

Acoustic variability within and across German, French, and American English vowels: Phonetic context effects

Winifred Strange,^{a)} Andrea Weber,^{b)} Erika S. Levy,^{c)} Valeriy Shafiro,^{d)} Miwako Hisagi,^{e)} and Kanae Nishi^{f)}

Ph.D. Program in Speech and Hearing Sciences, Graduate School and University Center, The City University of New York, 365 Fifth Avenue, New York, New York 10016-4309

(Received 30 January 2007; revised 14 May 2007; accepted 25 May 2007)

Cross-language perception studies report influences of speech style and consonantal context on perceived similarity and discrimination of non-native vowels by inexperienced and experienced listeners. Detailed acoustic comparisons of *distributions* of vowels produced by native speakers of North German (NG), Parisian French (PF) and New York English (AE) in citation (di)syllables and in sentences (surrounded by labial and alveolar stops) are reported here. Results of within- and cross-language discriminant analyses reveal striking dissimilarities across languages in the spectral/temporal variation of coarticulated vowels. As expected, vocalic duration was most important in differentiating NG vowels; it did not contribute to PF vowel classification. Spectrally, NG long vowels showed little coarticulatory change, but back/low short vowels were fronted/raised in alveolar context. PF vowels showed greater coarticulatory effects overall; back and front rounded vowels were fronted, low and mid-low vowels were raised in both sentence contexts. AE mid to high back vowels were extremely fronted in alveolar contexts, with little change in mid-low and low long vowels. Cross-language discriminant analyses revealed varying patterns of spectral (dis)similarity across speech styles and consonantal contexts that could, in part, account for AE listeners' perception of German and French front rounded vowels, and "similar" mid-high to mid-low vowels. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2749716]

PACS number(s): 43.70.Kv, 43.70.Fq, 43.71.Hw, 43.71.Es [ARB]

Pages: 1111–1129

I. INTRODUCTION

There has been ongoing interest over the last twenty years in investigating the perception of non-native vowels by naïve listeners (listeners with no experience with the target language) and second-language (L2) learners with varying amounts of L2 experience. For instance, several studies have investigated the perception by American English (AE) listeners of front, rounded vowels that contrast with both front, unrounded vowels and back, rounded vowels in German and French (AE contrasts only front, unrounded and back, rounded vowels). An early cross-language categorical perception study with synthetically generated isolated vowels suggested that these non-native vowel contrasts were not problematic perceptually (Stevens *et al.*, 1969). However, more recent studies with talker-generated vowels produced and presented in one or more phonetic contexts have shown

significant and persistent difficulties for both naïve listeners and L2 learners in perceptual differentiation of some front, rounded vs back, rounded vowel contrasts, as well as contrasts among front, rounded vowels (Best *et al.*, 2003; Flege, 1987; Gottfried, 1984; Levy and Strange, in press; Polka, 1995). These studies have also reported perceptual difficulties even for L2 vowel contrasts that also occur in the listeners' native language (L1), but are phonetically realized differently across languages. For instance, Gottfried (1984) reported that naïve American listeners had trouble discriminating French [i/e] and [e/ε].

In attempting to predict and explain these cross-language perceptual difficulties, it becomes clear that an abstract phonological description of vowels fails to capture significant cross-language differences in these phonetic categories. Rather, to characterize cross-language (dis)similarities as they affect perceptual performance, the vowels of each language must be described with respect to their actual articulatory (and resulting acoustic) realization in various phonetic contexts. This paper reports on a detailed acoustic analysis of vowel corpora produced by native speakers of North German, Parisian French, and New York English. Distributions of vowels produced in a variety of phonetic contexts were analyzed to document similarities and differences in their systematic acoustic variation in the three languages. Using linear discriminant analysis, the contribution of spectral and temporal parameters to the acoustic differentiation of vowels within each language was assessed. In addition, *cross-language* discriminant analyses were used to evaluate cross-language acoustic (dis)similarities and how they

^{a)}Electronic mail: strangepin@aol.com

^{b)}Present address: Saarland University, Building C7.1, Rm. 1.1, 66041 Saarbrücken, Germany.

^{c)}Present address: Department of Biobehavioral Sciences, Program in Speech and Language Pathology, Box 180, Teachers College, Columbia University, 525 W. 120th St., New York, NY 10027.

^{d)}Present address: Communication Disorders and Sciences, Rush University Medical Center, 203 Senn 1653 W. Congress Parkway, Chicago, Illinois 60612.

^{e)}Present address: Research Laboratory of Electronics and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Rm. 36–581, 50 Vassar St., Cambridge, MA 02139–4307.

^{f)}Present address: Boys Town National Research Hospital, 555 N. 30th St., Omaha, NE 68131

changed across different speech styles and phonetic contexts. The productions of male and female speakers were analyzed separately to mitigate issues of vocal tract normalization and to provide a replication of cross-language patterns of acoustic (dis)similarities.

The usefulness to speech perception/production research of such detailed acoustic descriptions covering complete phoneme inventories of languages needs no special advocacy. On the basis of such data, it is possible to provide insights into the patterns of perception, not only by native listeners of the languages but also by non-native listeners. Our present aim was to provide such essential acoustic descriptions for North German, Parisian French, and New York English. Previous research reported difficulties by AE listeners in discriminating German (Polka, 1995) and French (Gottfried, 1984; Levy and Strange, in press) vowels; thus, an acoustic analysis of the vowel inventories allowed us to determine the extent to which perceptual difficulties might be explained by cross-language patterns of acoustic (dis)similarity (Flege, 1995).

In most previous studies of cross-language vowel perception, stimuli were generated by recording (di)syllables spoken in lists (citation materials) in which the surrounding consonants were fixed or absent (e.g., Gottfried, 1984; Polka, 1995). A few studies have used materials in which the target (di)syllables have been produced and presented in short carrier sentences at a rate more closely approximating continuous speech (e.g., Levy and Strange, in press; Strange *et al.*, 2005). Citation productions can be considered a form of “clear speech” in which the vowels may be produced with more acoustic/articulatory distinctiveness than in continuous speech utterances (see Stack *et al.*, 2006). Following the classic study by Peterson and Barney (1952) of American vowels, the hVd context is often used. In English (and German), the initial /h/ can be considered a voiceless onset of the vowel; the tongue/lips/jaw assume the position for the vowel from the beginning of the utterance. (French does not have an initial /h/, but all vowels can occur in word/syllable-initial position.) The following /d/ shows some coarticulatory influence on the spectral characteristics of the vowel toward the end of slowly-produced syllables, but the midpoint formant values are often taken to be the “canonical target” specification of the vowel, independent of any coarticulatory influence from the surrounding consonants.

Studies of AE vowels (Hillenbrand, Clark, and Nearey, 2001; Stevens and House, 1963) have shown that midpoint formant frequencies of vowels produced in CVC syllables vary significantly with the place-of-articulation of the surrounding consonants. Other studies have documented the effects of consonantal context on mid-syllable formant values for vowels of other languages (see Bohn, 2004 for North Frisian; Steinlen, 2005, for North German, Danish, and British English; Strange *et al.*, 2005, for North German). In the present study, we recorded citation (di)syllables of the form (h)Vb(ə) to establish the target values for vowels in each language. The following /b/, because it does not require articulation of the tongue, was thought to be a better choice than the coronal /d/. The same speakers then produced vowels in multisyllabic nonsense words /cvCVC(ə)/ medially

within short carrier sentences; the immediately surrounding consonants included both labial and alveolar stops /bVb, bVp, dVd, dVt/ because previous studies had determined that these places of articulation encompassed the greatest variation in vowel formants in AE and North German.

A. Vowel inventories for North German, Parisian French, and American English

1. North German vowels

German is described as having 15 distinctive monophthongal vowels: front, unrounded [i:, ɪ, e:, ɛ, ɛ:]; back, rounded [u:, ʊ, o:, ɔ]; front, rounded [y:, ʏ, ø:, œ]; and low (back, unrounded) [ɑ:, a]. However, in earlier work with North German (NG), it was determined that [ɛ:] was only produced in hypercorrect speech; thus, it was excluded from the present analysis (Steinlen, 2005; Strange and Bohn, 1998). The remaining 14 vowels can be grouped into seven pairs of long (tense, close) vs short (lax, open) vowels. However, in NG, these long/short pairs differ significantly in tongue height as well as in vocalic duration, except for [ɑ:/a] (Steinlen, 2005; Strange *et al.*, 2004a; 2005). Thus, in terms of phonological descriptions, NG vowels can be said to contrast in tongue height, tongue position, lip rounding, and vowel length/tenseness. Phonetically, tongue height and vowel length redundantly distinguish six pairs: [i:/ɪ, e:/ɛ, u:/ʊ, o:/ɔ, y:/ʏ, ø:/œ], while [ɑ:/a] are distinguished only by length.

2. Parisian French vowels

French is described as having ten distinctive oral vowels: front, unrounded /i, e, ɛ/; back, rounded /u, o, ɔ/; front, rounded /y, ø, œ/; and the low central vowel /a/.¹ The mid and mid-low, front, rounded vowels [ø, œ] contrast in only a few French words, and have merged in present-day Parisian French (PF), with some allophonic variation as a function of phonetic context. Thus, in this study, a single symbol [ø] is used to indicate the mid, front, rounded vowel in all contexts, in contrast with the high [y]. The three front, unrounded vowels and the three back, rounded vowels differ in tongue height (high, mid, mid-low, respectively). Vowel length is arguably not phonetically functional in PF vowels. Only the pair [o:, ɔ] is said to vary in length systematically, and then only in closed syllables, which make up a small proportion of lexical items in French (Gottfried and Beddor, 1988). Furthermore, Gottfried and Beddor reported that native French listeners did not use vocalic duration to distinguish these vowels perceptually, although they did vary systematically in production (but see Miller and Grosjean, 1997; Miller *et al.*, 2000, for a different perceptual pattern in Swiss French). Thus, unlike German, in which vowel duration appears to have both a phonological and a phonetic function, French vowels appear not to be distinguished phonologically or perceptually by vocalic duration differences.

3. New York English vowels

AE is described as having 11 nonrhymic distinctive vowels: front, unrounded [i:, ɪ, e:, ɛ, æ:]; back, rounded [u:, ʊ, o:, ɔ:]; and the mid-low and low (back, unrounded) vowels

[ʌ, ɑ:]. It does not have distinctive front, rounded vowels, although in coronal consonantal contexts, there is allophonic “fronting” of back, rounded vowels (Hillenbrand *et al.*, 2001; Strange *et al.*, 2005). Front and back vowels vary in tongue height (high, mid-high, mid, mid-low, low) in similar ways to NG. While vowel length is not considered phonologically distinctive, AE vowels vary systematically in intrinsic duration, as indicated by the [ː] above (Peterson and Lehiste, 1960). Note that, while front vowels alternate in duration as a function of tongue height, the back vowels [oː, ɔː, ɑː] do not. In some dialects of AE (including Canadian English), [ɔː, ɑː] have merged into a single mid-low, back, slightly rounded [ɒː]. However, in New York English, these vowels are distinct, with [ɔː] rounded and higher than unrounded, low [ɑː]. The mid, front and back vowels are described as diphthongized [e¹, o¹] especially in open syllables, and several other vowels have been shown to have perceptually-relevant vowel-intrinsic spectral change (VISC) throughout the vocalic nucleus (Nearey and Assmann, 1986). These changes in formant structure within the vocalic nucleus are most apparent in slowly-articulated citation utterances, and vary across dialects of AE.

B. Language-independent and language-dependent coarticulatory patterns

From these traditional phonological/phonetic descriptions of vowels in the three languages, several speculations can be made about their cross-language (dis)similarities. The front, rounded vowels of German and French can be considered non-native for AE listeners. However, NG front, rounded vowels are sometimes described as more central than PF front, rounded vowels (Steinlen, 2005) and AE back, rounded vowels that are fronted in coronal contexts are phonetically quite similar to NG front, rounded vowels (Strange *et al.*, 2005). To our knowledge, no direct comparisons of the acoustic structure of these American, French, and German vowels, produced in a variety of consonantal contexts, have been reported (but see Hay *et al.*, 2006). Other vowels, which are often transcribed as the “same” in all three languages, may nevertheless differ in their relationships to other vowels in the language and in their dynamic spectral structure. Mid, front and back vowels /e/ and /o/ are monophthongal in German and French, while they tend to be diphthongized in AE. Furthermore, previous comparisons of NG and AE vowels have shown that these mid vowels differ in relative height, with the NG mid vowels higher (i.e., closer to NG high vowels) than for AE vowels. The low vowels in German [ɑː, a] and French [a] may differ in both tongue height and backness from AE low and mid-low [ɑː, ʌ]. Finally, some vowels vary across languages in their temporal structure; [ɔ] is short in German and French, while AE [ɔː] is long. Furthermore, the effect of tongue height on relative durations of vowels is also greater for NG than for AE vowels (Strange and Bohn, 1998; Strange *et al.*, 2004a, 2005). Thus, a comparison of the vowels of these three languages, produced in the same phonetic contexts, was necessary to establish perceptually-relevant phonetic (dis)similarities across languages.

Of specific interest here was how *contextual variability* in the spectral and temporal structure of vowels differed across languages. Classical theories of vowel coarticulation (phonetic vowel reduction) posit the concept of “target undershoot” to predict/explain the variation in mid-syllable formant values in different CVC syllables produced at different speaking rates (Lindblom, 1963). According to this simple dynamic model, as vowels are coarticulated with consonants, the specified target position of the vowel in vowel space remains invariant, but the temporal overlapping of commands to the articulators for consonant and vowel gestures leads to a failure of the articulators to reach the canonical target before the command for the following consonant takes effect. The amount of articulatory (and resulting acoustic) undershoot varies as a function of the difference between consonant loci and vowel target (locus/target distance) and increases at faster speaking rates. If these dynamic consonant/vowel interactions were the only influence on vowel production, variation in the spectral structure of vowels in the same consonantal contexts should be very similar across languages. However, Moon and Lindblom (1994) showed that the amount of target undershoot in AE vowels differed with speech style (clear vs citation) and other studies of AE (Fourakis, 1991; Stack *et al.*, 2006) and Dutch (van Son and Pols, 1992) have failed to show systematic formant changes in vowels over different speaking rates in read continuous speech utterances. With respect to the influence of phonetic context on the spectral structure of vowels, Steinlen (2005) reported that variability of both first (F1) and second formant (F2) values for vowels in the same consonantal contexts differed across languages as a function of the size of the vowel inventory, with Danish (20 vowels; 10 long/short pairs) showing the least variability and British English (11 vowels) showing the most variability. Bohn (2004) has proposed that the amount of contextual shrinking of the vowel space is inversely related to inventory size, although the relationship may not be linear.

Given these differences within and across languages, we hypothesized that contextual variability in the spectral structure of vowels is, to a significant extent, actively controlled by speakers. That is, much of the coarticulatory variation in vowels is due to learned patterns of production (i.e., language-specific coarticulatory constraints) that serve to maintain perceptual distinctiveness even at quite rapid rates of speaking (Diehl and Lindblom, 2004; Lindblom, 1990). We therefore expected to find differences across languages in the distributional characteristics of mid-syllable spectral structure when the vowels were produced in various phonetic contexts. As a consequence, these cross-language differences in variation would result in different patterns of *cross-language* spectral similarity in different phonetic contexts. Because of the different phonological/phonetic functions of vocalic duration across the three languages, we also expected to find language-specific patterns of temporal variation with changes in phonetic context and speech style (Hay *et al.*, 2006).

II. METHOD

A. Speakers

Three female and three male monolingual speakers of each language were selected from a larger set of speakers initially recorded, including 10 German speakers (6 females, 4 males); 17 French speakers (9 females, 8 males), and 11 American speakers (3 females; 8 males). Reasons for removing speakers were based on language background questionnaires and judgments of phonetically trained native speakers of NG, PF, and New York English (hereafter AE), respectively. Reasons included: (a) wrong dialect or too bilingual (8 speakers), (b) dyslexia or trouble reading the orthographic representations of vowels (6 speakers), and (c) “sloppy” speech (1 speaker). An additional 5 speakers were not used because of failure to follow directions or because of difficulties with acoustic analysis (unclear formants). The 18 selected speakers ranged in age from 20 to 48 years old (NG 22–29 years; PF 20–36 years; AE 33–48 years). Although all speakers were recorded in New York City, the German and French speakers were either visitors (9 speakers) or very short-term residents of the US (2 subjects, 1 month; 1 subject, 3.6 months). French and German speakers (and some Americans) had formal foreign language classes in high school and/or college, but reported that they were not able to converse in any non-native language. The language background questionnaires established that the speakers had spoken the appropriate dialect throughout their lives, with little exposure to other dialects or languages in their immediate families.

B. Stimulus materials

1. German

The 14 vowels were represented by standard German orthography, which specifies phonetic vowel identity unambiguously. For citation utterances, vowels were embedded in disyllables $hVb(\partial)$, with primary stress on the target vowels: /i/Hieba, /ɪ/Hibba, /e/Hehba, /ɛ/Hebba, /u/Huhba, /ʊ/Hubba, /o/Hohba, /ɔ/Hobba, /y/Hühba, /ʏ/Hübba, /ø/Höhba, /œ/Höbba, /ɑ/Hahba, /a/Habba. Each vowel was assigned a number (1–Hieba, 2–Hibba, etc.) in order to assure that the identity of the intended vowel was retained during acoustic analysis and to facilitate random sequencing of items. For sentence utterances, target nonsense trisyllables $/g\partial CVC\partial/$ were embedded in the sentence, “Ich habe fünf _____ gesagt.” (I said five _____).

Protocols consisted of randomized lists of 15 utterances; the first and last utterance contained the same target vowel and the final utterance was discarded to control for list-final intonation effects. Pages of the protocol were arranged such that the first page (two lists) was always the citation utterances, and the next 4 pages were the sentences, with the consonantal context fixed for each page.² The second half of the protocol was identical to the first half except that the vowels appeared in different random orders. Thus, speakers recorded two exemplars of each vowel in each context in the first half of the session, and then repeated the whole protocol for a total of four utterances of each vowel in each context.

2. French

Pilot testing indicated that French speakers stressed the target vowel only when it occurred in the final syllable of the nonsense word. However, final consonants are heavily released in French. In addition, the vowel /e/ does not occur in closed syllables. Thus, for citation utterances, the protocols contained words spelled in French orthography as follows: /i/ hibe, /e/ héb'a, /ɛ/ hèbbe, /u/ hoube, /o/ haube, /ɔ/ hobe (botte), /y/ hube, /ø/ heube, /a/ habe. These words were produced as $Vb(\partial)$ syllables with an audible, schwa-like release. For sentence materials, disyllables $/raCVC/$ (e.g., “rabibes, rabipes, radides, radites”) were embedded in the sentence, “J’ai dit neuf _____ à des amis” (I said nine _____ to some friends). The /r/ is produced as an uvular fricative in French, and so has a similar place of articulation to /g/ in German and English. Since French does not have reduced vowels, the preceding/following /a/ was chosen as comparable in front-back position to schwa in German and English. The organization of the protocols was the same as for German.

3. English

For citation disyllables, words were spelled: /i/ heeba, /ɪ/ hibba, /e/ hayba, /ɛ/ hebba, /æ/ habba (hat), /u/ hooba, /ʊ/ hUba (should), /o/ hoaba (road), /ɔ/ hawba, /ɑ/ hobba, /ʌ/ hubba. As noted, real words were used to aid in pronouncing three vowels with very ambiguous/variable spelling in English. For sentence materials, the trisyllables $/g\partial CVC\partial/$ were embedded in the sentence: “I said five _____ this time.” The spelling of the vowels was the same as above and real-word prompts were printed above the nonsense word in the sentences for [æ, ʊ, o]. Pages were arranged the same as for the German and French protocols.

C. Recording procedures

Participants were recorded while seated in an IAC chamber with the experimenter outside in visible contact through the window. An intercom allowed the experimenter to converse with the speaker while monitoring the recording input over headphones (Sennheiser HD565 Ovation). Output from a dynamic microphone (Shure SM48), placed about 15 cm from the speaker’s mouth, was fed to a microphone preamplifier (Earthworks Lab 101) and then directly into a Soundblaster Live Wave DF80 sound card of a PC computer (Dell Dimension XPS B800). The stimuli were digitized (22.05 kHz, 16-bit) as monaural computer files, using SoundForge™ 5.0 software. The preamplifier level was set such that input signals varied from the equivalent of –18 to –3 dB of maximum capacity to preclude peak clipping of any part of the signal. Signal/noise-floor ratios of digital recordings were thus about 30–40 dB.

The experimenter spoke only in the speaker’s native language throughout the entire session. Speakers were given practice with the protocols with feedback about the style in which they were to speak. For citation utterances, they read through the items and the experimenter answered questions about pronunciation of the nonsense words. The speakers were instructed to read the number preceding each nonsense

word, then pause, and read the nonsense word with falling intonation (demonstrated by the experimenter). The speakers then practiced reading the list of citation utterances while the experimenter set the gain control for the preamplifier. The experimenter corrected their productions if they clearly misread the nonsense items, failed to read the identifying number, or did not pause between the number and the target word. Finally, speakers were instructed to speak a little more loudly than they would normally in a quiet room, as if they were talking through the window to the experimenter. They were also instructed not to speak too carefully, but just to say the words normally. The speakers then read the first column of citation utterances; if any were clearly mispronounced or if they read the wrong number, the experimenter asked them to repeat that utterance (referred to by number). They then progressed to the second column of utterances, with corrections at the end if needed.

After the citation materials were recorded, speakers were familiarized with the sentence materials and were given practice reading them. They were instructed to read the sentences as if they were talking to a friend who was a native speaker of the language and not to emphasize the nonsense word or put any pauses in the sentence. After the speakers had practiced until they read the sentences fluently, recording began. Again, they were reminded to speak a little more loudly, but not too slowly or too over-correctly, and they were told that they could reread a sentence at any time (self correct). Speakers were allowed to take short breaks and drink water any time during the recording session. The entire session, including reading and signing the Human Subjects consent form and filling out the language background questionnaire took approximately 2 hours.² Subjects were paid for their participation.

D. Acoustic analysis procedures

Digitized strings of utterances were separated into individual files and after coding, the spoken identifying numbers were removed from the files. Vowel productions were independently verified as “good” exemplars of the intended vowels by a native speaker (not one of the experimenters or speakers) of each language. Acoustic analysis was accomplished in a two-step process, using a specialized Matlab™ 6.0 program (CVCZ by Valeriy Shafiro). First, the waveform and a time-synchronized wideband spectrogram (0–4 kHz) were displayed and target syllable onsets and offsets were determined manually by the experimenter: for citation utterances, onset was defined as the onset of voicing (i.e., not including the voiceless /h/ for NG and AE utterances), and offset was defined as the beginning of /b/ closure, determined from the offset of upper formant energy and decrease in waveform amplitude. For CVC syllables, onset was defined as the release burst of the /b, d/, and offset was defined as the beginning of closure of the final stop (and in the case of flap [r] for /d/ and /t/ in English, the point at which amplitude reached a minimum). Thus, vocalic duration included the entire gesture from release of the preceding consonant to the beginning of full closure for the following consonant.

Second, values for the first three formants (F1, F2, F3)

were determined for three 23-ms windows placed at 25%, 50%, and 75% through the vocalic duration of the syllable using FFT and LPC analysis. The three FFT spectra (512 points) were displayed from left to right, with formant center frequencies estimated from LPC algorithms (24 coefficients) superimposed as vertical lines on the spectra. When estimated formants were clearly in error, hand corrections were made; three kinds of errors occurred: (a) formants were merged when two formants were very close together, (b) the third formant was missed when amplitudes were low, and (c) occasionally, spurious formants were identified.³ Corrections were based on the experimenter’s judgments from comparisons of FFT spectra and LPC formant tracks superimposed on the spectrographic display. All hand-corrected formant frequencies were obtained independently by two experimenters and any discrepancies were resolved by a third experimenter (the first author). Approximately 5% of the measurements of F1, 10% of the measurements of F2, and 17% of the measurements of F3 were obtained by hand corrections. For each speaker, mid-syllable (50% point) formant frequencies of the four repetitions of each vowel in each context were plotted in F1/F2 and F1/F3 plots for a final check for measurement errors. For citation utterances, F1 and F2 formant trajectories (25%–50%–75%) were also plotted to note direction and extent of formant movement throughout the middle half of the syllables indicating diphthongization (vowel-inherent spectral change).

III. RESULTS

To provide an initial overview of the variability of vowels in the three languages, Fig. 1 presents scatter plots of mid-syllable formant frequencies (F1/F2 in Bark) for all tokens produced by the three male speakers of each language (female plots look very similar). Thus, there are 60 tokens of each vowel, including the citation utterances and the vowels in sentences in bVb, bVp, dVd, and dVt contexts (5 contexts \times 3 speakers \times 4 repetitions). Coordinates are arranged so that high vowels appear at the top of the plots, front vowels to the left.

As the plots show, there was extensive contextual variability in the mid-syllable formant frequencies of vowels in all three languages. In general, the back vowels varied more than front vowels in F2 (front-back dimension), while the low vowels varied more than the high vowels in F1 (tongue height dimension). There was considerable overlap in F2 of front, unrounded and rounded vowels for both PF and NG, with generally less overlap of front and back, rounded vowels. There was considerable overlap in F1 for so-called high vowels [i, y, u] and mid vowels [e, ø, o] in all three languages, between so-called mid-high [ɪ, ʏ, ʊ] and mid vowels in NG and AE, and between mid vowels and mid-low vowels [ɛ, ʌ, ɔ] in PF and AE.

In the following sections, comparisons of vowels in citation utterances are presented first (Sec. A). Then analyses of sentence materials (Sec. B) explore the effects of consonantal context on vowels within and across languages. In each section, figures presenting average formant data for male and female utterances are shown for each language;

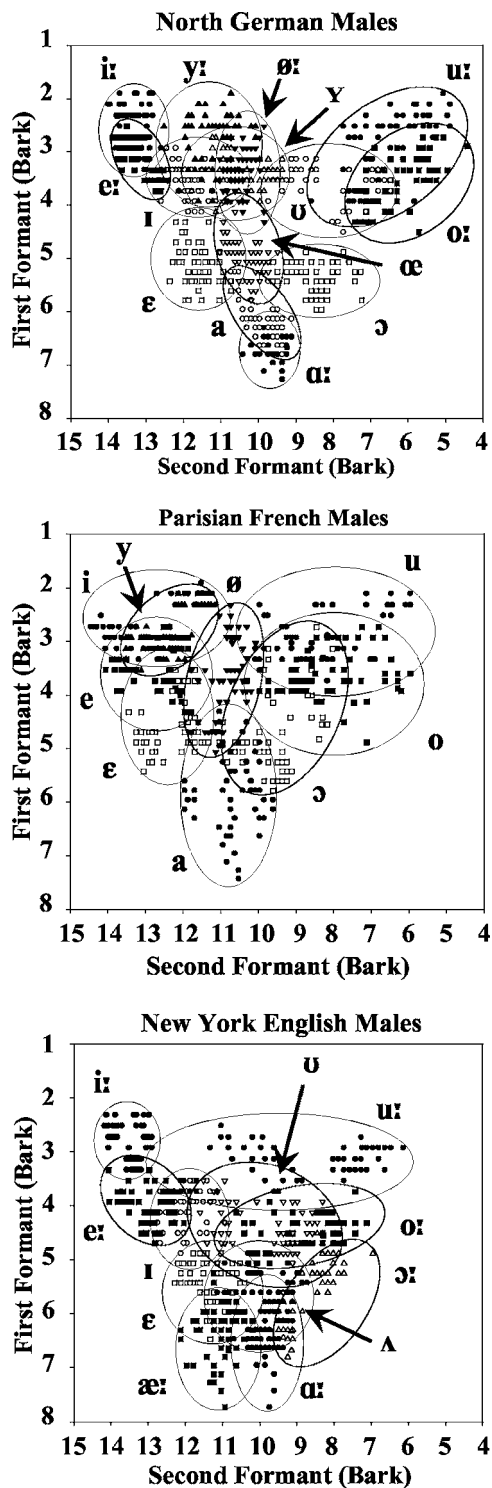


FIG. 1. F1/F2 plots (Bark) of all vowels produced by male speakers of North German (top), Parisian French (middle), and New York English (bottom). For North German and New York English, short vowels are plotted with open symbols; for Parisian French, mid-low vowels are plotted with open symbols. Ellipses were placed by hand to surround all tokens of each vowel.

then within-language discriminant analyses of each language are presented. These analyses quantify the amount of spectral overlap of mid-syllable formant frequencies and the extent to which vowels are differentiated by vocalic duration differences. Finally, cross-language discriminant analysis results

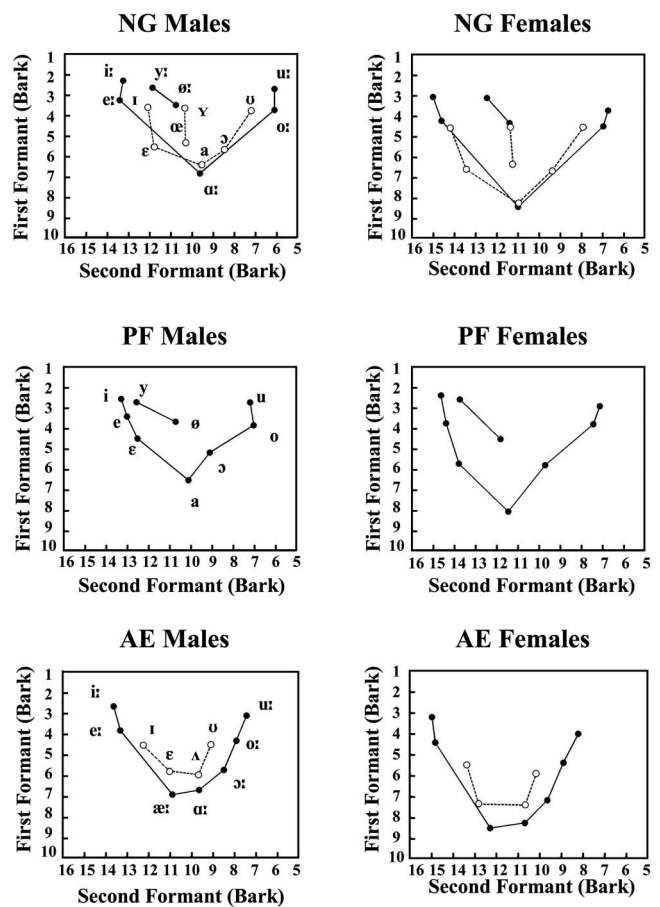


FIG. 2. Average mid-syllable F1/F2 values (Bark) for North German (NG), Parisian French (PF) and New York English (AE) male (left) and female (right) vowels produced in citation utterances. For NG and AE, short vowels are plotted with open circles.

are presented as a quantitative description of the acoustic similarity of NG and PF vowels to AE vowels.

A. Spectral and temporal structure of vowels in citation materials

Inspection of the results of acoustic analysis indicated that, for all three languages, the three speakers within each gender group showed quite similar patterns in the relative location of vowel targets (mid-syllable formant frequencies) in vowel space and similar relative duration differences.⁴ Thus, average Bark values for the first three formants at mid-syllable and vocalic durations (ms) were computed over all 12 tokens of each vowel in citation utterances (4 repetitions \times 3 speakers). Figure 2 presents the F1/F2 Bark plots for male (left) and female (right) utterances in all three languages. In these figures, the relative spacing of vowels in terms of tongue height (F1) and front-back position (F2) can be compared across languages. Appendices A–C list average Bark values for F1 and average Bark-difference values (F2-F1 and F3-F2) for each vowel for male and female tokens separately. In addition, the average vocalic durations of each vowel are given in the final two columns.

First, a comparison of the point vowels in the three languages indicates that differences in F1 values (tongue height) for high, front [i], and low [a, a] are quite comparable across

languages (F1 Bark differences for Male: NG=4.5, PF=4.1, AE=4.0; Female: NG=5.1, PF=5.7, AE=5.1). Thus, relative distances of other vowels from these vowels can be compared meaningfully. In comparing NG and AE, it is readily apparent that the so-called mid vowels [e:, ø:, o:] are as high as or higher (equal or lower F1 values) than the so-called mid-high vowels [ɪ, ʏ, ʊ]; these vowels are distinguished by vocalic duration; long/short (L/S) ratios ranged from 1.6 to 2.3 for NG, 1.4 to 1.7 for AE. In all three languages, high and mid, long vowels differed in F1 by less than 1.5 Bark for front, unrounded; front, rounded; and back, rounded pairs across both gender groups (NG range=0.9 to 1.4 Bark; PF range=0.9 to 1.3 Bark; AE range=1.2 to 1.4 Bark). These vowels were generally less differentiated by vocalic duration; mid/high duration ratios ranged from 1.1 to 1.3 for NG, 0.8 to 1.2 for PF, and 1.2 to 1.4 for AE.

In comparing the relative locations of vowels on the front-back dimension, languages appeared to differ systematically with respect to the high, back point vowel [u]. Thus, while [i] was very similar in F2 Bark values across languages (Male: NG=13.2, PF=13.3, AE=13.6; Female: NG=15.0, PF=14.6, AE=14.9), NG and PF [u] were further back (lower F2) than the AE [u] (Male: NG=6.0, PF=7.1, AE=7.4; Female: NG=6.7, PF=7.1, AE=8.2).

The front, rounded vowels of both NG and PF were closer to front, unrounded vowels than back, rounded vowels, with the high, front [y] having relatively higher F2 Bark values (Male: NG=11.8, PF=12.5; Female: NG=12.4, PF=13.7) than the other front, rounded vowels [ø, (ʏ, œ)] (Male: NG=10.7, 10.3, 10.2; PF=10.7; Female NG=11.3, 11.3, 11.2; PF=11.8). These comparisons also show that PF [y] had a higher F2 value on average than the “same” vowel in NG, with the PF [y] less than 1 Bark lower in F2 than PF front, rounded [i].

One final observation from these vowel spaces regards the “centralization” of short (lax, open) vowels, relative to long (tense, close) vowels in NG and AE. In general, the short vowels of AE were more centralized on both F2 and F1 dimensions than the short vowels of NG in these citation utterances. However, NG vowels tended to be more differentiated temporally than AE vowels in this context (see Appendices). Notice that for both languages, mid-high and mid-low short vowels are considerably lower in vowel space (higher F1) than their high and mid, long counterparts, respectively.

1. Within-language discriminant analyses

While these comparisons of central tendencies of vowel categories across languages begin to capture important (dis-)similarities across languages, a more quantitative comparison of *distributions* of vowels is needed to assess the contributions of mid-syllable formant frequencies and vocalic duration to the acoustic differentiation of vowel categories within each language. Linear discriminant analyses (Klecka, 1980) were performed on male and female vowel sets separately (12 tokens/vowel), using Bark values for F1/F2/F3 as input parameters in one set of analyses, and F1/F2/F3 plus vocalic duration in a second set of analyses of the three languages.⁵ This statistical technique optimizes the weighting of input parameters for maximum separation of categories, estab-

TABLE I. Overall results of within-language discriminant analyses (percent tokens correctly classified): Citation disyllables.

Language		Input Parameters	
		F1/F2/F3	Fs + Duration
NG	M	85	95
	F	80	90
PF	M	89	92
	F	93	92
NYE	M	83	93
	F	87	93

lishes a center-of-gravity in parameter space for each category, and then yields a classification matrix of all tokens. “Correct” classification indicates that a token of an intended vowel was classified as closest to the center-of-gravity for that category.

Table I presents the overall rates of correct classification for each gender group within each language; the first column gives the results of the analyses with spectral parameters only, while the second column shows the analyses when duration was added as an input parameter. Looking first at the analyses with formants only, overall correct classification rates were best for PF vowels, and somewhat lower for NG and AE vowels. For PF vowels, correct classification of individual vowel categories ranged from 58% to 100%. Confusions were mostly between front, unrounded and rounded vowels [e/y] and [i/y], between mid and mid-low, front [e/ε]; and between high and mid, back vowels [u/o]. For NG vowels, correct classification of individual vowel categories ranged from 58% to 100%. Errors were mostly confusions between long and short vowels [i:/i, e:/e, u:/u, o:/o, ø:/y, a:/a], but high and mid, long vowels [e:/i:, u:/o:, y:/ø:] were also sometimes confused, as were front, rounded vowels with front, unrounded or back, rounded vowels [y:/i, ʏ/ʊ]. For AE vowels, correct classification of individual vowel categories ranged from 50% to 100%. As with NG, the majority of misclassifications were confusions between long and short vowels [e:/i, æ:/ε, æ:/Λ, a:/Λ, o:/u, ɔ:/u, u:/u], but confusions also occurred between long vowels [e:/i:, a:/o:, u:/o:].

Correct classification improved for NG and AE vowels when vocalic duration was included as an input parameter, while classification of PF vowels remained almost unchanged, and misclassification patterns were very similar. For NG vowels, remaining misclassifications were primarily among back vowels [u:, ʊ, o:]; for AE vowels, remaining misclassifications were of [e:/i:], [a:/Λ/ɔ:], and [u:/o:].

In summary, distributions of vowels produced in citation utterances were acoustically differentiated quite well overall in all three languages on the basis of mid-syllable formant values and vocalic duration. As expected, vocalic duration was most important in differentiating vowels in NG and least important in PF. In all three languages, the high and mid, back vowels [u/o] were sometimes confused. In PF, front, unrounded and rounded vowels [e/y, i/y] were sometimes confused, while in NG, these confusions did not occur. In AE, confusions occurred between front vowels [i:/e:], while these vowels were not confused in either NG or PF. In PF, the front vowels [e/ε] were sometimes confused (cf.

TABLE II. Cross-language spectral similarity of NG, PF, and NYE point vowels: Discriminant analyses (F1/F2/F3 Bark) for citation disyllables. Male and Female results are combined.

NG vowels	Modal AE V	% tokens classified	Other Vs >10%
i:	i:	96	
ɑ:	ɑ:	96	
u:	u:	96	
a	ɑ:	79	ɔ:
PF vowels			
i	i	100	
a	æ:	41	ɔ:, ɑ:
u	u	100	

Gottfried, 1984), whereas in NG and AE, these vowels were never confused. These differences in patterns of spectral overlap reflect language-specific differences in the relative locations of vowels in vowel space.

2. Cross-language discriminant analyses

For comparisons across languages, we were most interested in how NG and PF vowels would be classified with respect to their acoustic similarity to AE vowels, as an acoustic basis of comparison with perceptual data from English listeners tested on French and German vowels (Levy and Strange, in press; Strange *et al.*, 2005; Strange, Levy, and Lehnhoff Jr., 2004). Thus, the AE vowel corpora served as the training set (upon which parameter weightings and centers-of-gravity were established), then NG and PF vowels served as the test sets. Male and female corpora were analyzed separately. Analyses using F1/F2/F3 Bark values and vocalic duration as input parameters produced poorer matches of “similar” vowels across languages than analyses using only formants. Thus, only the latter analyses that establish the *spectral* similarity of NG and PF vowel categories to AE vowel categories are presented here and in subsequent cross-language analyses.

The two point vowels, NG [i:, u:] and PF [i, u], were consistently classified as most similar to AE [i:, u:], respectively (shown in Table II for male and female data combined; 24 tokens/vowel). Thus, even though average F2 values for [u] differed across languages, NG and PF [u] were not more similar to any other AE vowel. While the NG low vowel [ɑ:] was a good fit to AE [ɑ:], the distribution of the PF low vowel [a] straddled the front and back AE low vowels [æ:, ɑ:], and nine tokens were more similar to AE [ɔ:]. In contrast, most tokens of NG [a] were classified as AE [ɑ:], with only five tokens classified as [ʌ, ɔ:].

Having established that [i, u] were acoustically similar across languages, we next asked how NG and PF front, rounded vowels compared acoustically to front and back AE vowels. Table III presents the results of cross-language discriminant analyses for these vowels, combined over analyses of male and female corpora. Overall percentages of tokens classified as AE front vowels, summed over front categories (listed in column 2) and back vowels, summed over back

TABLE III. Cross-language spectral similarity of NG and PF front rounded vowels to AE front and back vowels: Discriminant analyses (F1/F2/F3 Bark) for citation disyllables. Male and Female results are combined.

NG vowels	AE front vowels	% test tokens	AE back vowels	% test tokens
y:	ɪ, e:, i:	100		
ø:	ɪ	71	ʊ, o	29
ɤ	ɪ	63	ʊ	37
œ	ɛ	54	ʊ, ʌ	46
PF vowels				
y	e:, i:	100		
ø	ɪ, ɛ	88	ʊ	12

categories (listed in column 4) are given (columns 3 and 5). Results indicated that the NG front, rounded vowels were acoustically intermediate between front and back AE vowels except for [y], which was always classified as more similar to front AE vowels [ɪ, e:, i]. For PF front, rounded vowels, all tokens of [y] were classified as similar to AE front vowels, and only three tokens of [ø] were classified as more similar to AE back vowels. Thus, in general, PF front, rounded vowels were more consistently classified as falling within front AE vowel categories than were NG front, rounded vowels.

Third, it was of interest to look at classification patterns for the remaining NG and PF vowels, often transcribed with the same symbols as for AE counterparts. Table IV presents the results of cross-language discriminant analyses for these vowels, again collapsed over male and female analyses. In this table, the AE vowel category to which each NG and PF vowel was most often classified is given in column 2 with percentages of tokens so classified in column 3. Other AE categories to which at least 10% of the tokens were similar are presented in column 4. As these results indicate, three NG vowels [ɪ, e:, ʊ] and three PF vowels [ɛ, o, ɔ] were not

TABLE IV. Cross-language spectral similarity of NG and PF mid-high, mid, and mid-low vowels to AE vowels: Discriminant analysis (F1/F2/F3 Bark) for citation disyllables. Male and Female results are combined.

NG vowels	AE modal vowel	% test tokens	Other AE vs >10%
ɪ	e:	63	ɪ
e:	i:	54	e:
ɛ	ɛ	54	ɪ
ʊ	o:	54	u:
o:	o:	92	
ɔ	ɔ:	96	
PF vowels			
e	e:	63	i:
ɛ	ɪ	84	ɛ
o	u:	50	o
ɔ	ʊ	46	ɔ:, ʌ

classified as most similar to their AE transcriptional counterparts a majority of the time. In the case of NG [ʊ] and PF [ɔ], none of the tokens were classified as most similar to AE [ʊ, ɔ:], respectively. In contrast, NG [o:], [ɔ] were relatively good spectral matches to their AE counterparts, while NG [ɛ] and PF [e] distributions straddled AE [ɛ, ɪ] and [e:, i:] categories, respectively.

In summary, the majority of mid-high, mid, and mid-low NG and PF vowels that are transcribed as the “same” as AE vowels were found not to be spectrally similar to their transcriptional counterparts in AE. In general, NG and PF vowels were acoustically more similar to higher AE vowels, or were distributed such that they overlapped with two spectrally adjacent AE categories. The “non-native” PF front, rounded vowels were spectrally more similar to front than to back AE vowels, while NG front, rounded vowels, except for [y:], fell in between front and back AE vowels. Finally, NG [i:, a:], [ɔ, o:], [u:] and PF [i, u] were most similar spectrally to their transcriptional counterparts in AE.

B. Spectral and temporal structure of vowels in sentence materials

In this section, the acoustic characteristics of coarticulated vowels produced in sentences are reported. Of specific interest was a comparison of (dis)similarities across languages in the contextual variability of spectral and temporal parameters as a function of the consonantal context in which they were produced. Several comparisons were made: (1) Average mid-syllable formant frequencies of vowels produced in labial and alveolar consonant contexts were evaluated with respect to how they differed from canonical values derived from citation utterances, (2) effects of consonantal context on relative duration differences were compared within and across languages, (3) context-dependent and context-independent discriminant analyses established the spectral overlap of coarticulated vowel distributions and evaluated those distributions relative to canonical values, and (4) cross-language discriminant analyses evaluated differences in spectral similarity of distributions of NG and PF vowels to AE vowel categories across contexts.

1. Contextual variation in mid-syllable formant frequencies

Average F1/F2/F3 Bark values for vowels produced in each consonantal context (bVb, bVp, dVd, dVt) were computed for individual speakers of each gender/language group. Within groups, patterns of variation in formant structure across contexts were very similar for the three speakers; thus, average values were computed for comparison with canonical values obtained from citation materials. Inspection of the F1/F2 and F1/F3 plots for vowels produced in the four contexts revealed that patterns for bVb and bVp contexts were almost identical, as were patterns for dVd and dVt contexts. Thus, Figs. 3 and 4 present average F1/F2 values for vowels in labial and alveolar contexts, collapsed over the two labial and two alveolar final consonants, respectively (24 tokens/vowel: 2 contexts × 3 speakers × 4 repetitions). As a refer-

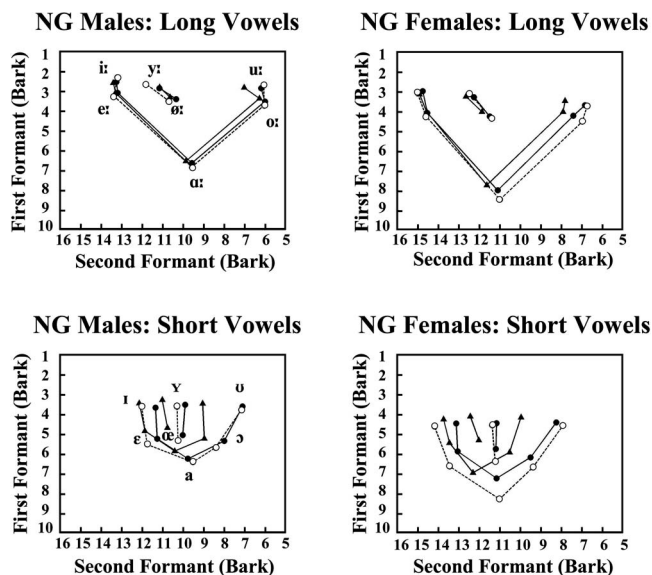


FIG. 3. Average mid-syllable F1/F2 values (Bark) for North German (NG) vowels produced in labial (closed circles) and alveolar (closed triangles) consonant contexts in sentence materials. Average values for long (top) and short (bottom) vowels for male (left) and female (right) speakers are plotted separately. For comparison, average values for citation utterances (open circles) are also shown (same as Fig. 2).

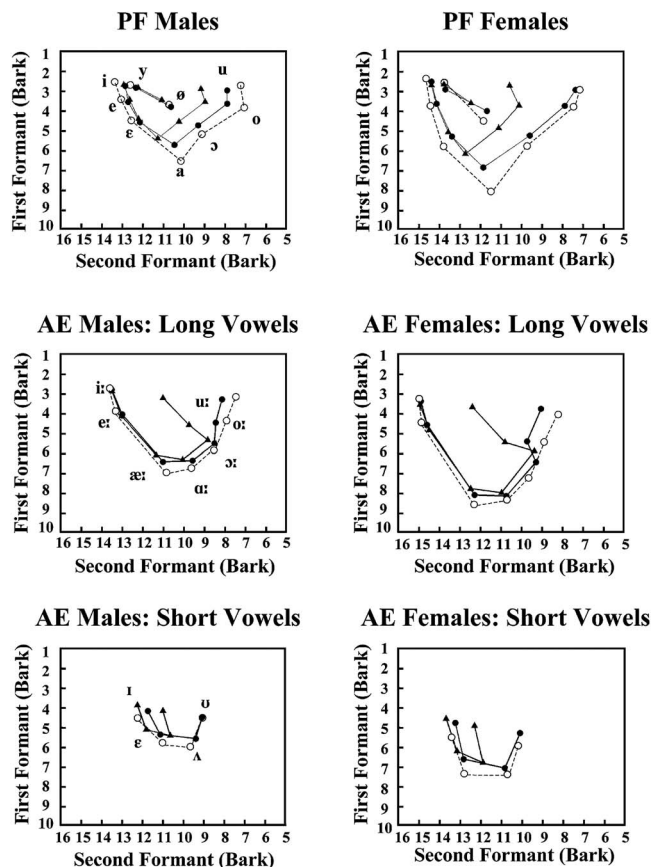


FIG. 4. Average F1/F2 values (Bark) for Parisian French (PF) vowels (upper) and New York English (AE) long vowels (middle) and short vowels (lower) produced in labial (closed circles) and alveolar (closed triangles) consonant contexts in sentence materials. For comparison, average values for citation utterances (open circles) are also shown (same as Fig. 2).

ence, the average values for citation utterances, shown in Fig. 2, are reproduced in these figures, indicated by open symbols and dashed lines.

Figure 3 presents the F1/F2 Bark plots for the NG corpora, with long vowels and short vowels plotted separately for clarity. It is immediately clear that variation in target formant frequencies as a function of coarticulation with labial and alveolar consonants was considerably greater for short vowels than for long vowels for both gender groups, and in general, greater in alveolar than in labial contexts. A second obvious trend is that contextual variation of low and back vowels was greater than for front vowels. Figure 4 presents the F1/F2 Bark plots for the PF corpora (top row), and the AE long vowels (middle row) and short vowels (bottom row), separately. As with the NG vowels, the PF low and back vowels showed greater contextual variation than the front vowels. Second, PF vowels varied considerably more than NG long vowels and perhaps even somewhat more than NG short vowels, especially in alveolar contexts. In contrast, AE long vowels, like NG long vowels, varied relatively little, except for [u:, o:], which showed extreme fronting (higher F2) in alveolar contexts, while AE short vowels showed somewhat larger changes from canonical values.

To quantify these shifts from canonical values more precisely, and to incorporate changes in F3 (not shown in Figs. 3 and 4), average F1/F2/F3 Bark values for each vowel in labial and alveolar contexts for each gender group (24 tokens) were rendered into Bark-difference values (F2-F1 and F3-F2). Then, the Euclidian distance between these values and the average Bark-difference values for citation utterances was computed for each vowel (see Appendices). These Euclidian distances thus represent shifts of vowels in a formant-ratio vowel space as a function of coarticulatory undershoot of canonical acoustic targets (see Fourakis 1991, for a similar analysis).

As Figs. 3 and 4 suggest, shifts from canonical values for coarticulated vowels in labial contexts were, on average, relatively small in all three languages. Euclidian distances for NG vowels across all vowels and both gender groups averaged 0.5 Bark (range=0.1 to 1.3 Bark). For PF vowels the Euclidian distances were also quite small (mean=0.6 Bark, range=0.2 to 1.7 Barks), as they were for AE vowels (mean=0.6, range=0.1 to 1.4 Barks). The largest shifts were for NG [ɪ, a], PF [a], and AE [o:, u:].

In contrast to the relatively small coarticulatory undershoot of vowels in labial contexts, all three languages showed consistently larger contextual effects for vowels in alveolar contexts. Table V presents mean Euclidian distances of vowels in alveolar contexts from canonical values for all three languages (male/female data shown separately). Front, unrounded vowels (rows 1–4) shifted relatively little from canonical values; only NG and AE [ɛ] shifted more than 1 Bark. NG front, rounded short vowels [y, œ] and female PF [ø] (rows 6–8) showed slightly greater fronting in this context. In contrast, most of the low and back vowels in all three languages showed much greater shifts from canonical values, and striking differences in coarticulatory patterns across languages were revealed.

The different patterns of coarticulatory change for back

TABLE V. Euclidian distances (Bark) from canonical values for NG, PF, and NYE vowels in alveolar context: Male/female data are presented separately (M/F).

Vowel	NG	PF	NYE
front			
i:	0.3/0.2	0.6/0.5	0.4/0.4
e:	0.2/0.4	0.5/0.2	0.5/0.7
ɪ	0.3/0.3	—	0.8/0.2
ɛ	0.8/1.2	0.5/0.6	1.6/1.5
y:	0.9/0.1	0.5/0.4	—
ø:	0.2/0.9	0.6/1.6	—
ɤ	1.2/1.7	—	—
œ	1.2/2.0	—	—
Low/back			
æ:	—	—	1.4/1.1
ɑ:	0.8/1.6	—	1.0/0.9
a	1.7/2.9	2.3/3.3	—
ʌ	—	—	1.9/2.2
ɔ/ɔ:	1.4/2.3	2.0/2.6	1.0/1.1
ʊ	2.9/3.2	—	3.2/3.8
o:	0.7/1.7	2.9/3.6	2.7/2.9
u:	1.3/1.7	2.5/5.0	5.0/6.3

and low vowels in alveolar contexts, illustrated in Figs. 3 and 4, are borne out in this analysis. While NG long vowels [ɑ:, o:, u:] varied relatively little from canonical values, the short vowels [a, ɔ, ʊ] shifted more. Thus, there was shrinking of the NG short vowel space, with low and back vowels raised and fronted, while front vowels remained relatively stable. This resulted in an increase in the relative distances between long and short, back and low vowels. For PF, all four low and back vowels differed from canonical values by 2 or more Barks. The shrinking of the vowel space was greater than for NG short vowels; however, the shape of the vowel space (i.e., relative locations of vowels) remained similar across these contextual variations. For AE vowels, the Euclidian distances from canonical values differed strikingly across vowels, illustrating that the *shape of the vowel space* changed for vowels produced in alveolar contexts. Whereas low and mid-low vowels [æ:, ɑ:, ʌ, ɔ:] shifted up and forward relatively little, the mid to high back vowels [o:, ʊ, u:] shifted drastically forward, becoming very similar to front, rounded vowels in NG.

2. Effects of speech style and phonetic context on vocalic durations

Mean durations of vocalic nuclei for male and female tokens were computed for vowels produced in bVb, bVp, dVd, and dVt context in sentences and were compared with durations of vowels produced in citation utterances. In addition, long/short (L/S) duration ratios for NG vowels (7 L/S pairs) and AE vowels (4 L/S pairs) were compared in citation and sentence utterances. These descriptive data are presented in Table VI. Regarding NG vowels, several generalizations can be noted. First, long vowels in sentences in bVb, dVd, and dVt contexts were only slightly shorter, on average, than these vowels in citation utterances, while short vowels were generally slightly longer in sentences than in citation utter-

TABLE VI. Vocalic durations (ms) and Long/Short ratios in citation and sentence utterances. Male and female data are listed separately (M/F).

Language/ context	Long (ms)	Short (ms)	Long/short ratio
NG citation	145/129	75/87	1.9/1.5
bVb	135/135	84/95	1.6/1.4
dVd	144/144	96/104	1.5/1.4
bVp	118/120	77/84	1.5/1.4
dVt	133/134	85/94	1.6/1.4
PF citation	113/159		
bVb		91/97	
dVd		102/107	
bVp		84/93	
dVt		96/110	
NYE citation	121/150	83/100	1.5/1.5
bVb	154/147	104/104	1.5/1.4
dVd	162/170	122/127	1.3/1.3
bVp	131/129	89/91	1.5/1.4
dVt	163/168	125/128	1.3/1.3

ances. Long and short vowels were somewhat shorter in bVp context. When L/S ratios were compared, there were only small differences across speech styles and phonetic contexts within gender groups, except for male citation utterances; on average, long/short differences were greater for male utterances.

Some of the same trends are seen in AE vowels, where long vowels were actually longer, on average, in bVb, dVd, and dVt utterances than in citation utterances. (For AE utterances, the final /d/ and /t/ were both produced as voiced flap [r] and vowel duration differences were totally neutralized for all six speakers.) L/S ratios for vowels in sentence materials tended to be smaller than in citation materials. Comparisons of NG and AE sentence materials reveal that NG L/S ratios tended to be larger than AE ratios. In addition, there were differences across NG and AE for the low vowel pairs. NG [ɑː, a] were relatively more differentiated temporally in all four contexts (L/S ratios from 1.6 to 1.8) than were AE [ɑː, ʌ] (L/S ratios from 1.3 to 1.5).

PF vowels were, on average, considerably shorter in sentence materials than in citation utterances, especially in labial contexts. In sentence materials, PF vowels were closer in duration to NG and AE short vowels than to long vowels. This could account, in part, for the relatively large shifts in mid-syllable formant values reported in the previous section for PF vowels [o, u], in comparison with the “same” NG vowels [oː, uː].

3. Within-language discriminant analyses

Given the (dis)similarities across languages in the extent of contextual variation in average mid-syllable formant frequencies (Sec. B 1) and relative vocalic durations (Sec. B 2), the next step was to examine the acoustic differentiation of distributions of coarticulated vowels across languages and contexts. In all, 18 different discriminant analyses were computed for the sentence materials, combining various consonantal contexts: Corpora of vowels produced in labial (bVb,

TABLE VII. Within-language discriminant analyses (% tokens correctly classified) for sentence materials. Male and female data are listed separately (M/F).

Contexts	F1/F2/F3 + duration	Re: citation
NG labial	93/93	83/86
NG alveolar	90/88	72/57
NG both	87/88	77/71
PF labial	81/94	63/87
PF alveolar	82/85	55/47
PF both	79/81	59/67
AE labial	91/96	82/87
AE alveolar	92/98	30/51
AE both	89/89	56/69

bVp) and alveolar (dVd, dVt) contexts were examined separately in (place) context-dependent analyses, and then vowels produced in all four contexts were entered into a single context-independent analysis. Again, corpora for male and female speakers were analyzed separately. Correct classification rates for each of these input sets were compared to determine the extent of acoustic overlap of vowel categories in each context and when labial and alveolar contexts were combined. Finally, to examine the extent of spectral and temporal *change* from canonical target values across gender/language groups, 18 additional analyses were computed in which the citation materials served as the training sets, while the sentence materials served as the test sets. These analyses allowed us to examine the extent to which distributions of particular vowels produced in sentence materials shifted so they were actually closer to canonical values for a *different* vowel category.

Table VII presents the overall rates of correct classification for the male and female sentence materials for the labial, alveolar, and combined corpora when the parameter weightings and centers-of-gravity were optimized for the sentence materials for each gender/language group (left). The right-hand columns give the overall classification rates for sentence materials (same test sets) when the citation materials (from the same language/gender group) served as the training sets. In all these analyses, F1/F2/F3 Bark values plus vocalic duration were included as input parameters. As in the analyses of citation materials reported above, the inclusion of duration as a parameter improved classification rates for NG and AE vowels, with almost no change for PF vowels.

When parameter weightings and category centers-of-gravity were established on the coarticulated vowel corpora themselves, correct classification rates, overall, were not drastically reduced from those reported for citation materials. For NG vowels, classification rates for vowels in labial context were comparable to those reported for citation materials (see Table I). There were more misclassifications of vowels in alveolar context, and when labial and alveolar contexts were combined (context-independent analyses). For the latter analyses, misclassifications were not evenly distributed across the 14 vowel categories (range=75% to 100% correct). There were confusions between high and mid vowels [iː/eː, uː/oː, yː/øː] and other height confusions [y/œ,

$\varepsilon/a]$, as well as a few confusions between front, unrounded and rounded vowels $[i/y, \varepsilon/\alpha]$. Front, rounded vowels were almost never misclassified as back vowels; back vowels were only rarely misclassified as front, rounded or unrounded vowels.

The analyses of PF vowels using parameter weightings and centers-of-gravity established for the coarticulated corpora themselves yielded generally lower correct classification rates than for NG (and AE) vowels and relative to PF citation materials. This reflects the greater contextual variation of PF vowels (see Fig. 4) and the greater shrinking of the vowel space. In looking at the context-independent analyses, the misclassifications were not evenly distributed across all nine PF vowels (range=58% to 100% correct). There were confusions between front, unrounded and rounded vowels $[\varepsilon/y, e/y, \varepsilon/\phi]$; between back and front, rounded vowels $[\sigma/\phi, u/\phi]$, as well as misclassifications in height $[a/\sigma, y/\phi, u/\sigma]$.

The analyses of AE vowels using parameter weightings and centers-of-gravity established on the coarticulated vowels themselves yielded results more like those of NG than of PF. Again, overall rates of correct classification were about the same for vowels in both labial and alveolar context-dependent analyses as they were for citation utterances (Table I), and only slightly lower for context-independent analyses. For the latter, however, there were large ranges in classification rates across individual vowels (range=50% to 100% correct). Almost all misclassifications were height confusions among front vowels and among back vowels $[i:/I, \varepsilon/I, a:/\sigma, \sigma:/o:, \Lambda/\sigma, o:/\sigma, u:/\sigma]$ with only a few confusions of back and front short vowels $[\sigma/I, \Lambda/\varepsilon]$. Thus, despite the striking change in location of back, rounded vowels in alveolar contexts, these vowel distributions did not overlap with front vowel categories. The pattern of height errors indicated that long/short duration differences differentiated front vowels well (except for $[i:/I]$), but spectrally adjacent back vowel distributions overlapped both spectrally and temporally.

Turning to the question of the extent to which distributions of coarticulated vowels shifted from canonical values in the three languages, the discriminant analyses in which citation utterances were the training sets and sentence utterances were the test sets revealed striking differences across contexts and across languages. For all three analyses within each language, correct classification rates decreased; that is, for all languages, the acoustic structure of vowels differed as a function of speech style and phonetic context sufficiently that some tokens of particular vowels became closer to canonical centers-of-gravity for different vowels. However, distributions of vowels produced in labial contexts were, in general, more similar to canonical values, as would be expected given the relative independence of the tongue and lip gestures.

The overall classification rates of coarticulated vowels with canonical centers-of-gravity as the reference reflect cross-language differences in coarticulatory influences. Overall, the distributions of NG vowels showed the least variation from canonical values in both labial and alveolar contexts. In labial contexts, the decrease in correct classifi-

cations was relatively small overall and equally distributed across long and short vowels. Vowels in alveolar contexts were misclassified more often, with more short vowels misclassified (Long=77% correct, Short=48% correct). Classification errors included many misclassifications of back, rounded as front, rounded vowels $[\sigma \rightarrow \alpha, \sigma \rightarrow y]$; front, rounded as front, unrounded vowels $[y \rightarrow i, \alpha \rightarrow \varepsilon]$; and the low vowel as a higher vowel $[a \rightarrow \alpha]$, reflecting the general shifts of back and low, short vowels toward more front and higher positions when they were coarticulated with alveolar consonants.

When PF coarticulated vowels were evaluated against canonical values, correct classification rates decreased relatively sharply, and showed a somewhat different pattern across gender groups. Misclassifications of male utterances were concentrated on three vowels $[o, a, \phi]$: $[o, a]$ were misclassified as higher vowels, $[\phi]$ as front, unrounded $[\varepsilon]$, in both labial and alveolar contexts. Female utterances appeared not to vary as much from canonical values in labial contexts, but in alveolar contexts, misclassifications of $[\sigma, a, u]$ reflected both fronting and raising patterns and $[y]$ was also often misclassified as $[e]$.

The cross-context analyses of AE vowels revealed yet another pattern. Distributions of vowels coarticulated in labial contexts did not vary much from canonical values, and misclassifications were distributed across vowel categories. However, in alveolar contexts, misclassification rates were very high, and were concentrated on the back vowels $[u:, \sigma, o:, \Lambda]$, which were classified as spectrally closer to front vowels $[i, \varepsilon]$ in these contexts. Thus, in alveolar contexts, these back vowels were fronted so drastically that the majority of tokens were classified as more similar to canonical front vowels. In addition, there were some misclassifications for $[e:, i, \varepsilon, \sigma:]$ as higher vowels.

In summary, for all three languages, within-language discriminant analyses indicated that distributions of vowels produced in sentence materials overlapped in vowel space when vowels produced in both labial and alveolar contexts were included in the analyses. However, *context-dependent* analyses showed that, in each context, vowel distributions remained fairly well separated on the front-back dimension in all three languages even though context-specific fronting occurred. There was somewhat more overlap in the height dimension, especially for high and mid (long) vowels in all three languages. When distributions of coarticulated vowels were evaluated against canonical values, significant cross-language differences were revealed, as was expected from comparisons of mean data (Sec. B 1). NG vowels showed smaller shifts overall than the other two languages, except for short vowels in alveolar context. PF vowels in both contexts shifted more from canonical values than for NG vowels. Misclassification patterns indicated significant raising and fronting of $[a, o]$ and fronting of $[\phi, u]$, especially in alveolar contexts. Finally, coarticulated AE vowels shifted relatively little in labial contexts, while in alveolar contexts $[u:, \sigma, o:]$ showed extreme fronting, and $[\Lambda]$ both raising and fronting.

TABLE VIII. Cross-language spectral similarity of NG, PF, and NYE point vowels: Discriminant analyses (F1/F2/F3 Bark) for Sentence Materials. Male and female data are listed separately (M/F).

Labial and alveolar			Labial		Alveolar	
NG vowels	Modal NYE V	% tokens classified	Modal NYE V	% tokens classified	Modal NYE V	% tokens classified
i:	i:	98/97	i:	92/94	i:	100/92
ɑ:	ɑ:	100/77	ɑ:	100/92	ɑ:	100/71
u:	u:	81/100	u:	79/100	ɔ: ɒ:	50/100
PF vowels						
i	i:	100/100	i:	96/100	i:	96/100
a	ʌ, ʊ	44/35	ʌ, ʊ	29/42	ʌ, ʊ	54/21
	ɪ, ɛ, æ:	50/65	ɪ, ɛ, æ:	71/58	ɪ, ɛ, æ:	46/79
u	u:	96/98	u:	100/100	u:, ʊ, ɔ:	33/79

4. Cross-language discriminant analyses

Given the marked differences across languages in the pattern of contextual variation reported above, we expected that *cross-language* similarity patterns would differ from those for citation utterances, especially for the back and low vowels, and for the front, rounded vowels. For the context-independent comparisons, the male and female AE corpora that included all four contexts (48 tokens/vowels: 4 contexts \times 3 speakers \times 4 repetitions) served as the training sets, with the analogous NG and PF corpora as the test sets. For labial and alveolar (context-dependent) analyses, the two AE corpora for each gender group served as the training sets (24 tokens/vowel: 2 contexts \times 3 speakers \times 4 repetitions). For all analyses, input parameters included only F1/F2/F3 Bark values (as for citation materials); thus, these analyses establish the cross-language *spectral* similarity of distributions of coarticulated vowels produced in sentence materials.

Table VIII summarizes the cross-language classification patterns for NG and PF point vowels (column 1) for the context-independent analyses (columns 2–3), and the two context-dependent analyses (labial, columns 4–5; alveolar, columns 6–7). As these data show, almost all tokens of NG and PF [i] were classified as most similar to AE [i] in both context-independent and context-dependent analyses, reflecting the fact that this vowel was very similar across all three languages and varied little across contexts. Most of the NG and PF [u] tokens were classified as similar to AE [u] in the context-independent analyses, for which centers-of-gravity were computed over both back (labial) and fronted (alveolar) AE allophones. However, when only vowels produced in alveolar contexts were compared across languages, all of the NG [u:] tokens were more similar to AE [ɔ:, ɒ:], because of their lower F2 values relative to the fronted AE [u:]. The majority of male PF [u] tokens were also more similar to other AE vowels [ʊ, ɔ:]. NG [ɑ:] was classified predominantly as most similar to its AE counterpart, whereas PF [a] in all three comparisons was almost never classified as AE [ɑ:]. Rather, both male and female tokens were classified as a variety of higher front and back AE vowels, reflecting the large contextual variation of this PF vowel in sentence materials (see Fig. 1).

Table IX presents the results of cross-language discrimi-

TABLE IX. Cross-language spectral similarity (% tokens classified) of NG and PF front rounded vowels to AE back vowels: Discriminant analyses for Sentence Materials. Male and female data are listed separately (M/F).

NG vowels	NYE vowels	Labial and alveolar	Labial only	Alveolar only
y:	u:, ʊ	96/94	0/17	100/100
ø:	u:, ʊ, ɔ:	96/98	8/75	100/100
ɣ	u:, ʊ, ɔ:	94/71	75/54	96/100
æ	ɔ:, ʊ, ʌ, ɔ:	83/94	92/88	92/100
PF vowels				
y	u:	25/21	0/0	33/17
ø	u:, ʊ	50/77	0/21	71/96

nant analyses for the coarticulated NG and PF front, rounded vowels. Classification rates for each of these vowels to AE back vowels, collapsed over height categories, are presented, with overall percentages of classifications as back vowels for the context-independent analyses (column 3) and the two context-dependent analyses (columns 4–5). Again, results for comparisons of male and female utterances are presented separately because of differences in outcomes across genders for some vowels. Looking first at the NG vowels, cross-language similarity patterns established by the context-independent analyses showed a markedly different pattern from citation utterances (Table III, columns 4–5), with nearly all tokens of all four vowels being classified as more similar to back than to front AE vowels. This was due to the shifts in centers-of-gravity for AE back vowel distributions that included the fronted allophones produced in alveolar contexts. The context-dependent analyses corroborated this pattern, with NG and AE vowels in labial contexts showing cross-language similarity patterns resembling those in citation utterances (see Table III). In contrast, almost all tokens of the four NG vowels produced in alveolar contexts were classified as spectrally similar to the (fronted) AE back vowels produced in this context.

Cross-language spectral similarity patterns for the coarticulated PF front, rounded vowels also changed relative to citation utterances. In the context-independent analysis, the majority of tokens of [y] were still classified as more similar to front than to back AE vowels; however, [ø] tokens straddled front and back AE vowel categories. Again, the context-dependent analyses showed that this PF vowel was similar to the (fronted) AE back vowels produced in alveolar context.

There were also changes in cross-language spectral similarity patterns for some of the mid-high, mid, and mid-low vowels, as shown in Table X. NG [ɪ, ɛ] were classified as their AE counterparts more consistently in context-independent analyses (shown in the table) and in both context-dependent analyses (not included in the table) than in citation materials (Table IV). However, a majority of tokens of NG [e:] and PF [ɛ] were still not classified as most similar to their counterparts in AE. NG back vowels [ʊ, ɔ:], and PF back vowels [ɔ, ɔ] were also not classified as most similar to their AE counterparts.

To summarize, cross-language similarity of distributions

TABLE X. Cross-language spectral similarity (% classified) of NG and PF mid-high, mid, and mid-low vowels to NYE vowels: Context-independent discriminant analyses (F1/F2/F3) for sentence materials. Male and Female results are listed separately (M/F).

NG vowels	Modal NYE vowel	% tokens classified	Other NYE vowels (>10%)
ɪ	ɪ	73/91	u:/
e:	i:	69/54	e:/e:
ɛ	ɛ	81/65	ɪ/ɪ
ʊ	u:	67/75	o:/o:
o:	u:	56/88	o:/o:
ɔ	ɔ:	94/73	/o:
PF vowels			
e	i:	48/67	ɪ, e:/e:
ɛ	ɪ	48/52	eɪ, ɛ/eɪ, ɛ
o	ʊ/u:	44/92	ʊ, o:/
ɔ	ʌ/ʊ	27/33	ɛ, u:, ʊ, o:/ /o:, u:

of NG and PF coarticulated vowels to category centers-of-gravity for AE coarticulated vowels differed considerably from patterns found for citation materials. All four NG front, rounded vowels and PF [ø] were more similar to (fronted) back AE vowels, while most tokens of PF [y] were still more spectrally similar to front AE vowels. Front, unrounded NG [i:, ɪ, ɛ] and PF [i] were spectrally quite similar to their AE counterparts in context-independent analyses, and in some cases, better matches than in citation materials. In contrast, the majority of tokens of low and back NG and PF vowels, except for NG [ɑ:, ɔ], were more similar to centers-of-gravity for other AE vowels than their transcriptional counterparts, due to marked differences in the patterns of fronting and raising of these vowels across languages.

IV. DISCUSSION

This study compared within- and cross-language (dis)similarities in the spectral and temporal structure of vowels in three languages as a function of speech style (citation vs sentence utterances) and phonetic context (labial vs alveolar preceding/following consonants) in sentence utterances. Mid-syllable formant frequencies in citation utterances established canonical acoustic targets for vowels in the three languages, and could provide an acoustic basis for interpreting results of previous perceptual studies using these sorts of materials. Then spectral relationships among distributions of vowels produced in sentence materials were compared with canonical values to document (dis)similarities in patterns of coarticulatory variation across phonetic contexts within and across languages. The contribution of vocalic duration to the acoustic differentiation of vowels in the three languages was also assessed. The results revealed that contextual variation differed markedly across languages, and therefore, cross-language spectral similarities of vowel categories varied as a function of speech style and consonantal context. In this section, we review the results for citation and sentence materials and discuss how they address hypotheses about the role of vowel inventory make-up and size in determining the nature

of the language-specific constraints on coarticulatory under-shoot. We point out some limitations of the present study and how they may be addressed in future research. Lastly, we explore the implications of these findings for explaining/predicting patterns of perceived cross-language similarity and discrimination of NG and PF vowels by AE listeners.

A. Within- and cross-language variation of vowels in citation utterances

The results of within-language discriminant analyses showed that vowels in all three languages were well differentiated by mid-syllable formant frequencies and vocalic duration in citation utterances, despite (minor) speaker differences within gender groups (Table I). As expected, vocalic duration was most important in the differentiation of spectrally-similar NG vowels, while duration did not contribute at all to the differentiation of PF vowels. There were only minor differences across gender groups in the relative locations of vowels in a vowel space defined by F1/F2/F3 Bark values (see Fig. 2 and Appendices). For both gender groups, F2-F1 and F3-F2 (Bark) ratios differentiated vowels along the front-back dimension in all three languages: front vowels had large F2-F1 Bark differences (>5 Barks) and small F3-F2 Bark differences (<5 Barks) (except for AE [æ:]), while back vowels had the reverse relationship, small F2-F1 differences, and large F3-F2 differences (cf., Syrdal and Gopal, 1986). Front, rounded vowels in both NG and PF tended to have somewhat smaller F2-F1 and larger F3-F2 Bark difference values relative to front, unrounded vowels, but were still more “front” than “back” in terms of the 5-Bark difference criteria given above (except for NG [œ]). F1 values were related to vowel height, although so-called mid-high, short vowels in NG and AE were acoustically equal to or a bit lower (higher F1 values) than so-called mid, long vowels (see Appendices). Thus, phonetically, all three languages contrasted four levels of height spectrally, with further differentiation by duration of AE and NG vowels in the mid to mid-high region.

For cross-language comparisons, only mid-syllable formant values were used to assess similarity relationships by linear discriminant analysis, since absolute and relative durations of “similar” vowels differed across languages and previous research had shown that AE listeners perceptually assimilated NG vowels more on the basis of spectral similarity than temporal similarity (Strange *et al.*, 2005). Results indicated that the point vowels, which defined the overall dimensions of the vowel spaces for all three languages, were roughly equivalent in these citation utterances (Table II): high, front [i] was almost identical in location and distribution across languages; NG and AE [ɑ:] were also very similar, while PF [a] straddled AE [æ:, ɑ:]; NG and PF [u], while most similar to AE [u:], were clearly more back in citation context. PF front, rounded vowels were more similar to front than to back AE vowels, while NG front, rounded vowels straddled front and back AE categories, except for [y:], which was closer to front AE vowels in this context (Table III). Distributions of other NG and PF vowels, usually transcribed as the “same” or “similar” to AE categories, were

found to vary greatly in their spectral similarity to their AE counterparts (Table IV). Most, but not all, NG and PF mid-high, mid, and mid-low vowels appeared to be located more toward the high extreme in vowel space relative to AE vowels.

The acoustic dissimilarities for “similar” AE and PF vowels might account for discrimination difficulties by AE listeners of French vowel contrasts that also occur in their native language (Gottfried, 1984). However, the cross-language spectral similarities of front, rounded vowels to front AE vowels do not account well for why AE listeners perceptually assimilate NG front, rounded vowels to back native categories even in citation utterances (Strange *et al.*, 2004a, 2004b), nor why AE listeners tend to confuse front vs back, rounded French vowels more than they do front, unrounded vs rounded vowels (Gottfried, 1984; Polka, 1995). However, cross-language spectral similarities of these vowels in sentence materials may account for these perceptual assimilation and discrimination patterns.

B. Within- and cross-language variation of vowels in sentence materials

Distributions of coarticulated vowels, produced in multisyllabic nonsense words at a rate more closely approximating continuous speech, were evaluated with respect to how spectral and temporal parameters varied from canonical target values. Duration differences for NG and AE long/short vowels were somewhat reduced in sentence materials, relative to citation materials, especially for NG male utterances (Table VI). This replicated earlier studies of NG and AE male productions (Strange *et al.*, 2004a, 2005) and extended the findings to female speakers and to a more homogeneous dialect group of AE speakers. In general, L/S duration ratios were larger for NG than for AE speakers, as was expected, although average differences between language groups varied with consonantal context and with gender. Discriminant analyses of the sentence materials, with and without duration as an input parameter, indicated that for both NG and AE, duration contributed substantially to the acoustic differentiation of spectrally-adjacent vowels. In contrast, PF vowels were not differentiated by relative duration differences; absolute durations were substantially reduced in sentence materials relative to citation utterances, especially for female utterances. In sentence materials, PF vowels were more similar in duration to NG and AE short vowels than to long vowels. Thus, in interpreting cross-language differences in spectral target undershoot, PF vowels were compared with NG and AE short vowels.

In all three languages, the front vowels [i, e] (and [y] in NG and PF) varied least across speech styles and phonetic contexts, providing a reference for comparisons of other vowels within and across languages. It is hypothesized that the relative stability in the acoustic structure of these vowels may be due to the fact that the sides of the tongue are in contact with the teeth and alveolar ridge for their production. The patterns of coarticulatory variation of mid-syllable formant frequencies for other vowels differed markedly across languages (Figs. 3 and 4). In general, there was greater coar-

tulatory undershoot in alveolar contexts than in labial contexts for other vowels, presumably due to the influence of tongue-tip gestures for the alveolar consonants on tongue body gestures for the vowels. In labial contexts, there was some raising of the low, short vowel [a] in NG, with even greater raising of this vowel in PF. These shifts might be attributable to less lowering of the jaw in the more rapidly produced sentence utterances because of the necessity for the lips to fully close for labial consonants.

In alveolar contexts, NG and PF [a] were even further raised and also fronted in both NG and PF utterances; other short NG vowels [ɛ, ʏ, œ, ɔ, ʊ] and PF [ɔ, o, u] also shifted up and/or forward (Table V), while long NG low and back vowels shifted relatively little. Thus, in alveolar context, the PF vowel space decreased markedly in size but relative positions of vowels remained similar. In NG, the relative positions of long and short NG vowels changed due to greater shifts for the latter (Figs. 3 and 4). The smaller shifts of long vowels in NG would be predicted if undershoot were a function of syllable duration (i.e., the extent of temporal overlap of consonant and vowel commands at syllable midpoint). However, the pattern of coarticulatory change for AE long vowels was very different from the other two languages, with allophonic fronting of both long and short, back vowels [u, ʊ, ɔ:], and relatively small shifts for mid-low and low, short and long, back vowels. This changed the shape of the vowel space, with the back, high to mid, rounded vowels becoming more like front vowels (large F2-F1; small F3-F2).

These different patterns of contextual change for the same vowels in the same contexts cannot be accounted for by a simple model of target undershoot, but rather support the claim that there are language-specific constraints on coarticulation that serve to maintain acoustic (and perceptual) differentiation of vowels in each language (Diehl and Lindblom, 2004; Lindblom, 1990). Because AE does not contrast front and back, rounded vowels, the back vowels are “free” to vary more in alveolar context. In NG and PF, where back and front, rounded vowels are contrasted, the back vowels move forward relatively less, and the front, rounded vowels (except for the very front [y]) move forward in conjunction with the back vowels to maintain distinctiveness. In PF, with fewer (and more fronted) front, rounded vowels than in NG, the back vowels are fronted more in alveolar contexts (Table V). Thus, context-dependent discriminant analyses showed that, even in alveolar contexts, vowel distributions remained fairly well separated on the front-back dimension in all three languages (Table VII).

Context-dependent discriminant analyses suggested that there was greater spectral overlap on the height dimension for coarticulated vowels, especially in NG and AE. In NG, relative duration differences differentiated so-called mid-high, (short) vowels and mid, (long) vowels, despite the somewhat smaller L/S ratios than in citation utterances. However, high and mid (long) vowels were sometimes misclassified in both contexts. In AE, relative duration differences helped to differentiate spectrally similar front vowels, but not back vowels. Coarticulated PF vowels (as well as citation utterances) were not acoustically differentiated by

duration. Thus, in both contexts, there was some acoustic ambiguity of adjacent height categories across speakers within gender groups.

These language-specific patterns of coarticulatory change in mid-syllable formant frequencies cannot be fully accounted for by the hypothesized correlation between vowel inventory size and amount of articulatory/acoustic undershoot of vowel targets (Bohn, 2004; Steinlen, 2005). In general, there were greater coarticulatory shifts (relative to canonical targets) in PF (9 vowels) than in NG (14 vowels), supporting the hypothesis; however, considerations of vocalic duration must be included in this analysis. Differences in the extent of target undershoot between NG short vowels and PF vowels (which were closer to NG short vowels in vocalic duration) were much less striking than between NG long and PF short versions of the “same” vowels [o, u, ø] (Table V). In comparing AE (11 vowels) with the other two languages, the allophonic fronting of high to mid, back vowels should not be considered in correlations of inventory size with overall spectral change across languages. We suggest that the *phonological function* of vowel length (and other features such as lip rounding), as well as the overall size of the vowel inventory, must be considered in attempts to predict/explain differences across languages in the contextual variation of vowels. Further studies comparing vowels in Russian, Japanese, and Spanish are underway in our laboratory to extend this investigation to languages with relatively small vowel inventories (5–6 spectral categories) that differ in the phonological function of vowel length (e.g., Law II *et al.*, 2006). Research on South German, which is reported to distinguish tense/lax vowel pairs almost exclusively by vocalic duration (i.e., contrasts only three heights) would also provide valuable insights about language-specific coarticulatory constraints, as they relate to inventory size and makeup.

C. Limitations of the study

Several limitations of the study should be addressed in future research. While the results for male and female corpora showed fairly similar patterns of within- and cross-language (dis)similarities, some gender differences were noted in the sentence materials. However, the number of speakers of each gender within each language was small; thus, differences may not be reliable. No inferential statistical comparisons of gender differences in shifts in mid-syllable formant frequencies or durations were computed, given the low power of such tests with these small sample sizes.⁶ To follow up on whether there are significant differences in the phonetic realization of vowels across genders, more speakers of each gender must be tested.

Conclusions about within- and cross-language (dis)similarities were based on quantitative analyses of distributions of utterances, summing over multiple tokens contributed by a small number of speakers selected for their dialect homogeneity (see Hillenbrand *et al.*, 2001). Generalization to speakers of other dialects of these languages should not be made. For example, conclusions about the relative distinctiveness of spectral and temporal parameters in Parisian French and North German do not apply to Swiss French (Miller *et al.*,

2000) and Southern dialects of German. While this narrows the scope of the conclusions, it is our opinion that dialect-specificity is necessary when investigating the phonetic realization of vowel categories unless it can be established, *a priori*, that vowels do not differ phonetically across the dialects represented in the sample.

In these analyses, only mid-syllable formant frequencies and vocalic durations were considered. Thus, other important acoustic information for specifying vowel gestures was not included (Strange, 1989). Future studies in which formant trajectories throughout the entire vocalic nuclei are tracked may uncover important differences in coarticulatory patterns across languages. For instance, NG speakers might back the place-of-constriction of alveolar consonants in the context of back vowels more than PF speakers do; these contextual variations in consonant production would be indicated by F2/F3 formant onset and offset frequencies.⁷ Second, in addition to vocalic duration differences in short (lax) and long (tense) vowels in NG and AE, these vowels are also distinguished by differences in the temporal location of F1 maximum (Stack *et al.*, 2006), and the shape of F1 temporal trajectories (Huang, 1985); tense vowels have relative symmetrical (and rapid) onglides and offglides, while lax vowels show asymmetrical temporal trajectories. For AE lax vowels (including [æ:]), F1 onglides are shorter than offglides (Peterson and Lehiste, 1960); for NG lax vowels, F1 onglides are longer than offglides (Strange and Bohn, 1998).⁸ To our knowledge, there are no data on the temporal properties of F1 trajectories for PF vowels. The perceptual relevance of these (dis)similarities across languages in the dynamic spectral information should be explored.

D. Implications for studies of perceived phonetic similarity and discrimination of vowels

A major motivation for comparing the acoustic structure of NG, PF, and AE vowels in this study arose from previous research investigating the difficulties that native speakers of AE have in discriminating French (Gottfried, 1984; Levy, 2004; Levy and Strange, in press) and German vowels (Polka, 1995) and the patterns of perceived similarity of NG and PF vowels to AE categories (Strange, 2007; Strange *et al.*, 2004a, 2004b; 2005). The comparisons of the patterns of acoustic variation of vowels in the three languages reported here provide some insights into the patterns of perception by naïve listeners and AE speaking L2 learners of German and French reported in the literature.

First, let us consider front, rounded vowels as they relate to AE categories. While both PF and (long) NG vowels in citation utterances are more similar acoustically to front than to back AE vowel categories, AE listeners routinely categorize them as more perceptually similar to back native categories and have more difficulty discriminating them from back, rounded vowels than from front, unrounded vowels. That is, acoustic similarity of citation utterances does *not* predict perceived similarity nor discrimination difficulty in categorial (name identity) tasks. This can be accounted for by considering the distributional characteristics of native vowel categories in continuous speech contexts with which the AE

listeners have had experience. As the data reported here show, AE front vowels vary little across prosodic and phonetic contexts, whereas AE back vowels vary extensively on the front-back dimension. Thus, NG and PF front, rounded vowels are quite deviant from any exemplars of front vowels in AE, while they are very similar to some allophones of back AE vowels. Thus, in judging perceived similarity and in discriminating non-native contrasts, front and back, rounded vowels are phonetic variants of the *same* phonological category in AE, whereas front, unrounded and rounded vowels are phonetic variants of *different* phonological categories in AE. The results of perceptual assimilation and discrimination studies support the conclusion that this experience with phonetic variation in phonological categories determines AE listeners' performance in perceptual assimilation and categorial discrimination tasks (cf., Best, 1995; Best and Tyler, 2007), and predicts L2 learners' perception (and production) problems in differentiating these categories in German and French.

Cross-language differences in how distributions of NG, PF, and AE vowels are arrayed in the height dimension, and how duration is or is not used to differentiate spectrally-adjacent categories may also account for why some vowel pairs in German and French that are also contrasted in AE are nevertheless problematic. Confusions between high and mid NG and PF vowels (e.g., [i/e, y/ø, u/o]) that do not differ in duration can be expected. Other height contrasts (e.g., [e/ɛ] problems reported by Gottfried 1984) may be more or less problematic depending upon the context in which they are produced and presented.

In a forthcoming paper, results of perceptual assimilation tests of PF and NG vowels by inexperienced AE listeners are presented. In two separate studies, tests of vowels produced/presented in citation disyllables and in sentences

with varying consonant contexts revealed differences across languages and speaking styles in the native AE categories to which the PF and NG vowels were assimilated and in the judged category goodness of the vowels. Some, but not all, of these variations in perceived cross-language similarity could be accounted for by contextual differences in spectral similarity reported in the present study.

In general, the results reported here indicate that predictions about perceptual difficulties with vowels by L2 learners, based on acoustic comparisons and perceptual studies using citation utterances may not be generalizable to continuous speech contexts. If our goal is to develop theories of non-native and L2 perception and production in "real world" communication situations, future research on vowels must employ speech materials that more closely resemble the speech input of the language learning environment. The sentence materials used here are still instances of "lab speech," and no doubt differ markedly from truly spontaneous speech in a conversational setting. However, these read sentence materials may approach the speech style that is used in a formal language-learning environment. The phonetic variation of vowels in sentence materials reported here help to explain patterns of difficulty found in the laboratory, and it is hoped, in the language learning environment of adult learners of non-native languages.

ACKNOWLEDGMENTS

The research reported here was funded by a grant to the first author (NIH NIDCD-RO1-00323). Thanks to the following people for help on the data analysis and the manuscript: James J. Jenkins, Franzo F. Law II; Yana Gilichinskaya. Portions of this research were reported in a poster at ASA (Strange *et al.*, 2002) and in a chapter by Strange (2007).

APPENDIX A: NORTH GERMAN CITATION MATERIALS

Mean F3-F2 and F2-F1 Bark differences, F1 (Barks), and syllable durations (ms); male (M) and female (F) data are listed separately.

NG vowels	F3-F2 (Barks)		F2-F1 (Barks)		F1 (Barks)		Duration (ms)	
	M	F	M	F	M	F	M	F
i:	2.3	1.6	10.9	11.9	2.3	3.1	118	112
e:	2.0	0.9	10.1	10.3	3.2	4.5	150	137
ɪ	2.4	1.4	8.4	9.6	3.6	4.6	64	80
ɛ	2.7	2.0	6.2	6.8	5.5	6.6	79	92
y:	1.4	2.0	9.2	9.3	2.6	3.1	132	122
ø:	2.6	3.1	7.2	6.9	3.5	4.4	151	140
ʏ	3.2	3.2	6.7	6.8	3.6	4.5	69	85
œ	3.6	3.7	4.9	4.8	5.3	6.4	84	91
u:	8.0	8.0	3.3	3.0	2.7	3.7	136	114
o:	8.6	8.1	2.3	2.4	3.7	4.5	152	128
ʊ	7.1	7.1	3.3	3.3	3.8	4.6	70	80
ɔ	6.1	5.6	2.7	2.6	5.7	6.8	82	92
ɑ:	4.9	4.1	2.8	2.5	6.8	8.5	173	147
a	4.9	4.2	3.1	2.7	6.4	8.3	7.3	90

APPENDIX B: PARISIAN FRENCH CITATION MATERIALS

Mean F3-F2 and F2-F1 Bark differences, F1 (Barks), and syllable durations (ms); male (M) and female (F) data are listed separately.

PF vowels	F3-F2 (Barks)		F2-F1 (Barks)		F1 (Barks)		Duration (ms)	
	M	F	M	F	M	F	M	F
i	2.7	2.3	10.7	12.2	2.6	2.4	108	143
e	2.1	1.5	9.6	10.7	3.4	3.7	87	107
ɛ	2.3	2.1	8.0	7.9	4.5	5.8	106	172
y	2.0	0.9	9.8	11.1	2.7	2.6	113	155
ø	3.3	3.3	7.0	7.3	3.7	4.5	136	173
u	6.6	8.0	4.5	4.2	2.7	2.9	106	167
o	7.2	8.1	3.2	3.6	3.8	3.8	126	183
ɔ	5.0	5.7	3.9	3.9	5.2	5.8	113	163
a	4.1	3.8	3.6	3.3	6.7	8.1	125	168

APPENDIX C: AMERICAN ENGLISH CITATION MATERIALS

Mean F3-F2 and F2-F1 Bark differences, F1 (Barks), and syllable durations (ms); male (M) and female (F) data are listed separately.

AE vowels	F3-F2 (Barks)		F2-F1 (Barks)		F1 (Barks)		Duration (ms)	
	M	F	M	F	M	F	M	F
i:	2.0	1.0	10.9	11.7	2.7	3.2	99	118
e:	1.6	0.9	9.4	10.3	3.9	4.5	125	163
ɪ	2.1	1.9	7.6	7.8	4.6	5.5	76	98
ɛ	3.0	2.5	5.2	5.4	5.8	7.3	85	101
æ:	3.4	3.0	3.8	3.8	7.0	8.5	126	153
u:	6.3	6.9	4.2	4.2	3.2	4.0	98	131
o:	6.1	6.3	3.5	3.5	4.4	5.4	120	152
ʊ	5.2	5.1	4.5	4.1	4.5	6.0	87	102
ɔ:	5.9	5.5	2.7	2.4	5.8	7.2	153	172
ɑ:	4.8	4.6	2.9	2.4	6.7	8.3	128	163
ɑ:	4.8	4.6	3.6	3.2	6.0	7.4	82	102

¹French nasal vowels /ɛ̃, œ̃, ã, õ/ were not considered in this study because their acoustic structure varies as a function of coupling with the nasal tract. Nasal vowels are perceived as quite distinct from oral vowels because of these nasal resonances.

²In addition to these conditions, speakers also recorded sentence materials (dVt context) in which sentence prosody (target nonsense word in narrow focus vs post-focus position) and the speaking rate (normal vs rapid rate) varied. Results of acoustic analyses of these materials will be presented in a separate paper.

³These were not distributed evenly over vowels: the largest problems were for F3 in NG [ʊr, ʊ] and PF [u], which had very low amplitude, and F2 and F3 for PF [y].

⁴For AE vowels, formant trajectories throughout the middle half of the syllables (25%–50%–75% Bark values) for individual speakers were also plotted to determine whether there were systematic patterns of VISC or diphthongization common to all speakers. Results yielded large inter-speaker differences in both the direction and extent of formant movement for several vowels, including the so-called diphthongized [e:, o:]. Therefore, in this and all subsequent analyses, trajectory information was not included in within-language or cross-language comparisons.

⁵Discriminant analyses using F1, F2-F1, F3-F2 Bark values were also computed. However, since the pattern of results was very similar to analyses with F1/F2/F3 Bark values as independent parameters, only the latter are reported here.

⁶With respect to tests of mean differences, it is our opinion that comparison of average values is less informative than are discriminative analyses, regardless of sample sizes. Significant differences in central tendencies only establish the reliability of those differences, not their size or perceptual relevance. The discriminative analyses establish relationships among distributions of tokens, and are thus more akin to estimates of d prime (d') or effect size.

⁷The first author wishes to thank Kenneth N. Stevens for suggesting that this may be a factor in cross-language differences in vowel coarticulation.

⁸These spectro-temporal patterns of formant change are not the same as the VISC reported by Nearey and colleagues which occur in Canadian and American CVC syllables and isolated vowels (Nearey and Assmann, 1986). For the AE vowels analyzed here, formant change over the middle half of the syllables, even in citation utterances, was not systematic across speakers.

Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language experience in second language speech learning: In honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 13–34.

- Best, C. T., Hallé, P. A., Bohn, O.-S., and Faber, A. (2003). "Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages," in *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M. J. Sale, D. Rescasens, and J. Romero (Causal Productions, Barcelona), pp. 2889–2892.
- Bohn, O.-S. (2004). "How to organize a fairly large vowel inventory: the vowels of Fering (North Frisian)," *J. Int. Phonetic Assoc.* **34**, 161–173.
- Diehl, R. L., and Lindblom, B. (2004). "Explaining the structure of feature and phoneme inventories: The role of auditory distinctiveness," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York), pp. 101–162.
- Flege, J. E. (1987). "The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification," *J. Phonetics* **15**, 47–65.
- Flege, J. E. (1995). "Second language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–277.
- Fourakis, M. (1991). "Tempo, stress, and vowel reduction in American English," *J. Acoust. Soc. Am.* **90**, 1816–1827.
- Gottfried, T. L. (1984). "Effects of consonant context on the perception of French vowels," *J. Phonetics* **12**, 91–114.
- Gottfried, T. L., and Beddor, P. S. (1988). "Perception of temporal and spectral information in French vowels," *Lang Speech* **32**, 57–75.
- Hay, J. F., Sato, M., Coren, A. E., Moran, C. L., and Diehl, R. L. (2006). "Enhanced contrast for vowels in utterance focus: A cross-language study," *J. Acoust. Soc. Am.* **119**, 3022–3033.
- Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (2001). "Effects of consonant environment on vowel formant patterns," *J. Acoust. Soc. Am.* **109**, 748–763.
- Huang, C. (1985). "Perceptual correlates of the tense/lax distinction in General American English," Master's thesis, MIT Cambridge, MA.
- Klecka, W. R. (1980). *Discriminant Analysis* (Sage, Newbury Park, CA).
- Law II, F. F., Gilichinskaya, Y. D., Ito, K., Hisagi, M., Berkowitz, S., Sperbeck, M. N., Monteleone, M., and Strange, W. (2006). "Temporal and spectral variability of vowels within and across languages with small vowel inventories: Russian, Japanese, and Spanish," *J. Acoust. Soc. Am.* **20**, Pt. 2, 3296.
- Levy, E. S. (2004). "Effects of language experience and consonantal context on perception of French front rounded vowels by adult American English learners of French." Ph.D. dissertation, City University of New York-Graduate School and University Center.
- Levy, E. S., and Strange, W., "Perception of French vowels by American English adults with and without French language experience," *J. Phonetics*.
- Lindblom, B. (1963). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* **35**, 1773–1781.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.
- Miller, J. L., and Grosjean, F. (1997). "Dialect effects in vowel perception: the role of temporal information in French," *Lang Speech* **40**, 277–288.
- Miller, J. L., Mondini, M., Grosjean, F., and Dommergues, J.-Y. (2000). "Dialect differences in the temporal characteristics of vowels: a comparison of standard (Parisian) and Swiss French," *J. Acoust. Soc. Am.* **108**, 2507.
- Moon, S.-J., and Lindblom, B. (1994). "Interaction between duration, context, and speaking style in English stressed vowels," *J. Acoust. Soc. Am.* **96**, 40–55.
- Nearey, T. M., and Assmann, P. F. (1986). "Modeling the role of inherent spectral change in vowel identification," *J. Acoust. Soc. Am.* **80**, 1297–1307.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Peterson, G. E., and Lehiste, I. (1960). "Duration of syllabic nuclei in English," *J. Acoust. Soc. Am.* **32**, 693–703.
- Polka, L. (1995). "Linguistic influences in adult perception of non-native vowel contrasts," *J. Acoust. Soc. Am.* **97**, 1286–1296.
- Stack, J. W., Strange, W., Jenkins, J. J., Clarke III, W. D., and Trent, S. A. (2006). "Perceptual invariance of coarticulated vowels over variations in speaking rate," *J. Acoust. Soc. Am.* **119**, 2394–2405.
- Steinlen, A. K. (2005). "*The influence of consonants on native and non-native vowel production: A cross-linguistic study*" (Gunter Narr, Tübingen).
- Stevens, K. N., and House, A. S. (1963). "Perturbation of vowel articulations by consonantal context: An acoustical study," *J. Speech Hear. Res.* **6**, 111–128.
- Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Öhman, S. (1969). "Cross-language study of vowel perception," *Lang Speech* **12**, 1–23.
- Strange, W. (1989). "Dynamic specification of coarticulated vowels spoken in sentence context," *J. Acoust. Soc. Am.* **85**, 2135–2153.
- Strange, W. (2007). "Cross-language phonetic similarity of vowels: theoretical and methodological issues," in *Language experience in second language speech learning: In honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 35–55.
- Strange, W., and Bohn, O.-S. (1998). "Dynamic specification of coarticulated German vowels: perceptual and acoustic studies," *J. Acoust. Soc. Am.* **104**, 488–504.
- Strange, W., Bohn, O.-S., Trent, S. A., and Nishi, K. (2004a). "Acoustic and perceptual similarity of North German and American English vowels," *J. Acoust. Soc. Am.* **115**, 1791–1807.
- Strange, W., Bohn, O.-S., Nishi, K., and Trent, S. A. (2005). "Contextual variation in the acoustic and perceptual similarity of North German and American English vowels," *J. Acoust. Soc. Am.* **118**, 1751–1762.
- Strange, W., Levy, E. S., and Lehnhoff, Jr., R. J. (2004b). "Perceptual assimilation of French and German vowels by American English monolinguals: Acoustic similarity does not predict perceptual similarity," *J. Acoust. Soc. Am.* **115**, 2606.
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., and Nishi, K. (2002). "Within- and across-language acoustic variability of vowels spoken in different phonetic and prosodic contexts: American English, North German, and Parisian French," *J. Acoust. Soc. Am.* **112**, 2384.
- Syrdal, A. K., and Gopal, H. S. (1986). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," *J. Acoust. Soc. Am.* **79**, 1086–1100.
- van Son, R. J. J. H., and Pols, L. D. W. (1992). "Formant frequencies of Dutch vowels in a text, read at normal and fast rate," *J. Acoust. Soc. Am.* **88**, 1683–1693.