CONTEXT AND SPEAKER EFFECTS IN THE PERCEPTUAL ASSIMILATION OF GERMAN VOWELS BY AMERICAN LISTENERS

Winifred Strange[1], Ocke-Schwen Bohn[2], Sonja A. Trent[1], Melissa C. McNair[1], & Katherine C. Bielec[1]

1. University of South Florida, 4202 E. Fowler Avenue, Tampa, FL, 33620 USA
2. University of Kiel, Olshausenstrasse 40, D-24098 Kiel, Germany

ABSTRACT

To directly assess the influence of consonantal context and speaker differences on cross-language perceptual similarity of vowels, speakers of American English (AE) were asked to categorize and rate the goodness of fit of North German (NG) vowels to native categories. Four speakers produced the 14 NG monophthongs in 5 CVC contexts in a carrier sentence. Twelve listeners were presented each speaker's utterances with vowels and consonantal contexts randomly sequenced. Overall perceptual assimilation patterns showed large variations in the perceived similarity of NG vowels even for those vowels which are considered phonetically similar across languages. The front rounded NG vowels, which do not occur as distinctive phonemes in AE, were almost always assimilated to back rounded AE vowels. Significant context and speaker effects were shown for most of the NG vowels. This suggests that context-free descriptions of cross-language phonetic similarity of vowels will not be adequate in predicting relative perceptual difficulty in learning to differentiate non-native vowels. These results also have implications for theories about the nature of the representation of native-language phonetic categories.

1. INTRODUCTION

Current theories of second-language (L2) speech learning assume that adult beginning learners perceptually assimilate non-native consonants and vowels to native-language phonetic categories on the basis of their "perceived phonetic similarity" (c.f. Best, 1995; Flege, 1992). Perceived similarity is often inferred on the basis of phonetic descriptions of the two languages, which presumably capture something about the articulatory/acoustic similarities of the segments in the languages of interest. For vowels, native (L1) and non-native (L2) categories are often compared with respect to their relative positions in the traditional articulatory "vowel space" or in the related F1/F2 acoustic target space. Few studies have *directly* assessed cross-language perceptual similarity of vowels (but see Flege, Munro, & Fox, 1994).

The study reported here is part of a larger project in which cross-language perceptual assimilation patterns are directly assessed for Japanese (J), American English (AE), and North German (NG) vowels. Of specific interest is the extent to which perceived phonetic similarity is influenced by the consonantal context in which the vowels are produced and presented, and by the individual speakers who produce the stimuli. Data on AE listeners' perceptual assimilation of NG vowels produced by four speakers in five CVC contexts are presented here.

2. METHOD

Four young adult male speakers of NG (Kiel dialect) each produced the 14 monophthongal vowels [iː, ɪ, eː, ɛ, a, ɑː, yː, ʏ, øː, œ, uː, ʊ, oː, ɔ] in five CVC contexts: /bVp, bVt, dVt, gVt, gVk/. The syllables were produced in the carrier sentence "Ich habe --- gesagt" at a relatively rapid speaking rate. Multiple repetitions of each utterance were recorded with a DAT recorder (in Germany) and then digitally transferred to computer for presentation in listening tests. For each speaker, two good instances (as judged by a phonetically trained native speaker of German) of each vowel in each context were selected for listening tests (14 X 2 X 5 = 140 stimuli per speaker). Listening tests consisted of 4 randomized blocks of the 140 stimuli, for a total of 560 test trials per speaker. Each trial consisted of two presentations of each stimulus.

A mixed design was employed to assess perceptual assimilation patterns by AE listeners, with consonantal context a repeated-measures variable and speakers a between-subjects variable. Each listener heard all tokens produced by one speaker, with the consonantal context and vowels randomly sequenced. Forty-eight undergraduate and graduate students (44 females, 4 males) from the USF Communication Sciences & Disorders Department who had completed a one-semester course in basic phonetics and IPA transcription were assigned to 4 groups of 12 listeners each. All testing was done individually, using an interactive computer program with stimuli presented over earphones. Subjects first completed a 44-item response training task in which they identified AE vowels spoken in CVC syllables embedded in a carrier sentence using 11 response categories, displayed on the screen with both /hVd/ key words (heed, hid...) and IPA symbols. Subjects were required to perform with fewer than 4 errors (except for /ɑ-ɔ/ confusions) in this task to remain in the experiment.

In the perceptual assimilation task, listeners first selected the response alternative "most similar" to the NG vowel presented; then after the second presentation of the same stimulus, they rated its goodness-of-fit to the chosen response on a 7-point scale (7 = best fit). Before each test, listeners heard 28 of the speaker's productions without responding, then 14 practice trials in which they categorized the target syllable and rated its goodness of fit to the selected AE response alternative. They then completed 4 test blocks of 140 trials each. Familiarization and testing were completed in two sessions held on separate days. Testing was self-paced, with sessions lasting about 60 to 90 minutes each.

3. RESULTS

NG	AE	Resp. 1 (%)	Median Rating	AE	Resp. 2 (%)	Median Rating
i:	i:	99	7	eɪ	<1	5
ɪ	ɪ	84	6	ɛ	9	5
e:	eɪ	64	6	i:	31	6
ɛ	ɛ	92	6	eɪ	6	5
u:	u:	88	5	ʊ	8	4
ʊ	ʊ	45	4	oʊ	32	4
o:	oʊ	69	5	u:	21	5
ɔ	ɔ:	53	4	ɑ:	28	5
ɑ:	ɑ:	75	5	ɔ:	24	5
a	ɑ:	56	5	ʌ	30	5
y:	u:	93	5	ʊ	6	3
ʏ	ʊ	74	4	ʌ	13	2
ø:	u:	43	5	ʊ	33	4
œ	ʌ	74	4	ʊ	19	4

Table 1: Overall assimilation patterns for modal and second most frequently selected response categories and corresponding median goodness-of-fit ratings summed across speakers and contexts.

First, overall perceptual assimilation patterns, averaged over speaker/listener groups and consonantal contexts were inspected, as presented in Table 1. German vowels are indicated in IPA symbols in the first column: front unrounded vowels (rows 1-4), back rounded vowels (rows 5-8), back and central unrounded vowels (rows 9-10), and front rounded vowels (rows 11-14). Columns 2-4 give the overall percentage of trials (summed over speakers and listeners) the most frequent response alternative (Response 1) was selected and the median goodness rating on those trials. Columns 5-7 give the percentages and median ratings for the second most frequently selected AE response alternative (Response 2). (For all NG vowels, more than two AE vowel categories were judged as most similar on some trials by some listeners in some contexts; thus, the percentages in columns 3 and 6 do not sum to 100%.)

As the modal response data indicate, the front unrounded NG vowels were perceived as quite similar to AE vowels (median rating = 6 or 7) and were quite consistently categorized as the equivalent AE vowels, except for the NG vowel [e:], which was categorized as either [i:] or, more rarely, [ɛ] in some context/speaker conditions (see below). In contrast, modal response data indicate a much less consistent pattern of assimilation of NG back rounded and unrounded vowels to AE categories, even for those vowels [u:, ʊ, ɑ:] which are typically described as phonetically equivalent across languages. Median ratings ranged from 4 to 5 and modal categorization rates ranged from 88% for [u:] to only 45% for [ʊ]. As the Response 2 data suggest, AE listeners were inconsistent in their use of the AE categories [ɑ:, ɔ:] in the perceptual similarity judgments of the NG [ɑ:, ɔ]. This was expected since AE listeners often confuse these AE vowels in identification tests (c.f. Hillenbrand, et al., 1995). When [ɑ:-ɔ:] response alternatives were combined for judgments of these NG vowels, consistency in assimilation exceeded 90%. Thus, for the 9 NG vowels [i:, ɪ, e:, ɛ, u:, ʊ, o:, ɔ] which are typically described as identical or very similar to AE vowels with respect to their position in articulatory/acoustic vowel space (i.e., in vowel "quality"), *perceived* similarity ranged from identical for [i:] to quite dissimilar for [ʊ].

The data for the front rounded vowels (rows 11-14), which are not distinctive in AE, showed a similar range in the consistency with which they were assimilated to AE vowels. All 4 vowels were assimilated to back rounded AE vowels except for [œ] in a small number of cases. Again, median ratings (4-5) suggest that these vowels were not perceived as highly similar to any AE vowel categories. NG high front [y:] was nevertheless categorized very consistently as most similar to [u:], while the mid front vowel [ø:] was a very poor fit to any one AE vowel category, with modal responses to AE [u:] reaching only 43%.

The overall pattern of responses shown in Table 1 indicate that two or more NG vowels were assimilated to each AE vowel category except for AE [ɪ] to which only NG [ɪ] was assimilated and [æ] to which no NG vowels were perceived as similar. One would predict then, that AE speakers would have difficulty differentiating many NG vowels, including even some of those which are transcribed as identical to AE vowels. Although patterns of assimilation were, in general, predictable on the basis of proximity in "vowel space," overall *consistency* in the choice of any particular AE category as most similar to each NG vowel was not well predicted on the basis of cross-language phonetic equivalence as captured by IPA transcription. Finally, the data in Table 1 suggest that, except for the NG vowels [ɔ, a], short (lax) NG vowels were almost always assimilated to short AE vowels, while NG long (tense) vowels were assimilated to AE long vowels.

Turning to the question of the influence of consonantal context on perceptual assimilation patterns, Table 2 presents modal response data for each of the NG vowels, ordered in terms of the overall consistency in the choice of the modal response alternative (right most column). Cell entries show the proportion of times the modal response was selected in each context condition (expressed as percentages of opportunities) summed over speaker groups. The range across contexts in percentage of modal responses varied from only 1% for [i:] to 47% for [ɪ]. Notice that large context effects on perceptual assimilation patterns were observed for NG vowels that, in general, were judged as quite similar to AE vowels [ɛ, ɪ] as well as those which were judged as quite dissimilar to AE vowels [a, ʊ]. Although not shown in Table 2, there were also differences as a function of consonantal context in Response 2 patterns for several vowels.

NG	AE	[b-p]	[b-t]	[d-t]	[g-t]	[g-k]	Overall
i:	i:	99	99	99	98	99	99
ɑ:	ɑ:,ɔ:	99	98	97	99	99	99
y:	u:	93	95	96	91	90	93
*ɛ	ɛ	98	96	98	97	70	92
u:	u:	90	89	92	85	82	88
*ɪ	ɪ	99	98	94	77	52	84
ɔ	ɑ:,ɔ:	74	74	84	88	85	81
ʏ	ʊ	68	73	73	83	73	74
œ	ʌ	77	69	75	78	72	74
o:	oʊ	65	70	72	69	68	69
*a	ɑ:,ɔ:	79	43	64	70	60	66
e:	eɪ	62	68	60	71	60	64
*ʊ	ʊ	39	59	36	61	32	45
ø:	u:	47	50	42	36	41	43

Table 2: Modal responses (in percentage of opportunities) as a function of consonantal context. Asterixes show vowels which produced the largest context effects.

Statistical analyses of context and speaker effects on assimilation patterns for each NG vowel were accomplished by distributing listeners' responses into 3 categories (modal response, second most frequent response, and all other responses) and entering those frequencies into a 5 (context) x 3 (responses) x 4 (speakers) matrix. To assess these frequency data, log-linear models which included contexts, speakers, listeners (within-speakers), and speaker-context interactions as factors in various combinations were then fit to the frequency data. The simplest model (least factors) with the best fit to the data was selected to determine significant effects ($p<0.05$). By these analyses, *all but three* NG vowels [i:, u:, y:] yielded significant context effects.

As shown in Table 2, the front unrounded NG vowels [ɪ, ɛ] were assimilated less consistently in one or both initial velar consonant contexts. Other contextual effects were not as systematic. There was no clear evidence that the front rounded vowels were assimilated better to back rounded AE categories in alveolar contexts, as would be predicted from coarticulatory patterns of AE (i.e. AE back rounded vowels are most fronted in alveolar contexts).

With respect to speaker effects, statistical analyses indicated that differences in perceptual assimilation patterns across speaker/listener groups were significant for 9 NG vowels [e:, a, ɑ:, ʏ, ø:, œ, ʊ, o:, ɔ]. This was somewhat surprising given all 4 speakers were drawn from a homogeneous dialect group. Extreme differences in assimilation patterns for [e:] and [o:] as produced by one speaker were due to his vowels being assimilated almost entirely to [i:] and [u:], respectively. Assimilation patterns for this speaker's productions of [ʏ,œ] were also atypical of the patterns shown for the other three speakers' productions, again indicating that these vowels were perceived as higher (more like [u:, ʊ]) than for the other speakers. Only 3 vowels [a, ø:, ʊ] produced significant speaker by context interactions. Finally, significant (within-speakers) listener differences were found for 9 of the 14 vowels, suggesting that perceptual assimilation patterns varied with the idiolect of the AE listener.

4. DISCUSSION

The results of this study strongly suggest that perceptual assimilation of non-native vowels to native-language phonetic categories is not well predicted on the basis of context-free descriptions of phonetic similarity. Direct assessments of perceptual assimilation patterns revealed that some NG vowels considered phonetically identical or very similar to AE vowels were nevertheless poorly assimilated to native AE categories. Furthermore, there were significant differences in perceptual similarity as a function of both individual speakers and the consonantal contexts in which the vowels were produced and presented. In the current study, speaker effects could not be considered independent of listener effects because of a confounding of speaker and listener groups. Future studies should be conducted using a repeated measures design on both variables.

A question remains about the origin of the consonantal context effects. That is, did the NG vowels differ across contexts in perceived similarity to AE vowels because of differences in their articulatory/acoustic realization, or did they differ because of AE listeners' expectations about coarticulatory patterns based on AE phonological rules, or both? We are now in the process of a detailed acoustical analysis of the stimulus corpus used in this study which will allow us to partial out speaker and listener contributions. Our previous research on the perceptual assimilation of NG vowels by AE listeners (Strange et al. 1993; Trent, et al. 1995; Fitzgerald et al. 1995) suggests, however, that at least some of the context effects are due to listeners' expectations, rather than to acoustic/phonetic differences in the realization of the vowels by NG speakers.

These results have both theoretical and practical implications for second-language speech learning. They suggest that knowledge about L1 contextual variation (i.e., allophonic rules of L1) influences the perceived similarity of L1 and L2 phonetic segments. Thus, predictions of relative perceptual difficulty with L2 segments *cannot* be based on context-free representations of native-language and non-native phonetic categories. Rather, predictions must take into account the influence of listeners' imposition of *language-specific* expectations about coarticulatory patterns on perceptual assimilation of non-native segments to native categories. This will limit the generalizability of cross-language studies which examine perception of segments produced and/or presented in only one context or in isolation. From a practical point of view, data such as those presented here can be used to structure corpora for perceptual training studies. It should be possible to sequence training materials in which the to-be-learned segments or contrasts are initially presented in easier contexts, then in more difficulty contexts until generalized learning of the categories is achieved. This should optimize performance and help to maintain motivation in perceptual training tasks.

5. REFERENCES

Best, C.T. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.) **Speech perception and linguistic experience: Issues in cross-language research**. Timonium, MD: York Press.

Fitzgerald, B.H., Trent, S.A., Stack, J.W., Ling, X., & Strange, W. (1995). Acoustic specification of North German vowels produced in citation and sentence contexts. **Journal of the Acoustical Society of America, 97**, 3419 (abstract).

Flege, J. E. (1992). Speech learning in a second language. In C. Ferguson, et al. (eds.) **Phonological development: Models, research and application**. Timonium, MD: York Press.

Flege, J.E., Munro, J.J. & Fox, R.A. (1994). Auditory and categorical effects on cross-language vowel perception. **Journal of the Acoustical Society of America, 95**, 3623-3641.

Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. **Journal of the Acoustical Society of America, 97,** 3099-3111.

Strange, W., Trent, S.A., Stack, J.W., Ling, X. & Rodriguez, A.I. (1993). Consonant context affects perceived similarity of North German and American English vowels. **Journal of the Acoustical Society of America, 94**, 1866 (abstract).

Trent, S.A., Fitzgerald, B.H., Crouse, S.M., & Strange, W. (1995) Perceptual assimilation of North German vowels to American English categories. **Journal of the Acoustical Society of America**, **97**, 3419 (abstract).

[Supported by NIH-NIDCD]