CROSS-LANGUAGE STUDY OF THE PERCEPTION OF COARTICULATED SPEECH

Patrice Speeter Beddor
University of Michigan, Ann Arbor, USA

ABSTRACT
Perception of coarticulated speech has been shown to be compensatory in that some of the acoustic characteristics of a target sound are attributed to coarticulatory context rather than to the target itself. But is perceptual compensation sensitive to language-specific coarticulatory patterns? This question was addressed through cross-language acoustic and perceptual study of two types of coarticulation that differ across languages, coarticulatory vowel nasalization (investigated in Thai and English) and vowel-to-vowel coarticulation (Shona and English). The perceptual differences across language groups were generally consistent with the patterns of coarticulation that emerged in the acoustic analyses. It is argued that language-specific coarticulatory structures give rise to listener expectations concerning the acoustic consequences of coarticulation and that such expectations lead, to a limited extent, to language-specific patterns of perceptual accommodation to coarticulated speech.

1. INTRODUCTION
Perceptual theories differ in their interpretation of the role of coarticulatory information in speech perception. In recent years, a controversial topic in the perception of coarticulated speech has been the phenomenon of ‘perceptual compensation’. The phenomenon itself is widely attested. Virginia Mann [9], for example, found that a /da-ga/ continuum elicited more /g/ responses from American English listeners when the stops were preceded by /l/ than when preceded by /s/. Because in natural coarticulated speech /l/ fronts a following velar consonant, Mann’s finding was compensatory in that listeners were willing to accept a wider range of stimuli as back /g/ in a (normally) fronting coarticulatory context. More generally, in compensatory responses, listeners’ judgments of a target sound appear to attribute some of the acoustic characteristics of that target to its coarticulatory context. This has been shown for consonants, vowels, and prosodic characteristics (see [6] and [14] for reviews), and for sine wave analogues of speech [8]. Perceptual compensation has also been demonstrated for various populations of listeners. For example, adult Japanese listeners, whose native language lacks an /l-/ contrast [10], young English-learning infants [6], and Japanese quail [7] have been shown to compensate for the effects of liquids on preceding stops.

The controversy over the phenomenon arises when theorists postulate the source of the compensatory context effects. That non-humans show, and that non-speech stimuli elicit, compensatory effects is taken by auditory theorists as evidence that the phenomenon is due to general auditory processes such as frequency contrast [7, 8]. Direct realists argue that the range of effects demonstrated in the literature is most consistent with the view that listeners ‘parse’ the acoustic signal along gestural lines, disentangling the properties that are due to coarticulatory overlap and assigning them to their source [5, 6].

My colleagues and I have been investigating the factors underlying compensatory effects from a cross-language perspective. The central question addressed is whether language-specific patterns of coarticulatory organization give rise to language-specific patterns of perceptual compensation. We have been conducting a series of experiments that explore two types of coarticulation: coarticulatory vowel nasalization and vowel-to-vowel coarticulation. Results indicative of a link between language-particular patterns of production and perception would argue against a purely auditory account of the perceptual effects. Although space does not permit a detailed presentation of the methods and results of these experiments, the main findings are described below.

2. COARTICULATORY VOWEL NASALIZATION
Languages differ in the extent of coarticulation — especially anticipatory — vowel nasalization [2, 3, 15, 17]. In an ongoing acoustic study in our laboratory, Chutamanee Onsuwan and I are comparing the temporal patterns of anticipatory nasalization in English and Thai. Results to date show that, in CVN(C) syllables (where C = oral consonant and N = nasal consonant), approximately 80% of the vowel was nasalized in American English speakers’ productions, while on average only 45% of the vowel was nasalized in Thai speakers’ productions of comparable syllables. Moreover, in Thai, but not English, longer vowels were proportionately less nasalized than shorter vowels. Thus anticipatory vowel nasalization in the two languages differ in both temporal extent and its constant (English) vs. variable (Thai) proportional nature.

Rena Krakow and I [11] investigated whether these language differences in patterns of coarticulatory vowel nasalization have consequences for the perception of vowel nasalization by native speakers of English and Thai. To assess perception of oral and nasal vowels in different coarticulatory contexts, we tested 16 native English-speaking listeners and 15 native Thai-speaking listeners in two perceptual paradigms, a nasality rating test and a nasality (4IAX) discrimination test. Both tests involved manipulating original [CVC] and [N[N] words (bed, bode, men, and moan) produced by a male native speaker of American English. Excising and cross-splicing techniques were used to create test stimuli in which oral and nasal vowels each occurred in an oral C-C context, a nasal N-N context, and isolation.

Consider first the basic design and results of the rating test. In this test, all possible pairings of the six stimulus types (CVC, ČVC, NVN, NVN, V, Y) within a vowel category (i.e., either /e/ or /o/) were created. For each pair, listeners decided which pair member had the more nasal vowel, or whether the vowels were equally nasal. Of particular interest were the pair types that included a nasal vowel in a nasal consonant context paired with a vowel in a non-nasal context: NNV-N, NNV-N, ČNV-CVC, and NNV-CVC. If the coarticulatory N-N context enabled listeners to compensate for nasal nasality, the vowels in NVN stimuli should
sound less nasal than the vowels in \( \tilde{V} \) and CVC, and possibly nearly equal in nasality to the vowels in V and CVC. Hence coarticulatory compensation should give rise to relatively poor rating accuracy on pairs with NVN. At the same time, we speculated that native English and Thai speakers might not have the same level of difficulty with these pairs. Perhaps Thai speakers, whose native language experience involves less extensive coarticulatory nasalization than that of English speakers, might offer less fully compensatory responses to the English stimuli. In this case, the expected decline in performance on pairs with NVN compared to other stimulus pairs might be greater for native English-speaking than for native Thai-speaking listeners.

The results for the two listener groups were in some respects more remarkable for their similarities than for their differences. First, both groups showed the same patterns of relative difficulty of the different pair types, with pairings involving NVN being judged (as predicted) the most inaccurately. Second, neither group showed “complete” compensation for coarticulatory context: performance on NVN-CVC pairs dropped to chance level, but no condition elicited systematically incorrect responses from listeners. (Note that complete compensation for vowel nasalization in a nasal context would mean that the vowel in NVN should consistently be judged as having the same nasality as the vowel in CVC or V.)

However, where the two language groups systematically differed, they differed in the direction expected on the basis of their patterns of coarticulation. Specifically, for pairs with NVN, Thai listeners were less likely than English listeners to make compensatory errors. That is, Thai listeners were less likely to judge the vowels in NVN and V or CVC as equally nasal, or to judge the vowel in NVN as less nasal than that in \( \tilde{V} \) or CVC. Figure 1 illustrates this outcome for the NVN-V and NVN-CVC pairs.

Turning to the discrimination task, the same set of six stimulus types were again used (CVC, CVC, NVN, NVN, V, \( \tilde{V} \)), but in this 4IAX design stimuli were grouped into two stimulus pairs, one whose vowels differed in nasality and one whose vowels had the same nasality (and in fact were acoustically identical). Listeners’ task was to select the pair whose vowels sounded “more different”. The trials of greatest interest here are those trials in which the vowels differing in nasality were both in coarticulatorily appropriate contexts. These trial types, all with NVN, are given in (1)-(4) of Table 1; (5)-(10) show the other test trial types. If listeners hear nasal vowels in a coarticulatory N-N context as being relatively non-nasal, then the vowels in, for example, NVN-\( \tilde{V} \) might (incorrectly) sound more different than the vowels in NVN-V. That is, listeners might base their similarity judgments more on coarticulatory appropriateness (left pairs in (1)-(4)) than on acoustic similarity (right pairs). However, if listener choices are based in part on language-specific patterns of coarticulatory vowel nasalization, English listeners might be more likely than Thai listeners to make such incorrect, contextually based choices.

As with the nasality rating results, the overall patterns in the nasality discrimination data were similar for the two language groups. For English and Thai listeners, error rates were highest on trials (1)-(4) (especially (3) and (4)), suggesting that both groups based their judgments in part on coarticulatory appropriateness. Contrary to our prediction, English listeners were, on average, no more likely than Thai listeners to offer compensatory responses. However, another cross-language difference did emerge that we would argue is also in keeping with Thai vs. English patterns of vowel nasalization: Thai listeners had greater difficulty discriminating the nasality of nasal vowels than that of oral vowels, independent of context; English listeners did not show this effect. (That is, Thai listeners performed more poorly on the odd-numbered trials in Table 1, where Pair B has nasal vowels, than on the corresponding even-numbered trials, where Pair B has oral vowels.) The context-independent difficulty of vowel nasalization for Thai listeners is consistent with the acoustic evidence that Thai listeners have less native language experience with vowel nasalization than do English listeners. Moreover, arguably, the two languages may also differ in the phonological status of vowel nasalization, with nasalization
being ‘phonologized’ in English [14, 17] but not Thai (recall, for example, the constant proportional nature of nasalization in English but not Thai).

To summarize, English and Thai listeners’ responses to tests of nasality rating and nasality discrimination show that listeners perceptually compensate — albeit partially rather than fully — for vowel nasalization under coarticulatorily appropriate conditions. The response patterns of the two language groups were similar in many respects, but the cross-language perceptual differences that were found appear to be linked to the differences in coarticulatory structure in the two languages.

3. VOWEL-TO-VOWEL COARTICULATION

Languages also differ in their patterns of vowel-to-vowel coarticulation [4, 11, 12, 13, 16]. James Harnsberger, Stephanie Lindemann and I are acoustically investigating vowel-to-vowel coarticulation in the productions of several speakers each of Shona and American English. Stimuli are disyllables with initial stress, CV.CV2, and trisyllables with medial stress, CV.CV2.CV3, where V = /i a e o u/ and C = /b/ for English and /p/ for Shona (Shona /p/ is relatively unaspirated; the voiced bilabial stop in Shona is implosive). Formant frequencies are measured at vowel onset, midpoint, and offset, and systematic formant differences that are conditioned by a flanking vowel are taken as acoustic evidence of vowel-to-vowel coarticulatory effects. Two differences in coarticulatory effects in the Shona and English data are particularly noteworthy. First, there is a tendency for the vowel-to-vowel effects in Shona to be more extensive for anticipatory coarticulation (V1 influenced by V2; trisyllables: also V2 influenced by V3) than for carryover coarticulation (V2 influenced by V1; trisyllables: also V3 influenced by V2). In contrast, this is not a general pattern for English speakers, some of whom show greater carryover than anticipatory coarticulation. Second, the coarticulatory effects on Shona vowels are nearly the same for stressed and unstressed vowels (e.g., for trisyllables, anticipatory effects on stressed V2 and on unstressed V3 are of comparable magnitude). But English speakers’ stressed and unstressed vowels are asymmetric in how they are affected by flanking vowels, with stressed vowels being more resistant to coarticulatory effects.

To examine listeners’ sensitivity to these anticipatory and carryover effects, we tested 16 native English-speaking and (to date) 9 native Shona-speaking listeners on an identification task. (Listeners are also being tested on a 4IAX discrimination paradigm similar in basic design to that described above for investigating coarticulatory nasalization; these results are not reported here.) The stimuli for the identification test were four synthetic /pV1V2p/ continua generated on the Klatt terminal-analog synthesizer (KLSYN88), with endpoints being composite representations of the productions of two female speakers of each of the languages. The endpoints of the four continua are given in Table 2. In all cases, the variable portion of the continuum was a 10-step vowel series from /a/ to /e/, varying in equal-sized F2 and F3 increments and F1 decrements across the series. The continua on the left were designed to test the anticipatory effects of V2 on the identification of V1: V2 was either /i/ or /a/ and was held constant within a continuum; V1 varied from /a/ to /e/. Similarly, the last two continua in Table 2 tested for carryover effects of V1 (always /i/ or /a/ within a continuum) on V2 (varying from /a/ to /e/). Listeners were asked to identify the vowel of either the first syllable (anticipatory continua) or the second syllable (carryover continua) as /a/ or /e/.

What is the predicted pattern of responses? Suppose a continuum vowel in isolation is ambiguous between /a/ and /e/. When that vowel is placed in the context of another vowel, compensation predicts that listeners should attribute some portion of the vowel quality of the ambiguous stimulus to the flanking vowel. Specifically, when the flanking vowel is the front vowel /i/, compensation should yield more back vowel — i.e., /a/ — responses; when the flanking vowel is /a/, compensation should yield more front /e/ responses. In addition, a compensation account sensitive to the language-specific coarticulatory patterns described above predicts that Shona listeners should compensate more than English listeners on the anticipatory continua (with a varying stressed vowel). Also, Shona listeners should compensate more on the anticipatory than the carryover continua. Because Shona and English exhibit fairly substantial carryover coarticulation, both language groups are predicted to show perceptual carryover effects.

Figure 2 gives the percent /a/ responses for the anticipatory continua for Shona and English listeners. As expected, listeners

<table>
<thead>
<tr>
<th>Anticipatory continua</th>
<th>Carryover continua</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stim. 1</td>
<td>Stim. 10</td>
</tr>
<tr>
<td>papi</td>
<td>pepi</td>
</tr>
<tr>
<td>papa</td>
<td>pepa</td>
</tr>
</tbody>
</table>

Table 2. Stimulus endpoints for the four identification continua. Symbols in bold are the sounds that vary across each continuum.

Figure 2. Anticipatory effects: Percent /a/ responses of Shona and English listeners to an /a-e/ continuum followed by a back vowel (diamonds) vs. a front vowel (squares). Dashed vertical lines show the location of the 50% identification crossover.
gave more /a/ responses when the following vowel was /i/ than when it was /a/. The cross-language perceptual difference predicted on the basis of Shona and English coarticulatory patterns was also found: the difference in the 50% category boundaries for the /p_ip/ and the /p_pa/ series was greater for the Shona than the English listeners’ responses, indicating more extensive perceptual compensation for that language group. Moreover, also as predicted on the basis of coarticulatory patterns, Shona listeners showed larger category shifts due to anticipatory than to carryover coarticulation. This is seen by comparing the Shona anticipatory boundary shift in Figure 2a with the Shona carryover boundary shift in Figure 3. One unexpected result is that English listeners did not show a consistent boundary shift for the carryover continua (not shown here); instead, their responses to the /pap_/ series indicated considerable perceptual ambiguity. (In general, the English identification functions were not as smooth as the Shona. A contributing factor was that, for the English listeners, the /a-e/ continuum passed through phonemic /e/; Shona’s five-vowel system did not give rise to this ambiguity for the Shona listeners.)

4. CONCLUSION
In these perceptual studies of coarticulatory nasalization and vowel-to-vowel coarticulation, not all expected language differences were obtained, and the findings show striking cross-language similarities in patterns of perceptual compensation. This is to be expected, especially for the types of coarticulation investigated here. Patterns of coarticulatory nasalization and vowel-to-vowel coarticulation are influenced by a range of segmental and prosodic factors (see also [1]). In addition, not all speakers of a language show particular patterns to the same extent. Listeners must be able to deal with both types of variability in coarticulation.

At the same time, systematic language differences emerged in these perceptual studies, and these differences support a theory of perceptual compensation that is sensitive to language-specific coarticulatory patterns. Speakers have, at some level, knowledge of the coarticulatory organization of their language. The theory of compensation proposed here is that language-specific patterns of coarticulation give rise to listener expectations concerning the acoustic consequences of coarticulation; such expectations in turn lead to language-specific patterns of perceptual accommodation to coarticulated speech.

ACKNOWLEDGMENTS
The research reported here was supported in part by NSF Grant SBR 9319597. Collaborators on the coarticulatory nasalization experiments are Rena Krakow and Chutamane Osumwan; collaborators on the vowel-to-vowel coarticulation experiments are James Harnsberger and Stephanie Lindemann. Sylvia Mupepi provided very helpful advice on the Shona portions of the experiments. Brad Rakerd generously made available facilities at Michigan State University for testing Shona participants.

REFERENCES