Published in final edited form as:

J Speech Lang Hear Res. 2012 February ; 55(1): 182–193. doi:10.1044/1092-4388(2011/10-0278).

Comparing identification of standardized and regionally-valid vowels

Richard Wright and

Department of Linguistics, University of Washington

Pamela Souza

Department of Communication Sciences and Disorders, Northwestern University

Abstract

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—We analyzed formant values of 8 monophthong vowels produced by 12 talkers from the region where the research took place and compared them to 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 rather than relying on standardized vowels in vowel perception tasks.

This study consists of two experiments with two related goals: 1) to determine the degree to which vowels in the Pacific Northwest region matched General American English vowels, and 2) to determine the degree to which acoustic dissimilarity has an effect on vowel identification accuracy. While any number of dimensions of the speech signal may be studied to examine differences across regional dialect we have chosen 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 to many consonants which 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 not only information about the vowel's quality itself but also about the flanking consonants through their formant transitions. Again, this may be particularly true under conditions of multiple distortions. Recent work by Webster (2002) has indicated 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, *et al.*, 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, particularly 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 pre-recorded 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 which varies by laboratory or clinic depending on geographic location, and which 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 nor 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. Labov, *et al*, 2006; Clopper, *et al*, 2005). 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, Detroit, Cleveland, and Buffalo, 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. Eckert, 1989; Clopper, *et al*, 2005; 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. While this last point may seem obvious to the researcher who is 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 in 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, while 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. Hagiwara, 1997; Clopper, *et al*, 2005; Labov, *et al*, 2006, Jacewicz, *et al*, 2007). 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 *et al*’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 and an unfamiliar accent. In the study subjects' response times were slower for words presented in the unfamiliar accent. The effect persisted even when listeners were exposed to 20 minutes of speech in the unfamiliar accent prior to the animacy decision task, indicating a lasting effect for unfamiliar accents. Similarly, Floccia *et al* (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 doesn't 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 (2008) 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 including Mid-Atlantic, Northern, and Southern regions. While not all dialectal regions were equally represented so true regional effects were not fully probed, on the whole their results indicate 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 we 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 counter evidence in that listeners seem to be able to adapt to foreign accented speech (e.g. Munro & Derwing, 1995a, 1995b; Clarke & Garrett, 2004; Ferguson, Jongman, Sereno, & Keum, 2010). Moreover, the native dialect effects that have been observed for speech perception and processing for Dutch (Adank & McQueen, 2007), French (Dufour, *et al*, 2007), and English (Cutler, *et al*, 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 F1 and F2 differences between dialects) on identification accuracy; therefore, it is unknown whether or not distance has a negative impact on vowel identification.

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": Peterson and Barney (1952) (PB) and Hillenbrand, Getty, Clark, and Wheeler (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 Pacific Northwest vowels to standard vowels

While 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: the first and second formants (F1 and F2 respectively) are typically measured at mid-point or steady-state. This measure was used in Peterson and Barney's (1952) study as well as Hillenbrand, *et al.* (1995) and other recent vowel studies (e.g. Hagiwara, 1997; Clopper, *et al.*, 2007). We predict that there will be study effects such that both PB and HGCW vowels will differ from the PNW vowels. Moreover, we predict that there will be vowel by study interactions, such that PB and HGCW vowels will differ from PNW vowels in study-specific ways.

Method

Talkers—Six adult males and six adult females participated in vowel recordings. All of the talkers had grown up in the Pacific Northwest (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 Pacific Northwest 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 (re: ANSI, 2004) at octave frequencies between .25 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 monophthong vowels (/i, ɪ, e, ɛ, æ, ɑ, ʊ, u/) were recorded using randomized wordlists. 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 is 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 *hewed*. 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 via a VU meter to ensure sufficient output levels without clipping. Tucker-Davis Technologies System 2 with an AP2 sound card and a Shure BG 1.0 omnidirectional microphone were utilized for all the recordings. Four talkers were recorded direct-to-disc at a rate of 44.1 kHz. The remaining 8 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—One representative token of each vowel was selected for each talker based on two criteria: (a) recording fidelity and clarity of each participant'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. Formant steady-state was determined using a wideband spectrogram and accompanying linear predictive coding

¹Other studies of the PNW accent have defined specific neighborhood regions of the Northwest and/or specific sounds (e.g. Ingle, *et al.*, 2005; Wassink, *et al.*, 2009).

(LPC) formant track with 12 coefficients. Pitch steady state was determined using an autocorrelation pitch track with a 25 ms window overlaid on a narrowband 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–F4 values and recordings were used in a separate study on hearing aid compression (Bor, *et al.*, 2008). In the present study only F1 and F2 were used. To minimize the risk of error, formant measurements were taken from an LPC spectrum overlaid on a fast Fourier transform (FFT) power spectrum with a sample window of 128 points and visually compared to its broadband spectrogram. In the event that there were LPC errors for a particular talker, the number of filter coefficients (poles) were 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 to HGCW and PB using the first formant (F1) and second formant (F2) values. Because we had a reasonably large sample (6 males and 6 females), we were able to compare 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 6 speakers from each of the larger studies (P&B, 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 by 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 ANOVAs with F1 and F2 as dependent variables, and Study (PNW, HGCW, PB) and Vowel (*i, ɪ, e, 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 indicate that there is the expected reliable effect for Vowel on F1 and F2 for both males and females indicating that in each study the vowels are reliably separated on both dimensions. More interestingly, there is an effect for Study on F2 for males and a Study by Vowel interaction for F1 and F2 in both the male and female data. The interactions indicate study-specific differences for individual vowels. To probe which vowels were contributing to the interactions, the formant values were submitted to a series of Bonferroni-Dunn post-hoc t-tests with an alpha of .05 (corrected to .0167). For PNW male F1 values, the vowel [æ] was different from its HGCW counterpart, but none were different from their PB counterparts. For PNW male F2, 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 PNW female F1, the vowels [i], [æ], [a], and [ʊ] were all different from their HGCW counterparts, but none were different from their PB counterparts. For PNW female F2, the vowels [e], [æ], [a], [ʊ], and [u] were all reliably different from their HGCW counterparts; and the vowels [i], [e], [ʊ], and [u] were all reliably different from their PB counterparts.

To illustrate the Study by Vowel effects, male and female vowel means are plotted in Figures 1 and 2 respectively with the PNW vowel plots overlaid on vowels from Hillenbrand *et al* (1995) (HGCW) and Peterson and Barney (1952) (PB). Notable differences include a lowered (higher F1) and backed (lower F2) PNW/æ/relative to the HGCW counterpart, a raised (lower F1) and backed (lower F2) PNW/ɑ/relative to the HGCW counterpart, 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, Euclidean distances were calculated from each PNW vowel's F1 by F2 point in the vowel space to its gender-matched counterpart from the PB and HGCW values². The equation in (1) was used to calculate distance. Throughout this study, this formula is used 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 $F1_p$ is an individual PNW speaker's F1 value for a particular vowel and $F1_i$ is a comparison value for the equivalent PB or HGCW vowel F1 mean, and where $F2_p$ is an individual PNW speaker's F2 value for a particular vowel and $F2_i$ is a comparison value for the equivalent PB or HGCW vowel F2 mean. Because we are interested in relating these differences to perceptual effects, we have also calculated the Euclidean distances in Bark using the formula published in Traunmüller (1990).

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

The results of the Euclidean distance measures are summarized in Table 4 which displays the mean distance and standard deviations 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 HGCW and the PB studies show large mean distances to the PNW vowels: HGCW at 333 Hz (1.67 Bark), and PB 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-hocs, in some regions of the vowel space the PB-PNW distances are larger while in others the HGCW-PNW distances are larger: the back vowels/ɔ, u/show greater distances to PB, while 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/ɔ, u/the PB-PNW distances are 106 Hz (0.67 Bark) and 91 Hz (0.41) greater than the HGCW-PNW distances.

Discussion—Results of Experiment 1 indicate that PNW regional vowels vary, sometimes substantially, from standardized vowels in terms of their formants. It also demonstrates that neither of the standardized vowel sets (HGCW, 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 those familiar with the sociolinguistic literature. After all, nearly all of participants in the HGCW study were from regions that participate in the well documented “Northern Cities” vowel shift. As described recently in detail in 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 to other regions and relatively fronted/a/vowel compared to 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 [ɔ] and [u] as noted by Eckert (1989) and as documented instrumentally by Hagiwara (1997) and Clopper *et al.* (2005). However, in agreement with

²For an alternative approach that incorporates consonant spectra as well, see Heeringa *et al.* (2009).

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, it remains to be seen whether such differences will lead to errors when non-regional vowels are presented in identification studies. Moreover, it remains to be answered 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. These questions were tested in Experiment 2 using a forced-choice vowel identification task in which participants identified synthetic vowels that varied by distance (as defined in Formula 1, above) in varying amounts of noise.

Experiment 2: Effect of dissimilarity distance on vowel identification

Method

Listeners—There were 15 listeners, 10 female and 5 male adults (mean age 24.6 years, age range 18–38 years). All were monolingual English speakers who were native to the Pacific Northwest. All listeners had bilaterally normal hearing, defined as pure-tone thresholds of 20 dB HL or better at octave frequencies between .25 and 8 kHz. Participants 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 described in more detail below. First, pairs of vowels were chosen for the identification task. Second, formant values from different parts of North America were compared to determine the distance steps and to identify individual vowel formant values. Third, stimuli were synthesized using the identified formant values.

Two vowel pairs/æ-ɛ, u-ʊ/were selected for the task based on three criteria. First, the vowels/æ-ɛ, u-ʊ/represent near neighbors in the acoustic F1 by 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, based on 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 7 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 1, we calculated a *dissimilarity distance* from the six PNW vowels to vowels as described in those studies. In calculating distances, previously published PNW values (Ingle, *et al.*, 2005) were used to ensure equivalency in the comparisons. After comparing all regions of North America to the PNW 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 3 – dissimilar to PNW (276–375 Hz, 1.46–1.70 Bark). Published values from previous studies were used 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, the F1 and F2 values from the published studies were used to create a set of synthetic stimuli at each of the three distance steps (0, 1, 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. F1 and F2 frequencies were taken from the published studies, F3 was estimated using published regression formulas (Nearey, 1989) and F4 was fixed at 3500 Hz. Formant bandwidths were calculated from the algorithm described in Johnson, *et al.* (1993). Table 6 summarizes the values used to synthesize the stimuli. To ensure equivalent SNRs across vowels, the stimuli were RMS normalized following synthesis.

Procedure—Throughout the experiment, listeners were seated in a double-walled sound booth. Subjects were first trained in the orthographic decision labels for vowels by associating 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, /*u*/u. When participants achieved 88% accuracy in the association of the labels with visually presented words they proceeded to the main experiment.

To measure vowel identification, stimuli were presented 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 300 Hz cut off frequency and 5 dB per octave spectral slope. Vowels were presented at signal-to-noise ratios of +2, +6, and +10 dB SNR. In each case the level of the vowel was fixed at 65 dB SPL and the level of the noise adjusted to the desired SNR.

Stimuli were presented blocked by vowel condition (front pairs, back pairs) and signal-to-noise ratio, creating 6 blocks total. The order of the blocks was randomized for each subject. A block consisted of 18 randomly-ordered trials (2 vowels, 3 distances, 3 repetitions). To mimic clinical (i.e., untrained) presentation, each block was presented once. Subjects responded to each trial in a 2-alternative forced-choice paradigm using a touch screen. The choices 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—Results were analyzed using a three-way ANOVA with factors Distance (0, 1, 2), SNR (2, 6, 10 dB), and vowel. The results of the ANOVA are reported in Table 7.

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 for SNR. There was an effect of Distance, and a significant interaction between Distance and SNR. To investigate further, data were collapsed across vowel and post-hoc analysis completed by comparing the distance scores within each SNR. The Distance by 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 ($p=.475$ and $p=.384$, respectively). Note that for SNR 10 there was a slightly higher score at Distance 1 compared to Distance 0, but given the nonsignificant p value this was considered to reflect measurement variability. At all SNRs, there was a significant decrease in score between distance 1 and 2 ($p<.005$ in each case). Post-hoc means comparisons were used to examine the Distance by Vowel interaction (Figure 4). Post-hoc analyses indicated that three of the vowels showed a significant effect of distance ($p<.005$ for *ɛ*, *æ*, *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 (u) 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 while distance-1 did not prove reliable. This may indicate that when comparing 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 sub-regional or sociolectal variation within dialect-specific vowel systems.

Our production results show that the grouping all of the regional accents into any one of the dialect regions (such as South or West) based on overall vowel similarity creates vowel specific mismatches; while 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 to descriptions of the West in general as defined by Clopper *et al*, 2005, and Labov *et al*, 2006, or compared to 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 broad for many purposes, and underscores the importance of considering regional-specific vowels rather than broad areas of the US (e.g., 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 in isolation and therefore the task is not directly representative of a listener's everyday experience with accent variation. It remains to be seen whether or not larger stretches of speech with context effects will show the same reliability. Moreover, the negative impact of dialect differences on speech perception may be mitigated by a variety of factors such as the experience of the listeners with other accents. For example, Evans and Iverson (2004) found that listeners are 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). There is evidence from a study by Sumner and Samuel (2009) 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 our ability to process stimuli presented in a non-native dialect.

Research on highly proficient non-native 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 suffer a more extreme decline in

perceptual performance in noise than seen for native listeners. 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 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 2008, 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 due to 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 vowels sets, there were some PNW vowels that 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. It is recommended that regionally validated vowels should be used rather than relying on standardized vowels in tasks that employ vowel perception.

Acknowledgments

Work was supported by NIH grant R01 DC60014. 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, JM. The effect of an unfamiliar regional accent on spoken-word comprehension. *Proceedings of the 16th International Congress of Phonetic Sciences*; Saarbruecken, Germany. Dudweiler: Pirrot (DVD); 2007. p. 1925-1928.
- Bor S, Souza P, Wright R. Multichannel Compression: Effects of Reduced Spectral Contrast on Vowel Identification. *Journal of Speech Language and Hearing Research*. 2008; 51(5):1315–1327.
- Clarke CM, Garrett MF. Rapid adaptation to foreign accented speech. *Journal of the Acoustical Society of America*. 2004; 116:3647–3658. [PubMed: 15658715]
- Clopper CG, Bradlow AR. Free classification of American English dialects by native and non-native listeners. *Journal of Phonetics*. 2009; 37:436–451. [PubMed: 20161400]
- Clopper CG, Pierrehumbert JB, Tamati TN. Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise. *Laboratory Phonology*. 2010; 1:65–92.
- Clopper CG, Pisoni DB, de Jong K. Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*. 2005; 118:1661–1676. [PubMed: 16240825]

- Cutler A, Smits R, Cooper N. Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication*. 2005; 47(1–2):32–42.
- Calandruccio L, Doherty KA. Spectral weighting strategies for hearing-impaired listeners measured using a correlational method. *Journal of the Acoustical Society of America*. 2008; 123(4):2367–78. [PubMed: 18397039]
- Dufour S, Nguyen N, Frauenfelder UH. The perception of phonemic contrasts in a non-native dialect. *Journal of the Acoustical Society of America*. 2007; 121(4):E131–E136.
- Evans BG, Iverson P. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*. 2004; 115(1):352–361. [PubMed: 14759027]
- Evans BG, Iverson P. Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*. 2007; 121 (6):3814–3826. [PubMed: 17552729]
- Eckert, P. *Jocks & Burnouts: Social Categories and Identity in the high school*. New York: Teachers College Press, Columbia University; 1989.
- Floccia C, Butler J, Goslin J, Ellis L. Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research*. 2009; 38:379–412. [PubMed: 19117134]
- Grabe, E. Intonational variation in urban dialects of English spoken in the British Isles. In: Gilles, P.; Peters, J., editors. *Regional variation in intonation*. Tuebingen, Germany: Niemeyer/Linguistische Arbeiten; 2004. p. 9-31.[Linguistic Work]
- Ferguson SH, Jongman A, Sereno JA, Keum KA. Intelligibility of foreign-accented speech for older adults with and without hearing loss. *Journal of the American Academy of Audiology*. 2010; 21(3):153–162. [PubMed: 20211119]
- Hagiwara R. Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*. 1997; 102:655–658.
- Hagiwara R. Vowel Production in Winnipeg. *Canadian Journal of Linguistics*. 2006; 51:127–141.
- Heeringa W, Johnson K, Gooskens C. Measuring Norwegian dialect distances using acoustic features. *Speech Communication*. 2009; 51:167–183.
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*. 1995; 97(5):3099–3111. [PubMed: 7759650]
- Ingle J, Wright R, Wassink AB. Pacific northwest vowels: a Seattle neighborhood dialect study. *Journal of the Acoustical Society of America*. 2005; 117(4):2459.
- Jacewicz, E.; Fox, RA.; Salmons, J. Vowel space areas across dialects and genders. Paper presented at the ICPHS; Saarbrücken. August 6–10, 2007; Dudweiler: Pirrot (DVD); 2007.
- Johnson K, Flemming E, Wright R. The hyperspace effect: Phonetic targets are hyperarticulated. *Language*. 1993; 69(3):505–528.
- Kewley-Port D, Burkle TZ, Lee JH. Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America*. 2007; 122(4):2365–2375. [PubMed: 17902871]
- Labov, W.; Ash, S.; Boberg, C. *Atlas of North American English*. Berlin: Mouton de Gruyter; 2006.
- Mayo L, Florentine M, Buus S. Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*. 1997; 40:686–693.
- Munro MJ, Derwing TM. Foreign accent comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*. 1995a; 45:73–97.
- Munro MJ, Derwing TM. Processing time, accent, and comprehensibility in the perception of foreign-accented speech. *Language and Speech*. 1995b; 38:289–306. [PubMed: 8816082]
- Nearey TM. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*. 1989; 85(5):2088–2113. [PubMed: 2659638]
- Nishi K, Kewley-Port D. Nonnative Speech Perception Training Using Vowel Subsets: Effects of Vowels in Sets and Order of Training. *Journal of Speech Language and Hearing Research*. 2008; 51(6):1480–1493.

- Rogers CL, Lister JJ, Febo DM, Besing JM, Abrams HB. Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*. 2006; 27:465–485.
- Peterson GE, Barney HL. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*. 1952; 24:175–184.
- Sumner M, Samuel AG. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*. 2009; 60:487–501.
- Traunmüller H. Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*. 1990; 88:97–100.
- Vance T. “Canadian Raising” in some dialects of the northern United States. *American Speech*. 1987; 62(3):195–210.
- Webster, G. Unpublished PhD Thesis. University of Washington; Seattle: 2002. Toward a psychologically and computationally adequate model of speech perception.
- Wassink, AB.; Squizzero, R.; Schirra, R.; Conn, J. Effects of Style and Gender on Fronting and Raising of /e/, /e:/ and /ae/ before /g/ in Seattle English. presentation at NWAV; Ottawa, ON. Oct. 22–25, 2009; 2009. p. 38
- Yund EW, Buckles KM. Multichannel compression hearing aids: Effect of number of channels on speech discrimination in noise. *Journal of the Acoustical Society of America*. 1995; 97:1206–1223. [PubMed: 7876443]

Figure 1.

First (F1) and second (F2) formant values for vowels spoken by female speakers. Circles and dashed lines represent the mean across 6 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.

Figure 2.

First (F1) and second (F2) formant values for vowels spoken by male speakers. Circles and dashed lines represent the mean across 6 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 to Figure 1, the scales are shifted to optimize the plot area but the frequency spacing is the same as in Figure 1.

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 $\pm$ one standard error about the mean.

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 $\pm$ one standard error about the mean

Table 1

ANOVA results for male vowels

Results for Male F1			
	Df	F value	Pr(>F)
study	2	1.739	0.1807
vowel	6	287.234	< 0.001
study:vowel	12	3.803	< 0.001
Results for Male F2			
study	2	15.134	< 0.001
vowel	6	277.375	< 0.001
study:vowel	12	7.073	< 0.001

Table 2

ANOVA results for female vowels

Results for Female F1			
	Df	F value	Pr(>F)
study	2	1.7194	0.184
vowel	6	154.024	< 0.001
study:vowel	12	9.8241	< 0.001
Results for Female F2			
study	2	2.7335	0.069
vowel	6	317.4220	< 0.001
study:vowel	12	11.3050	< 0.001

Table 3

Formants and standard deviations of the PNW vowels (Hz).

<i>Vowel</i>	<i>F1 (standard deviations)</i>		<i>F2 (standard deviations)</i>	
	Female	Male	Female	Male
i	327 (41)	290 (10)	2991 (131)	2346 (188)
ɪ	502 (22)	441 (34)	2357 (77)	1942 (167)
e	405 (22)	404 (42)	2792 (143)	2217 (192)
ɛ	687 (74)	556 (34)	2160 (113)	1791 (133)
æ	983 (69)	703 (76)	1884 (201)	1622 (56)
ɑ	817 (78)	679 (43)	1259 (77)	1091 (35)
ʊ	542 (41)	465 (31)	1699 (202)	1444 (190)
u	385 (29)	324 (32)	1450 (215)	1223 (95)

Table 4

Mean distance in Hz and in Bark from PNW values (HGCW, PB) by vowel (bold indicates greater distance of the pair)

<i>Vowel</i>	<i>Dist. Hz (st. dev.)</i>		<i>diff (Hz)</i>	<i>Dist Bark (st. dev.)</i>		<i>diff (Bark)</i>
	HGCW	PB		HGCW	PB	
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
Mean	333	276	57	1.67	1.38	0.29

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 N. Carolina	176	0.75
æ	2-Northern	376	1.46
e	0-PNW	0	0.00
e	1-S. California	138	0.56
e	2-Northern	276	1.52
o	0-PNW	0	0.00
o	1-Winnipeg	111	0.59
o	2-Northern	326	1.67
u	0-PNW	0	0.00
u	1-Winnipeg	152	0.77
u	2-Western N. Carolina	323	1.70

Table 6

Formants and bandwidths of the vowel stimuli

Vowel	F1	F2	F3	F4	B1	B2	B3	B4
æ-0	635	1579	2279	3500	71	50	170	200
æ-1	675	1750	2504	3500	73	69	230	200
æ-2	588	1952	2700	3500	66	101	284	200
ε-0	483	1834	2504	3500	59	92	232	200
ε-1	458	1698	2329	3500	58	76	185	200
ε-2	630	1600	2301	3500	71	53	176	200
o-0	423	1445	2146	3500	59	60	120	200
o-1	459	1340	2213	3500	65	63	117	200
o-2	469	1122	2300	3500	74	72	99	200
u-0	313	1176	2158	3500	60	74	77	200
u-1	313	1328	2102	3500	55	68	91	200
u-2	390	1490	2104	3500	55	60	118	200

Table 7

ANOVA results for effect of distance by SNR by vowel

	Df	F value	Pr(>F)
Vowel	3	6.261	<0.005
SNR	2	0.330	0.719
Distance	2	180.762	<0.005
Vowel:SNR	6	1.040	0.398
Vowel:Distance	6	12.895	<0.005
SNR:Distance	4	2.909	0.021
Vowel:SNR:Distance	12	1.343	0.190
Error	504		