JSLHR

Article

Comparing Identification of Standardized and Regionally Valid Vowels

Richard Wright[a] and Pamela Souza[b]

Purpose: In perception studies, it is common to use vowel stimuli from standardized recordings or synthetic stimuli created using values from well-known published research. Although the use of standardized stimuli is convenient, unconsidered dialect and regional accent differences may introduce confounding effects. The goal of this study was to examine the effect of regional accent variation on vowel identification.

Method: The authors analyzed formant values of 8 monophthong vowels produced by 12 talkers from the region where the research took place and compared them with standardized vowels. Fifteen listeners with normal hearing identified synthesized vowels presented in varying levels of noise and at varying spectral distances from the local-dialect values.

Results: Acoustically, local vowels differed from standardized vowels, and distance varied across vowels. Perceptually, there was a robust effect of accent similarity such that identification was reduced for vowels at greater distances from local values.

Conclusions: Researchers and clinicians should take care in choosing stimuli for perception experiments. It is recommended that regionally validated vowels be used instead of relying on standardized vowels in vowel perception tasks.

Key Words: dialect, identification, vowels

This study consists of two experiments with two related goals: to determine (a) the degree to which vowels in the Pacific Northwest region matched General American English vowels and (b) the degree to which acoustic dissimilarity has an effect on vowel identification accuracy. Although any number of dimensions of the speech signal may be studied to examine differences across regional dialect, we chose to make vowels the focus of this study for several reasons. They are high-intensity quasi-periodic components of the speech signal; moreover, they occupy lower frequencies (below 3 kHz) compared with many consonants that have aperiodic cues distributed in the higher frequencies (above 3 kHz). The frequency distribution and the high intensity of vowels make them important perceptual toeholds for hearers who are listening under less-than-optimal conditions, such as in noise or with hearing loss. Vowels are also important because their formants carry information about not only the vowel's quality itself but also the flanking consonants through their formant transitions. Again, this may be particularly true under conditions of multiple distortions. Recent work by Webster (2002) indicates that as noise increases, listeners shift their weighting of cues away from the aperiodic cues associated with consonants, increasing their reliance on formant transitions. Accordingly, vowels represent a methodologically important object of study (e.g., Kewley-Port, Burkle, & Lee, 2007; Nishi & Kewley-Port, 2008).

The exigencies of perceptual research and especially clinical testing encourage researchers and clinicians to use stimuli that have well-known properties—in particular, stimuli that have been widely normed. These stimuli are sometimes described as *General American English*. In North America, this has led to the widespread use of a relatively small set of prerecorded or synthetic stimuli drawn from well-known published data. Because the standard stimuli are taken from a small number of specific geographic regions at specific points in time, there is often a mismatch between the regional accent in the stimuli and the perceiver's regional accent. This mismatch introduces a potential problem: Even relatively subtle differences in accent may introduce a stimulus-goodness effect that varies by laboratory or clinic,

---

[a] University of Washington, Seattle
[b] Northwestern University, Chicago, IL
Correspondence to Pamela Souza: p-souza@northwestern.edu

Editor: Robert Schlauch
Associate Editor: Mark Hedrick
Received October 4, 2010
Revision received March 14, 2011
Accepted June 3, 2011
DOI: 10.1044/1092-4388(2011/10-0278)

depending on geographic location, and that may even vary by listener if the target population is dialectally diverse.

Regional dialect and accent effects are not uniform across the different regions in North America; neither are they uniform across speech sounds. Some accents are more similar to the standard stimuli than others, and some speech sounds vary more by region than others (e.g., Clopper, Pisoni, & de Jong, 2005; Labov, Ash, & Boberg, 2006). For example, in several dialects spoken in New England and in the northern cities, urban areas on the southern shores of the Great Lakes—such as Chicago, Illinois; Detroit, Michigan; Cleveland, Ohio; and Buffalo, New York—there is a phonemic contrast between an open-mid vowel /ɔ/ and a low-central vowel /a/ (as in the words *caught* and *cot,* respectively), whereas in much of the West and the South, this contrast is replaced with a single low-back vowel /ɑ/ (e.g., Clopper et al., 2005; Eckert, 1989; Labov et al., 2006). Moreover, different dialects typically have distinct allophonic processes for otherwise similar phonemes. Take, for example, a process often referred to as *Canadian raising*: the diphthong phoneme /aɪ/ (as in “ride” [ɹaɪd] or “rise” [ɹaɪz]) has a “raised” allophone [ʌɪ] before all voiceless coda consonants (as in *write* [ɹʌɪt] or *rice* [ɹʌɪs]; e.g., Vance, 1987). This allophonic variant occurs in several dialects spoken in the northern and eastern United States but not in other dialects spoken in the South and West. Thus, even with identical phonemic inventories, there may be context-specific (allophonic) differences in pronunciations of particular words or syllables.

Although not all regions have been fully documented, existing documentation indicates that regional dialects differ not only in the number of vowel contrasts and phonetic realization of vowel allophones but also in the acoustic values that phonemes have even if they are conventionally transcribed using the same phonetic symbol. Although this last point may seem obvious to researchers who are familiar with regional variation and transcription conventions, the symbols that are used to represent vowel categories are often taken as true values by researchers in clinical and psychological settings who are not trained in regional variation. As a result, it is common to assume that standardized vowels should be used in all regions of the country where the phonetic symbols match the symbols used in the standard descriptions. For example, whereas the symbol /u/ is used to represent a vowel that is present as a phoneme in all dialects of North American English, the acoustic realization of this vowel varies from a high-back variant in Wisconsin to a high-central variant in southern California and parts of the Deep South (e.g., Clopper et al., 2005; Hagiwara, 1997; Jacewicz, Fox, & Salmons, 2007; Labov et al., 2006). In a study of six regional vowel systems, Clopper et al. (2005) found that each region had its own realization of the vowel system even though the vowels are typically represented with the same symbols and phonetic descriptors. Their findings led them to conclude that vowels are best characterized in terms of regional systems rather than in terms of General American English, a finding that echoed Hagiwara’s (1997) observation about southern Californian English and Hillenbrand, Getty, Clark, and Wheeler’s (1995) observation about regional effects in their study of southern Michigan speakers.

These findings have potentially serious implications for researchers who use speech stimuli in perception experiments or as comparison points for speech production. In the absence of acoustic validation through regional sampling, there is no guarantee that a particular phoneme’s acoustic realization in North American English will be the same from one region to the next. Ignoring regional dialect and accent variation may result in regional asymmetries in category goodness of stimuli across listeners in perception studies.

Phonetic differences in stimuli may seem relatively unimportant if the task is not tapping into low-level phonetic perception, but there is evidence that accented speech may impose a processing delay or a cognitive load. For example, Clarke and Garrett (2004) found that initial exposure to foreign-accented speech introduced a temporary perceptual-processing deficit. More relevant to the current research, regional accent differences may have a cascading effect that percolates up to higher level tasks. Adank and McQueen (2007) conducted a noun-animacy (animate vs. inanimate) decision task in which listeners were presented with auditory stimuli in a familiar accent and an unfamiliar accent. Subjects’ response times were slower for words presented in the unfamiliar accent. The effect persisted even when listeners were exposed to 20 min of speech in the unfamiliar accent prior to the animacy decision task, indicating a lasting effect for unfamiliar accents. Similarly, Floccia, Butler, Goslin, and Ellis (2009) found that unfamiliar accents in certain tasks impose a processing delay that is long lasting; it continues after intelligibility scores reach ceiling. To compound the problem, prosody, intonation, syntax, and lexical choice vary across regions (e.g. Grabe, 2004); therefore, using sentence-length stimuli does not necessarily solve the potential problems at the phone level and may even make them worse. In a study of cross-dialect intelligibility in noise, Clopper and Bradlow (2009) found that dialects that were more similar to General American English, which included New England, the West, and the Midland regions, were more intelligible than regional dialects that were more distant from General American English, including mid-Atlantic, northern, and southern regions. Although not all dialectal regions were equally represented—so true regional effects were not fully probed—on the whole, Clopper and Bradlow's results indicated that there is a negative effect on intelligibility of mismatch between the listener's dialect and the dialect in the stimuli in noisy listening environments.

How concerned should researchers be about the effects of regional accents on category goodness, perceptual processing, and higher level processing in perceptual research? After all, there appears to be counterevidence in that listeners seem to be able to adapt to foreign-accented speech (e.g., Clarke & Garrett, 2004; Ferguson, Jongman, Sereno, & Keum, 2010; Munro & Derwing, 1995a, 1995b). Moreover, the native dialect effects that have been observed for speech perception and processing for Dutch (Adank & McQueen, 2007), French (Dufour, Nguyen, & Frauenfelder, 2007), and English (Cutler, Smith, & Cooper, 2005; Evans & Iverson, 2007; Floccia et al., 2009) involved dialect differences that were larger than those typically observed within North America. This uncertainty about regional effects has led to a general trend in North American research to treat a common set of vowel values as "standard" North American English in perception experiments and in clinical studies. No study has directly tested the effect of dialect vowel distance (as measured using first and second formant [F1 and F2, respectively] differences between dialects) on identification accuracy; therefore, whether distance has a negative impact on vowel identification is unknown.

To test whether vowel differences that are typical in North American English interfere with vowel identification, we conducted two experiments. In the first, we investigated vowels of the region where the research took place, the Pacific Northwest (PNW), to establish whether their formant values differ from two sets of formant values of vowels that are commonly referred to as General American English: (a) Peterson and Barney (1952; PB) and (b) Hillenbrand et al. (1995; HGCW). In the second experiment, we conducted a perceptual identification task using three synthesized vowel sets that differed in spectral distance (F1 and F2 differences in a Euclidean space) from the vowels of the subjects' regional dialect (0 = *PNW identical,* 1 = *PNW similar,* 2 = *PNW dissimilar*) to test the effect of vowel dissimilarity distance on identification accuracy.

Experiment 1: Comparing PNW Vowels With Standard Vowels

Although all dimensions of language vary across regional dialects, vowels are particularly well documented in English and therefore most easily compared. Moreover, there is a widely used comparison metric for vowels: F1 and F2 are typically measured at midpoint or steady state. This measure was used in Peterson and Barney's (1952) study as well as Hillenbrand et al.'s (1995) research and other recent vowel studies (e.g., Clopper et al., 2005; Hagiwara, 1997). We predicted that there would be study effects such that both PB and HGCW vowels would differ from the PNW vowels. We also predicted that there would be Vowel × Study interactions, such that PB and HGCW vowels would differ from PNW vowels in study-specific ways.

Method

Talkers

Six adult men and six adult women participated in vowel recordings. All of the talkers had grown up in the PNW (Washington, Idaho, Oregon) and were monolingual speakers of English. Talkers who had lived elsewhere or had significant experience with other languages were excluded from participation. The PNW was defined broadly to include the amount of local variation that is likely to occur in an experimental or clinical population.[1] None of the talkers had a history of speech therapy, and all had normal hearing, defined as pure-tone thresholds of 20 dB HL or better (see American National Standards Institute, 2004) at octave frequencies between 0.25 kHz and 8 kHz, bilaterally. All procedures were reviewed and approved by the local institutional review board, and talkers were reimbursed for their time.

Recording Procedure

Words containing eight monphthong vowels (/i, ɪ, e, ɛ, æ, ɑ, ʊ, u/) were recorded using randomized word lists. The mid-back vowel /o/ was excluded because it is a diphthong in the PNW region (Ingle et al., 2005). The mid-front vowel /e/ was included for qualitative comparison but excluded from the statistics because it was absent in the PB study. Talkers were seated in a double-walled sound-attenuated chamber during the recordings. All vowels except for /e/ and /ɑ/ were read in the /hVd/ context from a word list following the procedure of Hillenbrand et al. (1995). The vowels /e/ and /ɑ/ were spoken without the preceding /h/ in the words *aid* and *odd* due to unfamiliarity of the words *hayed* and *hawed*. The eight words were read in five randomizations to control for list effects on pronunciation, thus minimizing the need for a carrier phrase. All talkers were instructed to read the list of words at a natural pace and vocal intensity. Vocal level was monitored with a volume unit meter to ensure sufficient output levels without clipping. We used a Tucker–Davis Technologies System 2 with an

---

[1] Other studies of the PNW accent have defined specific neighborhood regions of the Northwest and/or specific sounds (e.g., Ingle, Wright, & Wassink, 2005; Wassink, Squizzero, Schirra, & Conn, 2009).

AP2 sound card and a Shure BG 1.0 omnidirectional microphone for all of the recordings. Four talkers were recorded direct to disc at a rate of 44.1 kHz. The remaining eight talkers were recorded at a rate of 22.05 kHz. All recordings were quantized at 16 bits and down-sampled to 11.025 kHz prior to signal processing and acoustic analysis.

Spectral Analysis

We selected one representative token of each vowel for each talker on the basis of two criteria: (a) recording fidelity and clarity of each subject's voice without hoarseness, pitch breaks, or other disfluencies, and (b) visual inspection of the vowel for F1 and F2 steady states and accompanying pitch steady state. We determined formant steady state using a wideband spectrogram and accompanying linear predictive coding (LPC) formant track with 12 coefficients. We determined pitch steady state using an autocorrelation pitch track with a 25-ms window overlaid on a narrow-band spectrogram. A formant was considered to be steady state if a straight line could be traced through the middle 50 ms of the vowel and the pitch remained constant over the same section.

Once the vowel was selected and the steady-state portion identified, the first four formants (F1, F2, F3, F4) were estimated at the center of the steady state. The F1 through F4 values and recordings had been used in a separate study on hearing aid compression (Bor, Souza, & Wright, 2008). In the current study, we used only F1 and F2. To minimize the risk of error, formant measurements were taken from an LPC spectrum overlaid on a fast Fourier transform power spectrum with a sample window of 128 points and visually compared with its broadband spectrogram. In the event that there were LPC errors for a particular talker, the number of filter coefficients (poles) was adjusted up or down for that entire talker's set of vowel measures. The sample window of the LPC was 25 ms, and the number of coefficients ranged between 10 and 12, depending on the talker.

Results

Vowels from the PNW study were compared with HGCW and PB vowels using the F1 and F2 values. Because we had a reasonably large sample (six men and six women), we were able to make comparisons within gender rather than normalizing the data, thus preserving vowel space shape and individual vowel variability. Because there were grossly different sample sizes between studies, we used a random sample of six speakers from each of the larger studies (PB, HGCW). To ensure that there were no sampling artifacts, we plotted the mean of the resulting sample within a 95% confidence interval ellipse representing all data within gender for adults. In all cases, our sample fell near the center of the F1 × F2 ellipse, indicating that our sample was representative of each study while still retaining variability within vowel. The results of the within-gender sample were submitted to a series of analyses of variance (ANOVAs) with F1 and F2 as dependent variables, and study (PNW, HGCW, or PB) and vowel (i, ɪ, ɛ, æ, ɑ, ʊ, u) as independent variables. ANOVA results are presented in Table 1 (male vowels) and 2 (female vowels).

The results of the ANOVAs indicated the expected reliable effect of vowel on F1 and F2 for both men and women, indicating that in each study, the vowels were reliably separated on both dimensions. More interesting is that there was an effect of study on F2 for men and a Study × Vowel interaction for F1 and F2 in the data for both the men and women. The interactions indicate study-specific differences for individual vowels. To probe which vowels were contributing to the interactions, we submitted the formant values to a series of Bonferroni–Dunn post hoc *t* tests with an alpha of .05 (corrected to .0167). For men's PNW F1 values, the vowel [æ] was different from its HGCW counterpart, but none were different from their PB counterparts. For men's PNW F2 values, the vowels [æ], [a], [ʊ], and [u] were all different from their HGCW counterparts, and the vowels [ʊ] and [u] were different from their PB counterparts. For women's PNW F1 values, the vowels [i], [æ], [ɑ], and [ʊ] were all different from their HGCW counterparts, but none were different from their PB counterparts. For women's PNW F2 values, the vowels [ɛ], [æ], [ɑ], [ʊ], and [u] were all reliably different from their HGCW counterparts, and the vowels [i], [ɛ], [ʊ], and [u] were all reliably different from their PB counterparts.

To illustrate the Study × Vowel effects, we plotted men's and women's vowel means in Figures 1 and 2, respectively, with the PNW vowel plots overlaid on HGCW and PB vowels. Notable differences include a lowered (higher F1) and backed (lower F2) PNW /æ/ relative to

Table 1. Analysis of variance (ANOVA) results for men's vowels.

Formant	*df*	*F*	*p*
F1			
Study	2	1.74	.1807
Vowel	6	287.23	< .001
Study × Vowel	12	3.80	< .001
F2			
Study	2	15.13	< .001
Vowel	6	277.38	< .001
Study × Vowel	12	7.07	< .001

Table 2. ANOVA results for women's vowels.

Formant	*df*	*F*	*p*
F1			
Study	2	1.72	.184
Vowel	6	154.02	< .001
Study × Vowel	12	9.82	< .001
F2			
Study	2	2.73	.069
Vowel	6	317.42	< .001
Study × Vowel	12	11.30	< .001

the HGCW counterpart, a raised (lower F1) and backed (lower F2) PNW /ɑ/ relative to the HGCW counterpart, and a fronted (higher F2) PNW /ʊ/ and /u/ relative to the PB and HGCW counterparts. The PNW /e/ also appears slightly raised (lower F1) and fronted (higher F2) relative to the HGCW counterpart. Formant values are summarized in Table 3.

To investigate the relative differences further, we calculated Euclidean distances from each PNW vowel's F1 × F2 point in the vowel space to its gender-matched counterpart from the PB and HGCW values.[2] We used Equation 1 to calculate distance:

$$s = \sqrt{(F1_p - F1_i)^2 + (F2_p - F2_i)^2} \qquad (1)$$

Throughout this study, we used this formula to calculate distance. In the formula, s is the distance between points in a two-dimensional Euclidian vowel space defined by F2 on the x axis and F1 on the y axis, and where $\mathrm{F1}_p$ is an individual PNW speaker's F1 value for a particular vowel and $\mathrm{F1}_i$ is a comparison value for the equivalent PB or HGCW vowel F1 mean, and where $\mathrm{F2}_p$ is an individual PNW speaker's F2 value for a particular vowel and $\mathrm{F2}_i$ is a comparison value for the equivalent PB or HGCW vowel F2 mean. Because we were interested in relating these differences to perceptual effects, we also calculated the Euclidean distances in Bark using the formula published in Traunmüller (1990).

The results of the Euclidean distance measures are summarized in Table 4, which displays the mean distance and *SD*s by vowel and by study in Hz and in Bark, respectively. The distance patterns are similar in Hz and Bark because in the frequency region of F1 and F2, there is a fairly linear relationship between Hz and Bark. Nevertheless, the transformations are presented to ease comparisons to other studies for the reader.

On the whole, the Hillenbrand et al. (1995) and the Peterson and Barney (1952) studies showed large mean distances to the PNW vowels: HGCW vowels at 333 Hz (1.67 Bark) and PB vowels at 275 Hz (1.38 Bark). Only a slight difference of 57 Hz (0.29 Bark) is seen between the average HGCW–PNW distance and the average PB–PNW distance. This accounts for the general lack of effect of study in the ANOVA. However, as indicated by the Vowel × Study interactions and subsequent post hoc tests, in some regions of the vowel space, the PB–PNW distances are larger, whereas in others, the HGCW–PNW distances are larger: the back vowels /ʊ, u/ show greater distances to PB, and the low vowels /æ, ɑ/ show greater distances to HGCW. For /æ, ɑ/, the HGCW–PNW distances are 357 Hz (1.66 Bark) and 257 Hz (1.44 Bark) greater than the equivalent PB–PNW distances. On the other hand, for /ʊ, u/, the PB–PNW distances are 106 Hz (0.67 Bark) and 91 Hz (0.41) greater than the HGCW–PNW distances.

Discussion

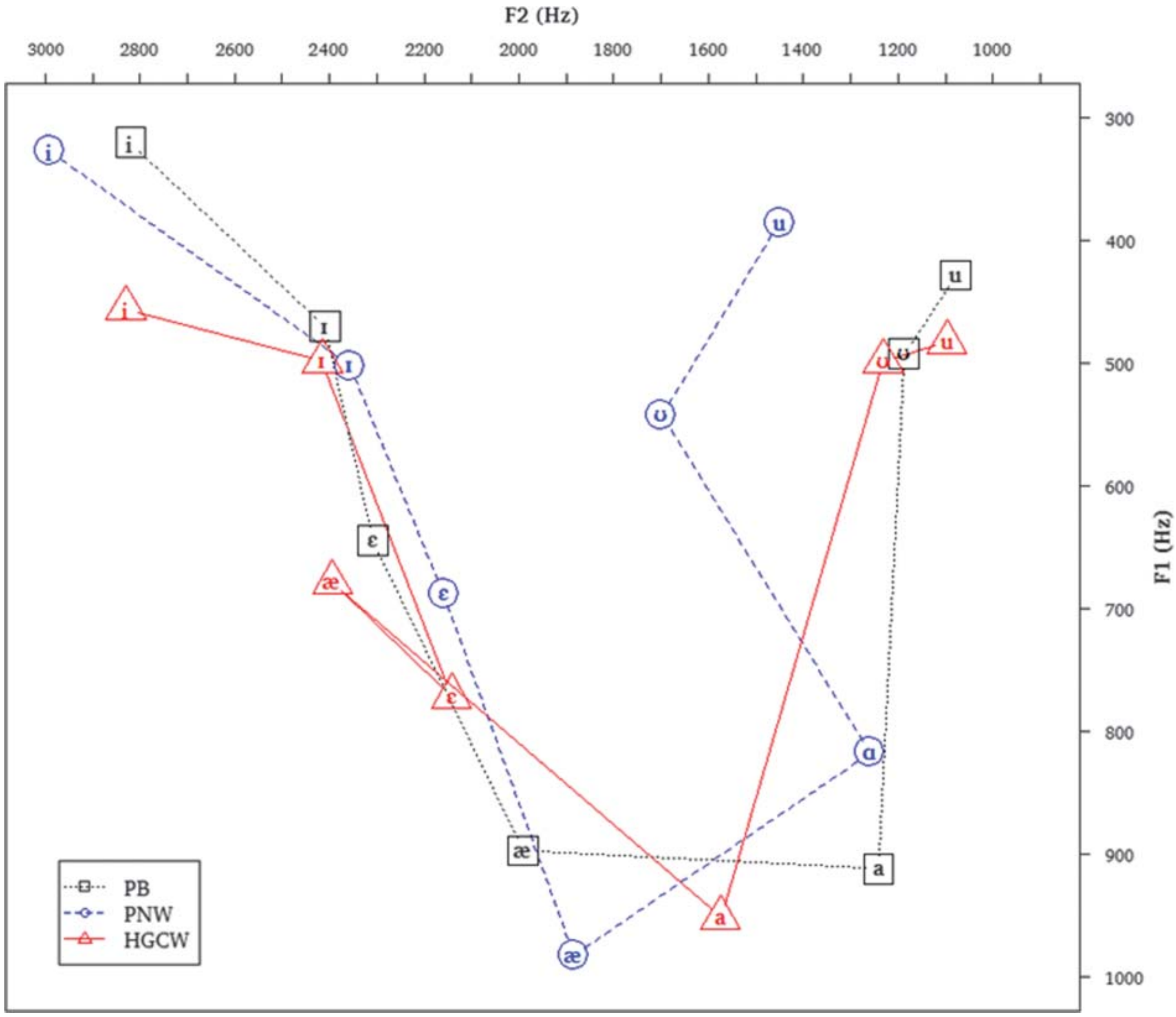
The results of Experiment 1 indicate that PNW regional vowels vary, sometimes substantially, from standardized vowels in terms of their formants. They also demonstrate that neither of the standardized vowel sets (HGCW or PB) is optimal in terms of the vowel mismatch because each has vowels that show larger distances than those of the other study. This result should be unsurprising to readers who are familiar with the sociolinguistic literature. After all, nearly all of the subjects in the Hillenbrand et al. (1995) study were from regions that participate in the well-documented "Northern Cities" vowel shift. As described recently in detail by Labov et al. (2006), Clopper et al. (2005), and Jacewicz et al. (2007), the Northern Cities vowel space is characterized by large differences in the low vowels; in particular, it has a raised /æ/ compared with other regions and a relatively fronted /a/ vowel compared with other regions' back /ɑ/. At the same time, much of the West is characterized by a fronting relative to other regions of the high-back vowels [ʊ] and [u], as noted by Eckert (1989) and as documented instrumentally by Hagiwara (1997) and Clopper et al. (2005). However, in agreement with Ingle et al. (2005), neither of these back vowels in our data show as much fronting as is seen in southern California.

Given the tolerance to variation in normal speech, whether such differences will lead to errors when nonregional vowels are presented in identification studies remains to be seen. Moreover, whether dissimilarity distance has an increasingly negative impact on perception such that at greater distances, recognition accuracy declines, or whether the negative effect is equivalent across distances, also remains to be determined. We tested these questions in Experiment 2 using a forced-choice

---

[2] For an alternative approach that incorporates consonant spectra as well, see Heeringa, Johnson, and Gooskens (2009).

Figure 1. First (F1) and second (F2) formant values for vowels spoken by female speakers. Circles and dashed lines represent the mean across six speakers for Pacific Northwest vowels. Squares and dotted lines represent published values for PB (Peterson & Barney, 1952) vowels, and triangles and solid lines represent published values for HGCW (Hillenbrand, Getty, Clark, & Wheeler, 1995) vowels.



vowel identification task in which subjects identified synthetic vowels that varied by distance (as defined in Equation 1) in varying amounts of noise.

Experiment 2: Effect of Dissimilarity Distance on Vowel Identification

Method

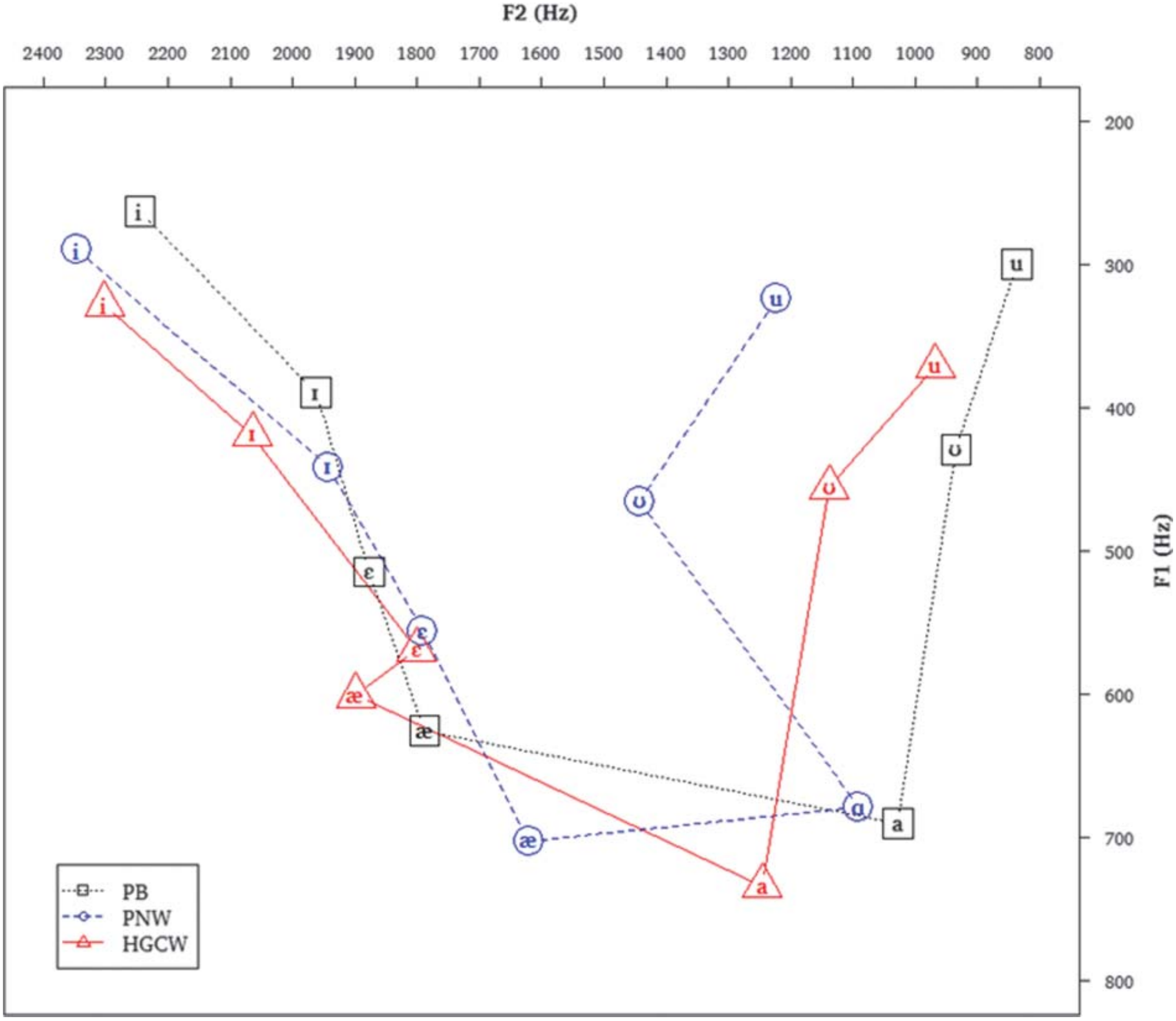
Listeners

There were 15 listeners: 10 women and five men (M age = 24.6 years, age range: 18–38). All were monolingual English speakers who were native to the PNW. All listeners had bilaterally normal hearing, defined as pure-tone thresholds of 20 dB HL or better at octave frequencies between 0.25 kHz and 8 kHz. Subjects from the production study were excluded from the identification study. All procedures were reviewed and approved by the local institutional review board, and subjects were reimbursed for their time.

Stimuli

There were three steps to creating the stimuli, each of which we describe in more detail below. First, we chose pairs of vowels for the identification task. Second, we compared formant values from different parts of North

Figure 2. F1 and F2 values for vowels spoken by male speakers. Circles and dashed lines represent the mean across six speakers for Pacific Northwest vowels. Squares and dotted lines represent published values for PB vowels, and triangles and solid lines represent published values for HGCW vowels. To facilitate comparison with Figure 1, we shifted the scales to optimize the plot area, but the frequency spacing is the same as that in Figure 1.



America to determine the distance steps and to identify individual vowel formant values. Third, we synthesized stimuli using the identified formant values.

We selected two vowel pairs /æ–ɛ, u–ʊ/ for the task on the basis of three criteria. First, the vowels /æ–ɛ, u–ʊ/ represent near neighbors in the acoustic F1 × F2 vowel space and, therefore, should be more easily confused than more distant pairs such as /ɑ–ʊ, i–u/. Second, these pairs represent vowels that set the West—and, therefore, the PNW—vowels apart from other regional vowels. Third, on the basis of our review of regional vowel studies across dialect regions of North America, these pairs represent sets for which monophthongs can be identified.

After identifying the two target vowel pairs, we examined published vowel formant values in a large number of sources representing all seven dialect regions: West, North Central, Midland, South, Inland North, Mid-Atlantic, New England (as defined in Clopper et al., 2005; Labov et al., 2006; and Jacewicz et al., 2007). Using the Euclidean distance formula in Equation 1, we calculated a *dissimilarity distance* from the six PNW vowels to the vowels as described in Clopper et al. (2005), Labov et al. (2006), Hagiwara (2006), and Jacewicz et al. (2007). In calculating distances, we used previously published PNW values (Ingle et al., 2005) to ensure equivalency in the comparisons. After comparing all regions of North America with the PNW

Table 3. Formants and *SD*s of the Pacific Northwest (PNW) vowels (in Hz).

	F1				F2			
	Women		Men		Women		Men	
Vowel	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
i	327	41	290	10	2,991	131	2,346	188
ɪ	502	22	441	34	2,357	77	1,942	167
e	405	22	404	42	2,792	143	2,217	192
ɛ	687	74	556	34	2,160	113	1,791	133
æ	983	69	703	76	1,884	201	1,622	56
ɑ	817	78	679	43	1,259	77	1,091	35
ʊ	542	41	465	31	1,699	202	1,444	190
u	385	29	324	32	1,450	215	1,223	95

Table 5. Region and Euclidean distances of the vowel stimuli.

Vowel	Distance—Region	Distance (Hz)	Distance (Bark)
æ	0—PNW	0	0.00
æ	1—Western North Carolina	176	0.75
æ	2—Northern United States	376	1.46
ɛ	0—PNW	0	0.00
ɛ	1—Southern California	138	0.56
ɛ	2—Northern United States	276	1.52
ʊ	0—PNW	0	0.00
ʊ	1—Winnipeg (Canada)	111	0.59
ʊ	2—Northern United States	326	1.67
u	0—PNW	0	0.00
u	1—Winnipeg (Canada)	152	0.77
u	2—Western North Carolina	323	1.70

results, we selected two non-PNW distances that varied by a consistent amount in dissimilarity distance (values that were closest in terms of differences in F1 and F2). These were designated as either 0 (*identical to PNW vowels*), 1 (*similar to PNW vowels*; 111–176 Hz, 0.52–0.77 Bark), or 2 (*dissimilar to PNW*; 276–375 Hz, 1.46–1.70 Bark). We used published values from previous studies to preserve the greatest similarity to existing regional vowels and to ensure the maximal naturalness of the subsequent synthetic stimuli, even though this meant some variation within the Distance 1 and Distance 2 groups. The resulting distances with their regional (study-based) identifier are summarized in Table 5.

Once formant distances had been established, we used the F1 and F2 values from the published studies to create a set of synthetic stimuli at each of the three distance steps (0, 1, and 2). The stimuli were 200 ms long and had a pitch contour that began at 130 Hz and remained steady state for 50 ms, gradually falling thereafter to 90 Hz. This created a male-sounding voice. We took the F1 and F2 frequencies from the published studies, estimated F3 using published regression formulas (Nearey, 1989), and fixed F4 at 3500 Hz. We calculated formant bandwidths from the algorithm described by Johnson, Flemming, and Wright (1993). The values used to synthesize the stimuli are summarized in Table 6. To ensure equivalent signal-to-noise ratios (SNRs) across vowels, the stimuli were root-mean-square normalized following synthesis.

Procedure

Throughout the experiment, listeners were seated in a double-walled sound booth. They were first trained in the orthographic decision labels for vowels by associating

Table 4. Mean distance (Dist.) in Hz and in Bark from PNW values (HGCW, PB) by vowel.

	Dist. Hz			Dist. Bark		
Vowel	HGCW	PB	Diff. (Hz)	HGCW	PB	Diff. (Bark)
i	**274** (122)	211 (114)	63	**1.27** (0.45)	0.76 (0.29)	0.52
ɪ	136 (85)	**189** (72)	53	0.59 (0.35)	**0.92** (0.32)	0.33
e	312 (173)			1.47 (0.60)		
ɛ	159 (106)	**209** (127)	50	0.76 (0.47)	**0.96** (0.58)	0.20
æ	**612** (245)	255 (156)	357	**2.87** (1.09)	1.21 (0.69)	1.66
ɑ	**371** (117)	96 (69)	275	**2.04** (0.69)	0.60 (0.44)	1.44
ʊ	427 (215)	**533** (214)	106	2.09 (0.91)	**2.76** (0.93)	0.67
u	349 (173)	**440** (168)	91	2.08 (0.80)	**2.49** (0.73)	0.41
M	333	276	57	1.67	1.38	0.29

Note. Numbers in parentheses are *SD*s. Boldface type indicates that there is a greater distance between the pair. HGCW = Hillenbrand, Getty, Clark, and Wheeler (1995); PB = Peterson and Barney (1952).

Table 6. Formants (Fs) and bandwidths (Bs) of the vowel stimuli.

Vowel	F1	F2	F3	F4	B1	B2	B3	B4
æ-0	635	1,579	2,279	3,500	71	50	170	200
æ-1	675	1,750	2,504	3,500	73	69	230	200
æ-2	588	1,952	2,700	3,500	66	101	284	200
ɛ-0	483	1,834	2,504	3,500	59	92	232	200
ɛ-1	458	1,698	2,329	3,500	58	76	185	200
ɛ-2	630	1,600	2,301	3,500	71	53	176	200
ʊ-0	423	1,445	2,146	3,500	59	60	120	200
ʊ-1	459	1,340	2,213	3,500	65	63	117	200
ʊ-2	469	1,122	2,300	3,500	74	72	99	200
u-0	313	1,176	2,158	3,500	60	74	77	200
u-1	313	1,328	2,102	3,500	55	68	91	200
u-2	390	1,490	2,104	3,500	55	60	118	200

visually presented words, such as *bet* and *bat,* with the orthographic decision labels. The following are the labels used in the experiment preceded by their IPA symbols: /ɛ/ eh, /æ/ ae, /ʊ/ uh, and /u/ *oo*. When subjects achieved 88% accuracy in the association of the labels with visually presented words, they proceeded to the main experiment.

To measure vowel identification, we presented stimuli monaurally to the right ear via an insert earphone. Each vowel was presented in a background of a masker noise with frequency spectrum matched to the long-term spectrum across all vowels. The masker consisted of a white noise that was low-pass filtered with a 300-Hz cutoff frequency and a 5-dB-per-octave spectral slope. Vowels were presented at SNRs of +2 dB SNR, +6 dB SNR, and +10 dB SNR. In each case, the level of the vowel was fixed at 65 dB SPL, and the level of the noise was adjusted to the desired SNR.

Stimuli were presented blocked by vowel condition (front pairs, back pairs) and SNR, creating six blocks total. The order of the blocks was randomized for each subject. A block consisted of 18 randomly ordered trials (two vowels, three distances, and three repetitions). To mimic clinical (i.e., untrained) presentation, each block was presented once. Subjects responded to each trial in a two-alternative forced-choice paradigm using a touch screen. The choices were presented on the screen as buttons labeled *eh* or *ae* for the front vowel block and *uh* or *u* for the back pair block. The location of response buttons on the touch screen was randomized on each trial to minimize response bias.

Results

We analyzed the results with a three-way ANOVA with the factors distance (0, 1, or 2), SNR (+2 dB, +6 dB, or +10 dB), and vowel. The results of the ANOVA are reported in Table 7.

Table 7. ANOVA results for the effect of Distance × Signal-to-Noise Ratio (SNR) × Vowel.

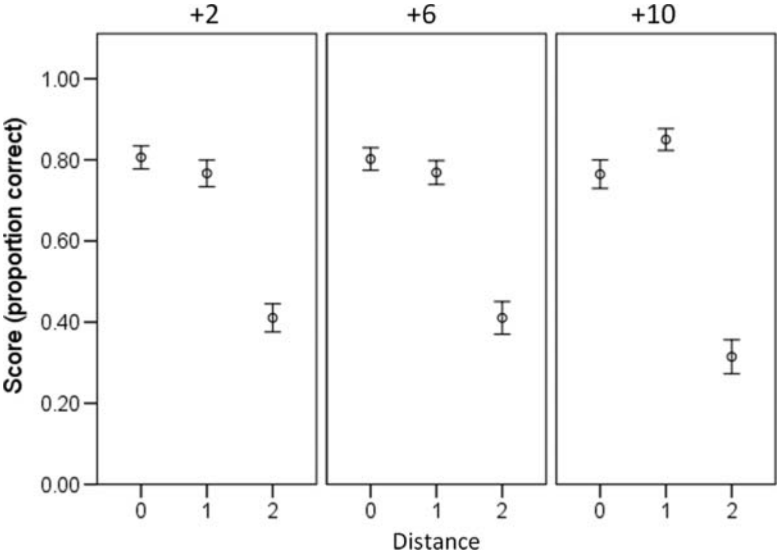
Variable	*df*	*F*	*p*
Vowel	3	6.26	< .005
SNR	2	0.33	.719
Distance	2	180.76	< .005
Vowel × SNR	6	1.04	.398
Vowel × distance	6	12.90	< .005
SNR × distance	4	2.91	.021
Vowel × SNR × distance	12	1.34	.190
Error	504		

The effect of distance for each SNR is shown in Figure 3. In general, scores were similar between distances 0 and 1 and dropped significantly for distance 2 for all SNRs. As suggested by Figure 3, there was no effect of SNR. There was an effect of distance and a significant interaction between distance and SNR. To investigate further, we collapsed the data across vowels and conducted a post hoc analysis by comparing the distance scores within each SNR. The Distance × SNR interaction was due to a small difference in the magnitude of the effect whereby the difference between distances 0 and 1 approached significance for SNR 10 ($p = .087$) but not for SNR 6 and 2 (ps = .475 and .384, respectively). Note that for SNR 10, there was a slightly higher score at distance 1 compared with distance 0, but given the nonsignificant p value, we considered this as reflecting measurement variability. At all SNRs, there was a significant decrease in score between distances 1 and 2 ($p < .005$ in each case). We used post hoc means comparisons to examine the Distance × Vowel interaction (see Figure 4). Post hoc analyses indicated that three of the vowels showed a significant effect of distance ($p < .005$ for ɛ, æ, and u). In each case, the difference between distances 0 and 1 was nonsignificant ($p > .050$), and the difference between distances 1 and 2 was significant ($p < .005$). For the remaining vowel (ʊ), the effect of distance was not significant ($p = .107$).

Discussion

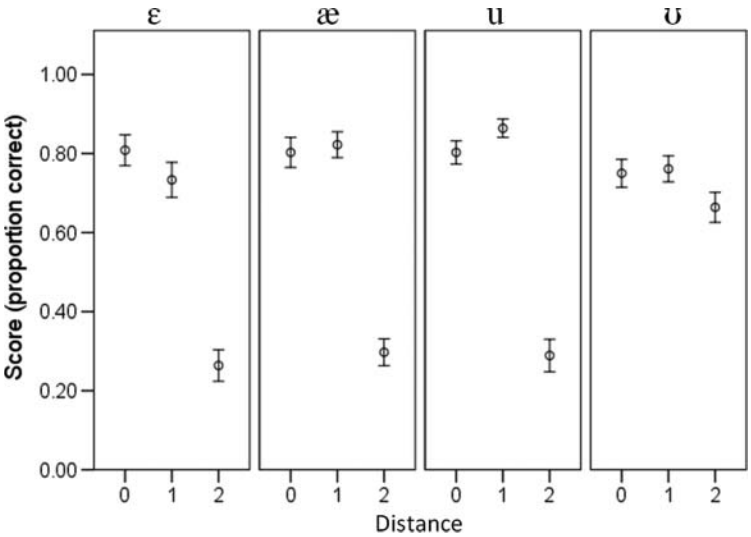
Although we focused on a specific region of the United States (PNW), we anticipate that the results seen here can be generalized to other regions of the country. In Experiment 2, the effect of distance 2 was robust, whereas distance 1 did not prove reliable. This may indicate that when making comparisons across dialects, small differences have little or no effect on identification, whereas larger differences have large negative effects. This finding needs to be tested more thoroughly by examining the perceptual effects of subregional or sociolectal variation within dialect-specific vowel systems.

Figure 3. Mean proportion correct vowel identification for three signal-to-noise ratios (+2, +6, and +10 dB) as a function of vowel distance. Error bars show ±1 *SE* of the mean.



Our production results show that grouping all of the regional accents into any one of the dialect regions (e.g., the South or West) based on overall vowel similarity creates vowel-specific mismatches: Whereas some PNW vowels are typical of other West coast varieties of English, such as the one spoken in southern California, other PNW vowels differ quite dramatically. For example, PNW /æ/ shows a Euclidean distance of 157 Hz (1.06 Bark) compared with descriptions of the West in general as defined by Clopper et al. (2005) and Labov et al. (2006), or compared with regionally specific values reported for southern California (e.g., Hagiwara, 1997; Johnson et al., 1993). PNW vowels also show less extreme fronting of /u/ than seen in the generic West or in the specific Southern Californian descriptions. If the distance effects in perception found in Experiment 2 extend to other vowels as predicted, then the treatment of the entire West as a single dialect group may be too

Figure 4. Mean proportion correct vowel identification for four vowels as a function of vowel distance. Data are collapsed across signal-to-noise ratio. Error bars show ±1 *SE* of the mean.



broad for many purposes, and this underscores the importance of considering regional-specific vowels rather than broad areas of the United States (e.g., the West, Midwest, South) as is typically done.

The perception results have both methodological and theoretical implications for studies of speech perception. They indicate that vowel category distances related to regional variation have a reliable negative impact on vowel identification. Stated another way, greater dissimilarity increases the risk that a vowel may be identified poorly in perception experiments; however, slight dissimilarities appear to have a negligible effect. It must be noted that in creating our stimuli, we chose a relatively modest distance for the distance 2 condition because we wanted to avoid vowels that were so different that no researcher would use them as stimuli. Accordingly, these results should represent the range of dialects that could be encountered by an individual in a realistic situation.

It is important to note that these were vowels presented in isolation, and therefore the task is not directly representative of a listener's everyday experience with accent variation. Whether larger stretches of speech with context effects will show the same reliability remains to be seen; moreover, the negative impact of dialect differences on speech perception may be mitigated by a variety of factors, such as the experience of the listener with other accents. For example, Evans and Iverson (2004) found that listeners were able to shift their perceptual targets to match that of the input accent over time. However, adaptation to an accent may require long-term exposure and may not occur for all individuals (Evans & Iverson, 2007). Sumner and Samuel's (2009) study provided evidence that there are dialect effects in both the immediate perceptual processing and long-term recall of speech stimuli. They concluded that there are individual experienced-based differences that affect one's ability to process stimuli presented in a nonnative dialect.

Research on highly proficient nonnative and bilingual listeners suggests that noise interacts with language background (e.g., Mayo, Florentine, & Buus, 1997; Meador, Flege, & MacKay, 2000; Rogers, Lister, Febo, Besing, & Abrams, 2006). In these studies, listeners appear native-like under ideal listening conditions but experience a more extreme decline in perceptual performance in noise than native listeners do. Age of acquisition also plays a role in these studies; the earlier the second language was acquired, the less noise seemed to affect their perceptual performance. The lack of interaction with noise shown in the present study may be due to an ability of native speakers to dynamically adapt to moderate levels of noise; the more extensive the experience with variation, the more robust the perceptual response in the face of noise. This is consistent with recent findings that the combined effects of dialect variation and noise on speech processing are lessened when the listener has extensive experience with a similar dialect (Clopper & Bradlow, 2009; Clopper, Pierrehumbert, & Tamati, 2010, Sumner & Samuel, 2009).

We believe that these findings also suggest that early and extensive experience with relevant variation creates a robust representation that listeners with normal hearing can draw on under difficult listening conditions, such as in background noise. Individuals with hearing loss may experience an increased difficulty adapting to dialect variability in noise because of the reduced redundancy in the received signal. There is a dearth of research on how listeners with hearing loss respond to dialect variation and other types of phonological distortion. Such work continues to be a focus of interest in our laboratories.

Conclusion

The results of the production study indicate that even when vowels are labeled with the same phonetic symbol (by convention), there can be large acoustic differences between one region's vowels and another's; moreover, there are vowel-specific effects such that some vowels are quite similar across studies and others vary widely. On the whole, the PNW vowels were more similar to the PB vowels than to the HGCW vowels, but for each of the standard vowel sets, some PNW vowels showed larger distances. When presented as stimuli in noise, vowel distance had a reliable (negative) impact on the listener's ability to correctly identify the target vowel. These findings indicate that researchers and clinicians should take care in choosing stimuli for perception experiments. We recommend that researchers use regionally validated vowels instead of relying on standardized vowels in tasks that use vowel perception.

Acknowledgments

This work was supported by Grant R01 DC60014 from the National Institutes of Health. The authors are grateful to Stephanie Bor, Star Reed, Daniel McCloy, and Kerry Witherell for their assistance with stimulus preparation, data collection, and analysis.

References

Adank, P., & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on spoken-word comprehension. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany* (pp. 1925–1928). Dudweiler, Germany: Pirrot.

American National Standards Institute. (2004). *Specifications for audiometers* (ANSI S.36-2004). New York, NY: Author.

Bor, S., Souza, P., & Wright, R. (2008). Multichannel compression: Effects of reduced spectral contrast on vowel identification. *Journal of Speech, Language, and Hearing Research, 51,* 1315–1327.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 116,* 3647–3658.

Clopper, C. G., & Bradlow, A. R. (2009). Free classification of American English dialects by native and non-native listeners. *Journal of Phonetics, 37,* 436–451.

Clopper, C. G., Pierrehumbert, J. B., & Tamati, T. N. (2010). Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise. *Laboratory Phonology, 1,* 65–92.

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America, 118,* 1661–1676.

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication, 47,* 32–42.

Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2007). The perception of phonemic contrasts in a non-native dialect. *The Journal of the Acoustical Society of America, 121,* E131–E136.

Eckert, P. (1989). *Jocks and burnouts: Social categories and identity in the high school*. New York, NY: Teachers College Press.

Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *The Journal of the Acoustical Society of America, 115,* 352–361.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America, 121,* 3814–3826.

Ferguson, S. H., Jongman, A., Sereno, J. A., & Keum, K. A. (2010). Intelligibility of foreign-accented speech for older adults with and without hearing loss. *Journal of the American Academy of Audiology, 21,* 153–162.

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research, 38,* 379–412.

Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (pp. 9–31). Tübingen, Germany: Niemeyer/Linguistische Arbeiten.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America, 102,* 655–658.

Hagiwara, R. (2006). Vowel production in Winnipeg. *Canadian Journal of Linguistics, 51,* 127–141.

Heeringa, W., Johnson, K., & Gooskens, C. (2009). Measuring Norwegian dialect distances using acoustic features. *Speech Communication, 51,* 167–183.

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

Ingle, J., Wright, R., & Wassink, A. B. (2005). Pacific Northwest vowels: A Seattle neighborhood dialect study. *The Journal of the Acoustical Society of America, 117,* 2459.

Jacewicz, E., Fox, R. A., & Salmons, J. (2007, August). *Vowel space areas across dialects and genders*. Paper presented at the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69,* 505–528.

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America, 122,* 2365–2375.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English*. Berlin, Germany: Mouton de Gruyter.

Mayo, L., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40,* 686–693.

Meador, D., Flege, J. E., & MacKay, I. R. A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition, 3,* 55–67.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45,* 73–97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of foreign-accented speech. *Language and Speech, 38,* 289–306.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America, 85,* 2088–2113.

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research, 51,* 1480–1493.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24,* 175–184.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27,* 465–485.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60,* 487–501.

Traunmüller, H. (1990). Analytical expressions for the Tonotopic Sensory Scale. *The Journal of the Acoustical Society of America, 88,* 97–100.

Vance, T. (1987). "Canadian raising" in some dialects of the northern United States. *American Speech, 62,* 195–210.

Wassink, A. B., Squizzero, R., Schirra, R., & Conn, J. (2009, October). *Effects of style and gender on fronting and raising of /ɛ/, /e:/ and /æ/ before /g/ in Seattle English*. Paper presented at the 38th annual meeting of New Ways of Analyzing Variation, Ottowa, Ontario, Canada.

Webster, G. (2002). *Toward a psychologically and computationally adequate model of speech perception* (Unpublished doctoral dissertation). Department of Linguistics, University of Washington, Seattle.

Comparing Identification of Standardized and Regionally Valid Vowels

Richard Wright, and Pamela Souza

J Speech Lang Hear Res 2012;55;182-193; originally published online Dec 22, 2011;
DOI: 10.1044/1092-4388(2011/10-0278)