types of coarticulatory patterns.

In Experiment 3, Shona and English listeners' identification responses to synthetic vowels embedded in different vowel contexts also exhibited their sensitivity to native V-to-V coarticulation. As expected, Shona listeners' /a-e/ boundary shifted more due to (perceived) anticipatory than carry-over influences of flanking /i/ and /a/, with the magnitude of the perceptual shift mirroring the magnitude of the coarticulatory effects of these vowels in Shona. English listeners' perception of the /a-e/ series was apparently heavily influenced by the phonological behavior of unstressed /a/ in English, but their identification of an /o-e/ continuum (which avoided this problem) showed boundary shifts that again paralleled the corresponding acoustic shifts found in Experiment 1.

That perception of coarticulatory effects of vowels on each other is influenced by native(-like) coarticulatory patterns contributes to the literature demonstrating that listeners are highly attuned to the coordination of articulatory gestures. Our findings suggest that listeners' assessment of these effects does not rely on the transient acoustic signal alone, but rather also depends on listeners' prior experience. This outcome conflicts with Fowler *et al.*'s (1990) language-general interpretation of perceptual compensation (which was based in part on evidence of compensation by prelinguistic infants), but builds on previous findings that mature listeners are especially attuned to the phonetic and phonological structures of their native language (see Strange, 1995, for a review).

Given the parallels found in important aspects of the production and perception data, we view our results as particularly compatible with a coarticulatory interpretation of listeners' discrimination and identification of V-to-V influences, with listeners ascribing more or less of the acoustic effects of coarticulation to their source depending on native-language coarticulatory structures. At the same time, our approach was not designed to tease apart coarticulatory accounts of perceptual compensation from auditory accounts of compensatory-like responses, and the findings allow for other interpretations. However, our data argue against any approach—coarticulatory or auditory—that does not incorporate experiential learning due to listeners' exposure to systematic variation in speech. Within both coarticulatory and auditory approaches, aspects of our data point toward listeners having learned patterns of coarticulation/covariation and interpreting the input signal relative to these patterns.

An aspect of our data that may seem troubling for a coarticulatory account is that, even when listeners were responding to native coarticulatory patterns, perceptual compensation was not perfect. For example, in discrimination, consistent compensation should have led to *below*-chance (50%) accuracy. However, accuracy on any trial rarely fell below 55% correct. Apparently, for both Shona and English listeners, while some of the acoustic influences of a context vowel on a target were (appropriately) assigned to the context, there remained a coarticulatory residue which enabled listeners to accurately discriminate vowel differences more often than not. Such partial compensation for coarticulatory influences is consistent with Fowler & Smith's (1986) results for V-to-V coarticulation, as well as with Beddor & Krakow's (1999) and Fowler & Brown's (2000) findings for nasal coarticulation.

We regard partial perceptual compensation as compatible with, and quite likely a consequence of, the considerable within-language variability in coarticulation. For example, Experiment 1 provided evidence of variability in production due to different speakers, segmental factors (e.g., vowels differed in their likelihood to undergo and to trigger coarticulation), and prosodic (e.g., stressed *vs.* unstressed) context; variability in production is also a common theme in the V-to-V coarticulation literature reviewed in the Introduction. The picture that emerges is that listeners are sensitive to language-specific coarticulation, but there is sufficient variability in coarticulation that perfect compensation does not occur. This outcome is the perceptual counterpart to Lindblom's (1990) observation that most articulatory compensations (e.g., in bite-block speech) to maintain perceptual constancy are partial, not perfect. Despite compensation in both domains, acoustic equivalence (real or perceived) is not reached in either domain.

The relation between patterns of coarticulatory organization and perceptual compensation has implications for theories of the factors underlying phonological behavior. Ohala (1981, 1986, 1993) speculated that perceptual compensation in natural communicative interactions might not be perfect, with listeners attributing too little or too much of the acoustic structure of a target sound to coarticulatory context. If too little were ascribed to context, these coarticulatory effects would be heard as intrinsic to the target, potentially leading to phonologization—i.e., systematic incorporation into the sound system—of coarticulatory variation. We and others (e.g., Ohala, 1994; Beddor & Yavuz, 1995; Majors, 1998; Beddor, Krakow & Lindemann, 2001) have suggested that V-to-V coarticulation, in which the articulation of one vowel has transconsonantal effects on other vowels, may evolve into phonological vowel harmony systems, in which vowels within a root or word necessarily share one or more features, such as lip rounding or tongue body height. (Along these lines, we would suggest that height harmony in Shona verb roots, and more generally the height harmony system widespread in Bantu languages, had its origins in V-to-V coarticulation, even though the phonological process may no longer be directly reflected in current coarticulatory patterns.) Our overall finding that compensation is partial and attuned to native-language patterns of coarticulation offers a possible listener-mediated mechanism for the evolution of vowel harmony in specific phonological systems.

The research reported here was supported in part by NSF Grant SBR 9319597 to Patrice Beddor, and a University of Michigan Department of Linguistics/Graduate School Research Partnership to Stephanie Lindemann and Patrice Beddor. A subset of these data was presented to the 14th International Congress of Phonetic Sciences (1999) and appears in those proceedings. We thank Calisto Mudzingwa (University of Zimbabwe), Chutamanee Onsuwan, and Alicia Beckford Wassink (University of Michigan) for data collection assistance and Kathy Welch for statistical assistance. We are also grateful to Juliet Thondhlana (University of Zimbabwe) and Brad Rakerd (Michigan State University) for the generous use of their facilities. Finally, this work has benefited from helpful discussions with, or comments from, José Benkí, Laura Downing, San Duanmu, Kathleen Evans-Romaine, Stefan Frisch, editor Jonathan Harrington, Rena Krakow, Sharon Manuel, Sylvia Mupepi, Scott Myers, Terry Nearey, and Marija Tabain.

References

Alfonso, P. J. & Baer, T. (1982) Dynamics of vowel articulation, *Language and Speech*, **25,** 151–173.

Beckman, J. N. (1997) Positional faithfulness, positional neutralization and Shona vowel harmony, *Phonology*, **14,** 1–46.

Beddor, P. S. & Krakow, R. A. (1999) Perception of coarticulatory nasalization by speakers of English and Thai: evidence for partial compensation, *Journal of the Acoustical Society of America*, **106,** 2868–2887.

Beddor, P. S., Krakow, R. A. & Lindemann, S. (2001) Patterns of perceptual compensation and their phonological consequences. In *The role of perceptual phenomena in phonology* (E. Hume & K. Johnson, editors), pp. 55–78. New York: Academic Press.

Beddor, P. S. & Yavuz, H. K. (1995) The relation between vowel-to-vowel coarticulation and vowel harmony in Turkish. In *Proceedings of the thirteenth international congress of phonetic sciences* (K. Elenius & P. Branderud, editors), Vol. 2, pp. 44–51. Stockholm, Sweden: KTH and Stockholm University.

Bell-Berti, F. & Harris, K. S. (1976) Some aspects of coarticulation, *Haskins Laboratories Status Reports on Speech Research*, **SR-45/46,** 197–204.

Boyce, S. E. (1990) Coarticulatory organization for lip rounding in Turkish and English, *Journal of the Acoustical Society of America*, **88,** 2584–2595.

Bradlow, A. R. (1995) A comparative acoustic study of English and Spanish vowels, *Journal of the Acoustical Society of America*, **97,** 1916–1924.

Byrd, D., Kaun, A., Narayanan, S. & Saltzman, E. (2000) Phrasal signatures in articulation. In *Acquisition and the lexicon: papers in laboratory phonology V* (M. B. Broe & J. B. Pierrehumbert, editors), pp. 70–87. Cambridge, U.K.: Cambridge University Press.

Choi, J. D. & Keating, P. (1991) Vowel-to-vowel coarticulation in three Slavic languages, *UCLA Working Papers in Phonetics*, **78,** 78–86.

Clark, M. & Sharf, D. J. (1973) Coarticulation effects of post-consonantal vowels on the short-term recall of pre-consonantal vowels, *Language and Speech*, **16,** 67–76.

de Jong, K., Beckman, M. E. & Edwards, J. (1993) The interplay between prosodic structure and coarticulation, *Language and Speech*, **36,** 197–212.

Evans-Romaine, D. K. (1998) *Palatalization and coarticulation in Russian*. Doctoral dissertation, University of Michigan, Ann Arbor.

Farnetani, E. (1990) V-C-V lingual coarticulation and its spatiotemporal domain. In *Speech production and speech modelling* (W. J. Hardcastle & A. Marchal, editors), pp. 93–130. Dordrecht, The Netherlands: Kluwer Academic.

Farnetani, E., Vagges, K. & Magno-Caldognetto, E. (1985) Coarticulation in Italian /VtV/ sequences: a palatographic study, *Phonetica*, **42,** 78–99.

Finney, D. J. (1971) *Probit analysis*. Cambridge, U.K.: Cambridge University Press.

Fortune, G. (1955) *An analytical grammar of Shona*. London: Longmans, Green and Co.

Fowler, C. A. (1981) Production and perception of coarticulation among stressed and unstressed vowels, *Journal of Speech and Hearing Research*, **46,** 127–139.

Fowler, C. A. (1986) An event approach to the study of speech perception from a direct-realist perspective, *Journal of Phonetics*, **14,** 3–28.

Fowler, C. A. (1996) Listeners do hear sounds, not tongues, *Journal of the Acoustical Society of America*, **99,** 1730–1741.

Fowler, C. A., Best, C. T. & McRoberts, G. W. (1990) Young infants' perception of liquid coarticulatory influences on following stop consonants, *Perception and Psychophysics*, **48,** 559–570.

Fowler, C. A. & Brown, J. M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels, *Perception and Psychophysics*, **62,** 21–32.

Fowler, C. A., Brown, J. M. & Mann, V. A. (2000) Contrast effects do not underlie effects of preceding liquids on stop — consonant identification by humans, *Journal of Experimental Psychology: Human Perception and Performance*, **26,** 877–888.

Fowler, C. A. & Smith, M. R. (1986) Speech perception as "vector analysis": an approach to the problems of segmentation and invariance. In *Invariance and variability of speech processes* (J. Perkell & D. Klatt, editors), pp. 123–139. Hillsdale, NJ: Erlbaum.

Harnsberger, J. D. (2001) On the relationship between identification and discrimination of non-native nasal consonants, *Journal of the Acoustical Society of America*, **110,** 489–503.

Keating, P., Cho, T., Fougeron, C. & Hsu, C-H. (in press) Domain-initial articulatory strengthening in four languages. In *Papers in laboratory phonology VI* (J. Local, R. Ogden & R. Temple, editors), pp. 143–161. Cambridge, U.K.: Cambridge University Press.

Kondo, Y. & Arai, Y. (1998) Prosodic constraint on V-to-V coarticulation in Japanese. In *Proceedings of the fifth international conference on spoken language processing (ICSLP 98)* (R. H. Mannell & J. Robert-Ribes, editors), pp. 1755–1758. Sydney, Australia: Australian Speech Science and Technology Association.

Krakow, R. A., Beddor, P. S., Goldstein, L. M. & Fowler, C. A. (1988) Coarticulatory influences on the perceived height of nasal vowels, *Journal of the Acoustical Society of America*, **83,** 1146–1158.

Ladefoged, P. (1984) 'Out of chaos comes order'; physical biological, and structural patterns in phonetics. In *Proceedings of the tenth international congress of phonetic sciences* (M. P. R. van den Broecke & A. Cohen, editors), pp. 83–95. Dordrecht, The Netherlands: Foris.

Ladefoged, P. (1989) Representing phonetic structure, *UCLA Working Papers in Phonetics,* **73,** 1–79.

Lehiste, I. & Shockey, L. (1972) On the perception of coarticulation effects in English VCV syllables, *Journal of Speech and Hearing Research*, **15**, 500–506.

Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H&H theory. In *Speech production and speech modelling* (W. J. Hardcastle & A. Marchal, editors), pp. 403–439. Dordrecht, The Netherlands: Kluwer Academic.

Lindblom, Guion, S., Hura, S., Moon, S.-J. & Willerman, R. (1995) Is sound change adaptive?, *Rivista di Linguistica*, **7,** 5–36.

Lotto, A. J. & Kluender, K. R. (1998) General contrast effects in speech perception: effect of preceding liquid on stop consonant identification, *Perception and Psychophysics*, **60,** 602–619.

Lotto, A. J., Kluender, K. R. & Holt, L. L. (1997) Perceptual compensation for coarticulation by Japanese quail, *Journal of the Acoustical Society of America*, **102,** 1134–1140.

Magen, H. S. (1984) Vowel-to-vowel coarticulation in English and Japanese, *Journal of the Acoustical Society of America*, **75,** S41.

Magen, H. S. (1997) The extent of vowel-to-vowel coarticulation in English, *Journal of Phonetics*, **25,** 187–205.

Majors, T. (1998) *Stress-dependent harmony: phonetic origins and phonological analysis*. Doctoral dissertation, University of Texas, Austin.

Mann, V. A. (1980) Influence of preceding liquid on stop-consonant perception, *Perception and Psychophysics*, **28,** 407–412.

Mann, V. A. (1986) Distinguishing universal and language-dependent levels of speech perception: evidence from Japanese listeners' perception of "l" and "r", *Cognition*, **24,** 169–196.

Mann, V. A. & Repp, B. H. (1980) Influence of vocalic context on perception of the [ʃ]-[s] distinction, *Perception and Psychophysics*, **28,** 213–228.

Manuel, S. Y. (1987) *Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages*. Doctoral dissertation, Yale University, New Haven, CT.

Manuel, S. Y. (1990) The role of contrast in limiting vowel-to-vowel coarticulation in different languages, *Journal of the Acoustical Society of America*, **88,** 1286–1298.

Manuel, S. Y. (1999) Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts. In *Coarticulation: theory, data and techniques* (W. J. Hardcastle & N. Hewlett, editors), pp. 179–198. Cambridge, U.K.: Cambridge University Press.

Manuel, S. Y. & Krakow, R. A. (1984) Universal and language particular aspects of vowel-to-vowel coarticulation, *Haskins Laboratories Status Reports on Speech Research*, **SR-77/78,** 69–78.

Marchal, A. & Hardcastle, W. J. (1993) ACCOR: instrumentation and database for the cross-language study of coarticulation, *Language and Speech*, **36,** 137–153.

Martin, J. G. & Bunnell, H. T. (1981) Perception of anticipatory coarticulation effects, *Journal of the Acoustical Society of America*, **69,** 559–567.

Martin, J. G. & Bunnell, H. T. (1982) Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences, *Journal of Experimental Psychology: Human Perception and Performance*, **8,** 473–488.

Ohala, J. J. (1981) The listener as a source of sound change. In *Papers from the parasession on language and behavior* (C. S. Masek, R. A. Hendrick & M. F. Miller, editors), pp. 178–203. Chicago, IL: Chicago Linguistic Society.

Ohala, J. J. (1986) Phonological evidence for top-down processing in speech perception. In *Invariance and variability of speech processes* (J. S. Perkell & D. H. Klatt, editors), pp. 386–401. Hillsdale, NJ: Erlbaum.

Ohala, J. J. (1993) Coarticulation and phonology, *Language and Speech,* **36,** 155–170.

Ohala, J. J. (1994) Towards a universal, phonetically-based theory of vowel harmony. In *Proceedings of the third international conference on spoken language processing (ICSLP 94)*, pp. 491–494. Yokohama, Japan: IEEE Press.

Öhman, S. E. G. (1966) Coarticulation in VCV utterances: spectrographic measurements, *Journal of the Acoustical Society of America*, **39,** 151–168.

Parush, A., Ostry, D. J. & Munhall, K. G. (1983) A kinematic study of lingual coarticulation in VCV sequences, *Journal of the Acoustical Society of America*, **74,** 1115–1125.

Petersen, N. R. (1986) Perceptual compensation for segmentally conditioned fundamental frequency perturbation, *Phonetica*, **43,** 31–42.

Polka, L. (1992) Characterizing the influence of native language experience on adult speech perception, *Perception and Psychophysics*, **52,** 37–52.

Pruitt, J. S. (1995) *The perception of Hindi dental and retroflex stop consonants by native speakers of Japanese and American English*. Doctoral dissertation, University of South Florida, Tampa.

Recasens, D. (1987) An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences, *Journal of Phonetics*, **15,** 299–312.

Recasens, D. (1999) Lingual coarticulation. In *Coarticulation: theory, data and techniques* (W. J. Hardcastle & N. Hewlett, editors), pp. 80–104. Cambridge, U.K.: Cambridge University Press.

Recasens, D., Pallarès, M. D. & Fontdevila, J. (1998) An electropalatographic and acoustic study of temporal coarticulation for Catalan dark /l/ and German clear /l/, *Phonetica*, **55,** 53–79.

Rosner, B. S. & Pickering, J. B. (1994) *Vowel perception and production*. Oxford: Oxford University Press.

Smith, C. L. (1995) Prosodic patterns in the coordination of vowel and consonant gestures. In *Phonology and phonetic evidence: papers in laboratory phonology IV* (B. Connell & A. Arvaniti, editors), pp. 205–222. Cambridge, U.K.: Cambridge University Press.

Strange, W. (1995) Cross-language studies of speech perception: a historical review. In *Speech perception and linguistic experience* (W. Strange, editor), pp. 3–45. Baltimore, MD: York Press.

U.S. Foreign Service Institute. (1965) *Shona: basic course*. Washington, DC: Foreign Service Institute, Department of State.

van Heuven, V. J. & Dupuis, M. C. (1991) Perception of anticipatory VCV-coarticulation: effects of vowel context and accent distribution. In *Proceedings of the twelfth international congress of phonetic sciences,* Vol. 4, pp. 78–81. Aix-en-Provence, France: Université de Provence.

Vayra , M., Fowler, C. A. & Avesani, C. (1987). Word-level coarticulation and shortening in Italian and English speech, *Studi di Grammatica Italiano*, **13,** 249–269.

Whalen, D. H. (1981) Effects of vocalic formant transitions and vowel quality on the English [s]-[ʃ] boundary, *Journal of the Acoustical Society of America*, **69,** 275–282.