

Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English

Robert Allen Fox and Ewa Jacewicz^{a)}

Department of Speech and Hearing Science, The Ohio State University, 1070 Carmack Road, Columbus, Ohio 43210, USA

(Received 28 June 2016; revised 25 January 2017; accepted 7 February 2017; published online 31 July 2017)

Vowel space area (VSA) calculated on the basis of corner vowels has emerged as a metric for the study of regional variation, speech intelligibility and speech development. This paper gives an evaluation of the basic assumptions underlying both the concept of the vowel space and the utility of the VSA in making cross-dialectal and sound change comparisons. Using cross-generational data from 135 female speakers representing three distinct dialects of American English, the first step was to establish that the vowel quadrilateral fails as a metric in the context of dialect variation. The next step was to examine the efficacy of more complete assessments of VSA represented by the convex hull and the concave hull. Despite the improvement over the quadrilateral, both metrics yielded inconsistent estimates of VSA. This paper then explores the possibility that regional variation can be characterized more effectively if formant dynamics and the resulting spectral overlap were also considered in defining the space. The proposed formant density approach showed that the working space may be common to all dialects but the differences are in the internal distribution of spectral density regions that define dialect-specific “usage” of the acoustic space. The dialect-inherent distribution of high and low density regions is largely shaped by sound change.

© 2017 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4991021>]

[CGC]

Pages: 444–459

I. INTRODUCTION

There is a long tradition in acoustic phonetic research to categorize vowels with reference to their relative positions in a two-dimensional plane. The two acoustic dimensions, the first (F1) and the second formant frequency (F2), have been generally accepted as representing the basic high/low and front/back articulatory dimensions, corresponding generally to tongue and vocal tract configurations during vowel production (Stevens and House, 1955). More recent work exploring the relationship between the acoustic dimensions and tongue kinematics provided further evidence that F1 reflects a relatively good approximation of tongue height (and openness of the vocal tract), whereas F2 is related in more complex ways to tongue variations in both dimensions, height and advancement (Lee *et al.*, 2016). Positioned in the two-dimensional plane, the “point” or “corner” vowels have received special consideration as representing the *periphery* of the vowel system. Presumably, a language-specific acoustic vowel space formed by these corner vowels encompasses the remaining vowels in the system and the area generated from F1 and F2 measurements reflects the articulatory working space (Neel, 2008).

In American English, vowel space area (VSA) computed on the basis of the corner vowels—having the geometric shape of either a triangle (/i, u, α/) or a quadrilateral (/i, u, α, æ/)—has been utilized in both basic and clinical research. As a metric, VSA can characterize variations in speaking style, speech development, and speech disorders

reasonably well. In particular, VSA tends to be reduced in casual speaking style and expanded in more careful intelligibility-enhancing productions such as when speaking to an infant or a hearing-impaired individual (Ferguson and Quené, 2014; Kondaurova *et al.*, 2012; Liu *et al.*, 2003). A general reduction in the size of the VSA occurs during speech-language development in children as a function of the developmental increase in vocal tract length. The development-related reduction of the VSA has been measured primarily cross-sectionally (Flipsen and Lee, 2012; Vorperian and Kent, 2007) as longitudinal studies are still rare (McGowan *et al.*, 2014). In speech-language pathology, vowel-space compression (relative to healthy controls), usually correlated with reduced intelligibility, is often related to impairment of articulatory function in disorders such as dysarthria (Higgins and Hodge, 2002; Weismer *et al.*, 2001) and Down syndrome (Bunton and Leddy, 2011).

Importantly, the observed compression vs expansion of the vowel space as a function of speaking style, speech-language development and impairment has been of the *within-talker* variety. That is, talkers enlarge their vowel spaces when speaking clearly and slowly, and produce smaller-sized spaces when speaking casually and at a faster rate (Bradlow *et al.*, 1996; Fourakis, 1991). This systematic variation, also exhibited in disorders (Lam and Tjaden, 2016; Tjaden *et al.*, 2013), influences the dispersion of vowels within the system so that vowels become more peripheral and acoustically distinct in the larger vowel space, and more centralized and overlapped in a smaller space. Command of this underlying articulatory behavior appears to be still maturing in typically developing older

^{a)}Electronic mail: jacewicz.1@osu.edu

children and adolescents, who otherwise reduce their VSA with age (Pettinato *et al.*, 2016).

VSA has also been successful in making *cross-linguistic* comparisons to characterize cross-language differences in adults and children (Al-Tamimi and Ferragne, 2005; Chung *et al.*, 2012) and to document the development of two distinct vowel systems in bilinguals (Yang *et al.*, 2015). However, the effectiveness of the VSA metric is limited in the context of regional variation, where neither the within-dialect variations nor across-dialect differences can be adequately characterized by an area encompassed by a small number of peripheral vowels. As shown previously (Fox and Jacewicz, 2008; Jacewicz *et al.*, 2007), inadequacies of existing vowel space estimates are significant. It is the case that the vowel triangle, although still used as a metric (Chung *et al.*, 2012; Liu *et al.*, 2003), severely underestimates the actual size of a working vowel space in American English. Recognizing this limitation, the four vowels /i, u, a, æ/ have increasingly been used in the literature to define the corners (e.g., Vorperian and Kent, 2007). However, although the resulting quadrilateral space is practical and convenient, it also fails as a metric in the assessment of both within-dialect (e.g., related to sound change) and between-dialect variation because, in some dialects, many vowels are found outside the quadrilateral and in other dialects the quadrilateral includes areas in which no vowels are found (Fox and Jacewicz, 2008).

The limited utility of the existing VSA metrics in the face of extensive flexibility in the American English vowel system as a function of both regional dialect and cross-generational sound change motivated the current study. In this paper, our efforts are directed (1) at revisiting the assumption that the size of the VSA is to be geometrically constrained by a specific (and relatively limited) number of distinct vowel categories, and (2) at redefining the vowel space by utilizing spectral content that is variable across vowels and dialects. We expect these current data-driven explorations to stimulate further interest in statistical modeling and development of improved algorithms for estimation of VSA in regional dialects.

Interest in the development of a more sensitive methodology to characterize the working vowel space has been growing in recent years. It reflects both the changing research focus from the study of speech phenomena produced under controlled laboratory conditions as opposed to spontaneous productions, and the rapid growth and availability of large corpora of conversational speech. To meet the current demands of analyzing large numbers of vowels produced by many speakers, automated estimations of VSA are becoming increasingly popular. The automated methods are appealing because algorithms can now detect vowel margins (onsets and offsets) and extract formant information from voiced sections of speech, thus reducing the need for the resource-intensive hand segmentation (Aylett and Turk, 2006; Sandoval *et al.*, 2013). However, despite its efficiency, current automated estimations of vowel space follow the same assumptions as the traditional VSA metrics computed from hand-segmented speech. Namely, both approaches conceptualize the vowel space as being shaped by phonologically distinct vowel categories and VSA as

an area bounded by a smaller or greater number of distinct peripheral vowels (Aylett, 1998; Sandoval *et al.*, 2013). However, as will be shown below, VSA computed as a planar convex polygon defined by n corners also fails as a metric in the context of regional variation. Despite the improvement over the quadrilateral, the convex hull area formed by all perimeter vowels in the vowel system still does not represent well the boundaries of a dialect-specific space.

The problems with defining the vowel space and its differential use by regional dialects has led us to rethink the basic idea of positional distinctiveness of vowels in the two-dimensional plane. In previous work, we demonstrated that dialects of American English differ in the way they utilize the dynamic vowel structure (Fox and Jacewicz, 2009; Jacewicz *et al.*, 2011a). Sampling of formant trajectories at multiple time points and not only at the vowel's center showed that the nature and amount of spectral change over time varies for individual vowels and that this variation is also dialect-specific. This insight has not yet been implemented in existing VSA computation methods. In fact, diphthongal vowels tend to be avoided in automatic spectral analysis in large speech corpora, mostly due to their moving spectral targets and difficulties in estimating the steady state formant values within the vowel.

The current study explores the possibility that regional variation in the acoustic vowel space—both within- and between-dialects—can be characterized more effectively if spectral characteristics represented in formant dynamics are also considered in defining the space. By necessity, such vowel-inherent spectral changes, operationally termed VISC (Morrison and Assmann, 2013; Nearey and Assmann, 1986), create regions of spectral concentration (or spectral overlap) in the acoustic space. We hypothesize that if dialects differ in their use of spectral dynamics, they must also differ in the distribution of the regions of spectral overlap created by these spectral changes. Presumably, cross-dialectal use of the vowel space has more to do with dialect-specific utilization of the regions of spectral overlap rather than with spectral separation of individual vowel categories. This possibility is explored in the current study on the basis of empirical data from a large corpus of regional variation in American English.

II. METHODS

Speech samples were selected from a large cross-dialectal corpus of American English collected to elicit variable vowel productions across several generations of speakers. Our previous work found significant dialectal and generational differences in the spectral and temporal structures of vowels (cf. Fox and Jacewicz, 2009; Jacewicz *et al.*, 2011a,b). In the current study, a subset of this speech material was utilized for the analysis of VSAs. Relevant descriptions pertaining to spectral characteristics of the vowels examined here—including their dynamic formant patterns—can be found in Jacewicz *et al.* (2011a).

A. Speakers

Recordings of 135 female speakers were selected. Each speaker was born, raised and resided in one of three distinct dialect regions in the United States: 45 were from Western North Carolina (NC) (Jackson County) and spoke the local Southern variety of American English typical of Inland South, 45 were from Central Ohio (OH) (Columbus and suburbs) and spoke the Midland variety, and 45 were from Southeastern Wisconsin (WI) (Madison area) and spoke the Midwestern English typical of Inland North. The participants represented four generations of speakers from each respective speech community: children (A0) ranging in age from 8 to 14 years and three groups of adults who, based on their age, could be parents of each successive generation, 27–47 (A1), 50–65 (A2), and 70–91 years (A3). Details about the characteristics of each group are summarized in Table I. The four age groups created the desirable condition for detection of cross-generational sound change in apparent time within the theoretical framework of Labov’s model of transmission and incrementation (Labov, 2001). Accordingly, transmission of the vowel properties that define the acoustic vowel space is the product of language acquisition by children and is implemented via transfer of features from each older to each younger generation. The participants were recorded in years 2005–2009.

B. Stimuli and procedure

Because of the exploratory nature of this study, only citation form vowels in the /hVd/ context were used for VSA computations. Although the citation form vowels exhibit far less of the contextual acoustic variability found in rapid speech, they provide important information about the inherent dynamic vowel structure, which is of immediate interest here. Admittedly, spectral dynamics undergo substantial alterations as a function of both consonant environment and temporal and prosodic variations, but this extensive variability can compromise detection of cross-generational changes related to sound change. For example, in temporally reduced vowels, consonant context may dominate formant trajectory

TABLE I. Group characteristics of study participants as a function of dialect and age (in years).

Dialect	Age group	Number of speakers	Age range	Age mean (standard deviation)
North Carolina	A0	12	9–14	11.2 (1.6)
	A1	12	33–47	39.8 (5.6)
	A2	12	53–62	58.3 (3.4)
	A3	9	70–91	77.0 (8.0)
Ohio	A0	12	9–13	10.1 (1.5)
	A1	12	28–43	38.8 (3.9)
	A2	12	50–63	55.9 (3.7)
	A3	9	71–88	77.4 (6.0)
Wisconsin	A0	12	8–12	9.3 (1.3)
	A1	12	27–47	39.4 (6.0)
	A2	12	51–65	59.2 (4.4)
	A3	9	72–90	79.0 (6.6)

patterns, obscuring the “inherent” aspect of formant change (see Nearey, 2013, for modeling efforts incorporating effects of consonantal contexts).

The speech samples were obtained using a common data collection protocol at all three testing sites in NC, OH and WI. Each participant was seated in front of a computer monitor and read the randomly presented prompts *heed, hid, hayed, head, had, hod, heard, hawed, hoed, hood, who’d, hide, hoyd, howed*, corresponding to 14 vowels /i, I, e, ε, æ, α, ɜ, ɔ, o, u, u, aI, oI, aU/. The items were read in isolation with the aim of eliciting citation form vowels and producing a highly homogeneous sample. Three repetitions of each token were collected for a total of 5670 vowels analyzed in the study (14 × 3 × 135). The tokens were recorded using a head-mounted Shure SM10A dynamic microphone and digitized directly onto a hard disk drive at a 44.1-kHz sampling rate with 16-bit quantization. The experiment was controlled by a custom program in MATLAB (Mathworks, Natick, MA).

C. Formant analysis

Prior to acoustic analysis, the digitized tokens were downsampled to 11.025 kHz. Vowel onsets and offsets were located by hand and defined using standard segmentation criteria (details in Fox and Jacewicz, 2009). The first two formants were sampled at the times corresponding to 20%-80% of vowel duration at increments of 15% (the 20%-35%-50%-65%-80%-points in the vowel). F1 and F2 values were based on linear predictive coding (LPC) and were extracted automatically using MATLAB. A 25-ms Hanning window was centered at each temporal point. All formant values were checked and hand-corrected if necessary using the formant tracking option in TF32 (Milenkovic, 2003) and adjusting the number of LPC coefficients. To ensure reliability of all measurements, a separate program written in MATLAB displayed the numerical values along with the fast Fourier transform (FFT) and LPC spectra and wideband spectrogram of the vowel. This reliability check was done on 100% of the tokens. Any disagreements in the analysis were resolved between the authors and hand-corrected prior to data processing.

D. Vowel space area computations

Two approaches to the assessment of VSA were tested in the current study in order to evaluate their efficacy in characterizing cross-dialectal and cross-generational variation.

The first approach is based on the traditional computation methodology assuming distinct vowel categories and calculating VSA as the area of a polygon defined by a set of corner vowels. The mean F1/F2 values at 50%-point for each of the corner vowels were used to compute the areas of (a) the quadrilateral formed by the vowels /i, u, α, æ/, (b) the convex hull, and (c) the concave hull. All computations were done in MATLAB using the “polyarea” function. This function computes the area of a polygon specified by the coordinates of its vertices, in this case, the coordinates of the perimeter vowels. Similar results could be produced by dividing the polygon into component triangles and using Heron’s formula

to determine the area of each triangle and summing the resulting areas (Fox and Jacewicz, 2008).

The area of a convex polygon with n corners, the convex hull, represents the spatial boundaries defined by all perimeter vowels used by each dialect and generation. The exact number of these “corner” vowels is expected to vary as a function of either variable (or both). The convex hull metric tends to maximize the shape of the vowel space and is viewed as a more complete assessment of VSA than the quadrilateral (Sandoval *et al.*, 2013). The area of a concave polygon with n corners, the concave hull, was defined by all perimeter vowels assuming that, unlike in a convex polygon, the interior angles can be greater than 180° . The concave hull is a more conservative metric than the convex hull and tends to eliminate unused regions at the periphery of the vowel space.

Selection of the perimeter vowels was done using a custom MATLAB program which displayed the mean vowel trajectories for each individual speaker and allowed a researcher to determine the perimeter vowels (and their coordinates) for that speaker’s vowel space following three criteria: (1) to include the greatest possible number of vowel midpoints at the periphery of the F1/F2 plane, and to ensure that (2) the interior angles of the vertices connecting the midpoints are $\leq 180^\circ$ for the convex hull and (3) allowed to be $>180^\circ$ for the concave hull. The aim was to define vowel space broadly in the computation of the convex hull and to minimize the amount of “empty space” while including the maximum number of vowel midpoints in the computation of the concave hull. A second researcher reviewed all such selections and any discrepancies were resolved prior to the final decision. This method was chosen because of the recognized difficulties in implementing an algorithmic approach in defining a convex or concave hull in computational geometry (see de Berg *et al.*, 2000).

The second approach departs from the traditional area calculations on the basis of midpoints of individual vowel categories. Rather, the formant space is partitioned into formant density regions that seem to have distinct importance for different dialects or generations of speakers within a particular dialect. This approach assumes that the numeric values along the two quality dimensions, F1 and F2, typically exhibit variations over the course of a vowel’s duration. Inevitably, these time-varying F1/F2 values create spectral overlap of vowels in a common region of the formant space. To create areas of overlap (or formant density regions), F1 and F2 formant frequencies were first measured

at five equidistant time points (20%-35%-50%-65%-80%) of each of the 14 vowels, which excluded immediate influence of consonant transitions. Next, to approximate denser formant sampling, the F1/F2 values between the five data points were estimated by linear interpolation in MATLAB so that each vowel contributed 21 formant “points.” The regions of overlap were created without linking these points to individual vowel categories or to specific measurement locations in the vowel (e.g., onsets, offsets or midpoints); these formant points merely served to indicate that a particular location in the acoustic formant space was “used” by the speaker.

Spatial formant density estimation was done in MATLAB by first deriving a three-dimensional histogram of the underlying distribution of the z -score normalized data points (Lobanov, 1971) contained in vowel trajectories. The normalization was done to minimize differences related to vocal tract lengths of children and adults. The frequency count in separate F1/F2 bins arranged over a two-dimensional F1 by F2 grid was the third dimension. There were 31 bins for each formant dimension (31×31) in the range -3.0 to $+3.0$. The histogram was then smoothed using interpolated mesh grid. To enhance a graphical interpretation of the mesh grid, a two-dimensional contour graph was then derived. The consecutive steps are illustrated in Fig. 1. In this and subsequent figures (as applicable), the z -score values for both F1 and F2 were multiplied by -1 in order for the vowel displays to match the standard IPA configuration (i.e., /i/ in upper left quadrant, /u/ in upper right quadrant).

III. RESULTS

A. Vowel quadrilateral

We begin with the analysis of the quadrilateral VSA. Figure 2 displays the variation in the shape and size of the /i, u, a, æ/ quadrilateral (in red) for the three dialects and four age groups, superimposed on the formant space generated from dynamic formant trajectories of all individual productions. Reiterating, each individual speaker produced 3 exemplars of each vowel category for a total of 42. The dynamic formant trajectories (in blue) represent each individual production from each speaker. Each formant trajectory contributed 21 F1/F2 points.

The quadrilateral VSAs (in squared z -scores) were analyzed using two-way analysis of variance (ANOVA) with the between-subject factors dialect and age group. Following the significance of a main effect, Scheffé’s multiple comparisons were used as *post hoc* tests. In addition to indicating

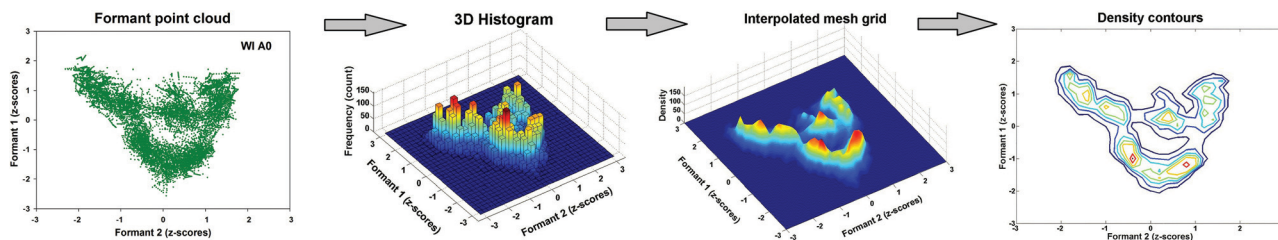


FIG. 1. Deriving spatial density areas in the formant space: (a) a scatterplot of the distribution of F1/F2 values from all individual productions, (b) the corresponding three-dimensional histogram showing the frequency (count) in separate F1 by F2 bins, (c) grid-based interpolation of the data (smoothing) in a mesh grid format, and (d) computation of density contours from the gridded data.

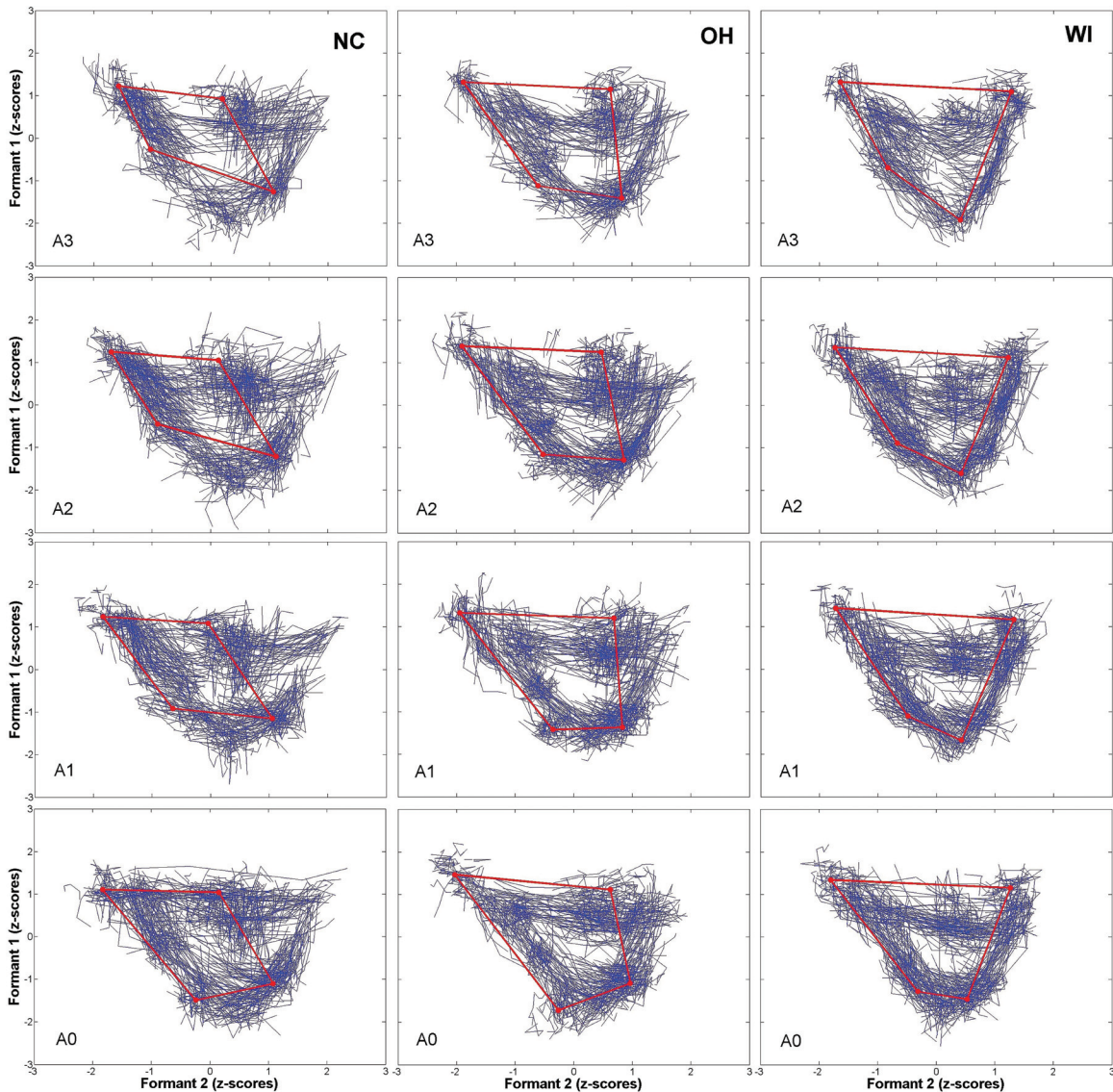


FIG. 2. Average quadrilateral vowel spaces encompassing four vowels /i, u, a, æ/ (clockwise from the upper left corner) superimposed on the dynamic formant trajectories from all individual productions. The displays are for three dialects (NC, OH, WI) and four age groups representing four generations of speakers ranging from old adults (A3) to children (A0).

p -values for specific F -tests, partial eta-squared (η_p^2) values are provided as a measure of effect size. Partial eta-squared values for an experimental factor represent the proportion of total variation attributable to that factor, excluding other factors from the non-error variation (Pierce *et al.*, 2004). The analyses were performed with IBM SPSS Statistics, version 21. It needs to be emphasized that age is treated in this study as a social and not a biological variable (Eckert, 1997). That is, the biological age, per se, is of no direct interest unless it carries a social meaning, in this case generational group membership.

The main effect of dialect was significant [$F(2, 123) = 104.06, p < 0.001, \eta_p^2 = 0.629$]. The NC space was significantly smaller ($M = 3.65, p < 0.001$) than either OH ($M = 5.05$) or WI ($M = 5.29$), and the latter two did not differ significantly from one another (Scheffé). There was also a significant main effect of age group [$F(3, 123) = 5.93, p = 0.001, \eta_p^2 = 0.126$]. Children had the largest VSAs and the spaces of each successive older generation were progressively smaller

(the means were 4.98, 4.73, 4.50 and 4.44, respectively). Multiple comparisons showed that the children's spaces differed significantly only from the two oldest adult groups, A2 ($p = 0.004$) and A3 ($p = 0.002$) and was not significantly different from the space of young adults (A1). There were no significant differences between any of the adult pairs.

Importantly, the interaction between dialect and age group was not significant, suggesting that the obtained cross-generational pattern is common to all three dialects, irrespective of the relative size and shape of the quadrilaterals. Tentatively, we can interpret these results as a manifestation of cross-generational sound change, which triggered a significant expansion of VSA in younger speakers and children, predominantly due to the lowering of the vowel /æ/ in the lower left corner of the quadrilateral. We reported on this common sound change across American English dialects in our previous work analyzing productions from 239 speakers (Jacewicz *et al.*, 2011a). Crucially, this common new development in the dialects examined was possible to detect when

all speakers produced common speech material so that variable contextual influences on spectral structure (whether related to immediate consonant or prosodic environments) could be minimized. It is also of note that the nature and degree of /æ/-lowering varies among the dialects. Applying the simple criterion proposed by Thomas (2001) that if /æ/ has a lower F1 than /a/, /æ/ is raised, we still find /æ/-raising in WI children despite the cross-generational lowering of the vowel in this dialect. No corresponding raising can be found in either NC or OH children.

Although the sound change-based interpretation of the quadrilateral VSA is intuitively appealing, we nonetheless cannot ignore the fact that substantial parts of the actual working vowel space were excluded from analysis, particularly in NC dialect (compare Fig. 2). For this reason, a convex polygon rather than the quadrilateral promises to be a better approach to VSA estimation as it should expand the spatial boundaries and thus maximize the area.

B. Convex hull

Convex areas were calculated on the basis of the midpoint values of F1 and F2 of the perimeter vowels. As shown in Fig. 3, the exact number of the perimeter vowels can vary as a function of dialect and age group, reflecting dialect-related positional differences in vowel dispersion pattern on one hand and cross-generational variation related to sound change on the other. For example, only three midpoints for the vowels /æ, aɪ, a, / in the lower part of the space were utilized in NC A3 group whereas their number increased to five /æ, aɪ, aʊ, a, ə / in NC A0 group.

A two-way ANOVA of VSA returned a significant main effect of dialect [$F(2, 123) = 3.85, p = 0.024, \eta_p^2 = 0.059$]. The convex area of the NC space increased relative to the area of the quadrilateral and was now not significantly different from the OH space (means were 5.82 and 5.87, respectively) but, despite the increase, the NC space was still significantly smaller than the WI space ($M = 6.08, p = 0.035$,

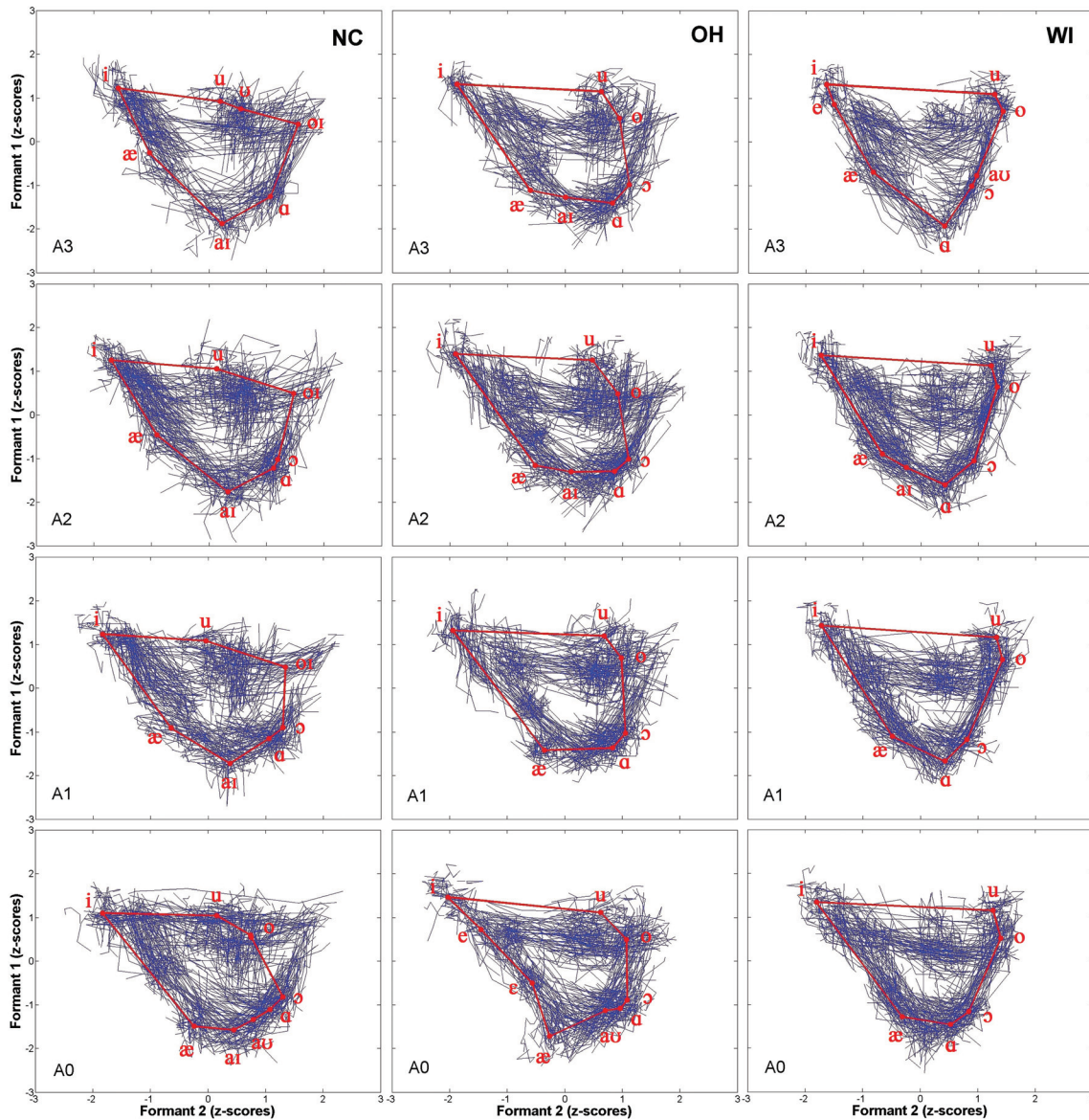


FIG. 3. Average convex hull spaces superimposed on the dynamic formant trajectories from all individual productions. The displays are for three dialects (NC, OH, WI) and four age groups representing four generations of speakers ranging from old adults (A3) to children (A0).

Scheffé). Unlike for the quadrilateral, the effect of age group was not significant.

A weak but significant interaction arose between dialect and age group [$F(6, 123) = 2.24, p = 0.044, \eta_p^2 = 0.099$]. Since our main interest was in the cross-generational change in the VSAs in each dialect, the interaction was explored using multiple comparisons. For OH, significant differences were found only between children and A2 adults ($p = 0.047$) and between children and A3 adults ($p = 0.042$). In each case, the children's VSAs were comparatively larger. There were no significant differences among any of the WI groups. For NC, there was only one significant difference between children and A1 adults ($p = 0.009$). As can be seen in Fig. 4, the general pattern revealed by this interaction was that children's mean VSA was larger when compared with the adults in OH dialect and, conversely, it was smaller relative to the adults in NC variety. However, significant differences between children and adults were obtained only for selected adult groups. No cross-generational changes in VSAs could be observed for WI. These results need to be interpreted with caution considering the weakness of the dialect by age group interaction. In particular, the above comparisons were significant only in LSD tests and were not significant using the more conservative Scheffé's procedure.

C. Concave hull

The third approach to calculating vowel space as the area of a polygon, the concave hull, used the midpoints of perimeter vowels allowing interior angles greater than 180° . The number of vowels also varied with dialect and age group. As shown in Fig. 5, this conservative approach was able to eliminate some of the unused regions at the periphery (such as for WI A3) but still excluded some working regions (such as for NC A0). The same set of statistical analyses was performed as for the convex hull.

The results of two-way ANOVA showed a significant main effect of dialect [$F(2, 123) = 19.58, p < 0.001, \eta_p^2 = 0.242$]. Surprisingly, the NC space was significantly larger ($M = 5.21, p < 0.001, \text{Scheffé}$) than either OH or WI ($M = 4.66$ and $M = 4.42$, respectively) whereas the latter two

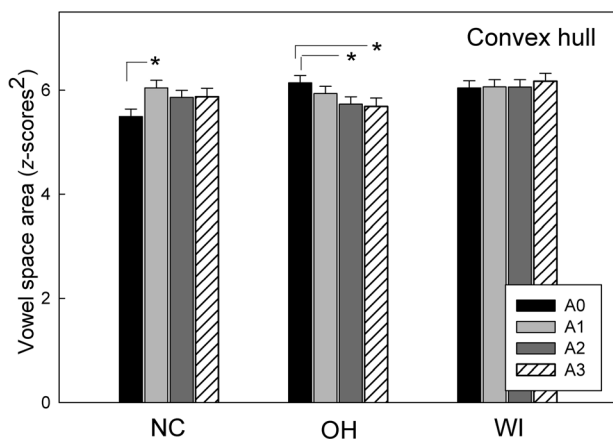


FIG. 4. Pairwise comparisons for convex hull vowel space area as a function of dialect (NC, OH, WI) and generation (A0-A3) are illustrated. The error bars represent one standard error.

did not differ significantly one from the other. The main effect of age group was not significant but there was a significant dialect by age group interaction [$F(6, 123) = 3.75, p = 0.002, \eta_p^2 = 0.155$].

The comparatively stronger interaction, displayed in Fig. 6, was explored using the same set of *post hoc* analyses as for the convex hull (including LSD tests). For OH, a significant difference was found only between children and A2 adults ($p = 0.004$). There was also one significant difference for WI: A1 adults had significantly smaller VSAs than A2 adults ($p = 0.041$). For NC, significant differences were between children and A1 adults ($p = 0.024$) and between children and A2 adults ($p = 0.034$). The difference between children and A3 adults narrowly missed significance ($p = 0.053$) but in each case, the children's VSAs were comparatively smaller. The general pattern for the concave hull was in part consistent with that for the convex hull. In particular, the VSAs in OH children were the largest among all OH groups, whereas the VSAs in NC children were the smallest. However, unlike for the convex hull, there were notable differences among adult groups in OH and WI as illustrated in Fig. 6.

D. Summary and discussion of the polygon areas

The quadrilateral provided a seemingly unified and straightforward account of variation in VSA as a function of dialect and age group. One could reason that some dialects utilize larger or smaller vowel spaces than others and that dialect-specific VSAs have been steadily expanding over generations due to changes in the relative positions of corner vowels (notably /æ/) in the process of sound change. However, positioned in a more complete display of formant trajectory points, the quadrilateral defined by the four "corner" vowels turned out to be a poor representation of the actual working vowel space. Despite its simplicity and natural appeal, the quadrilateral failed as a metric of VSA as it did not appropriately define the periphery of the acoustic space utilized in each dialect.

The convex polygon offered only a partial solution. The convex hull did increase areas by including more perimeter vowels. For example, the severely underestimated quadrilateral VSA in NC increased from 3.65 to 5.82 (in squared z-scores), approximating the spaces of the two other dialects. The expansion of VSAs in each dialect is in agreement with previous work showing that the convex hull metric resulted in consistently larger VSA estimates (Sandoval *et al.*, 2013). However, notwithstanding the improvement over the quadrilateral, the areas defined by the convex hull still excluded regions actually utilized and so did not represent well the periphery of a dialect-specific space. While expanding the areas, the convex hull tended to minimize the differences among dialects, complicating a meaningful interpretation of the relation between dialect-specific dispersion of vowels in the F1 by F2 plane and the size of VSA. We also found contradicting results as a function of age group so that OH VSAs tended to expand with each younger generation (a result consistent with the quadrilateral), but NC VSAs tended to decrease and WI VSAs remained unchanged.

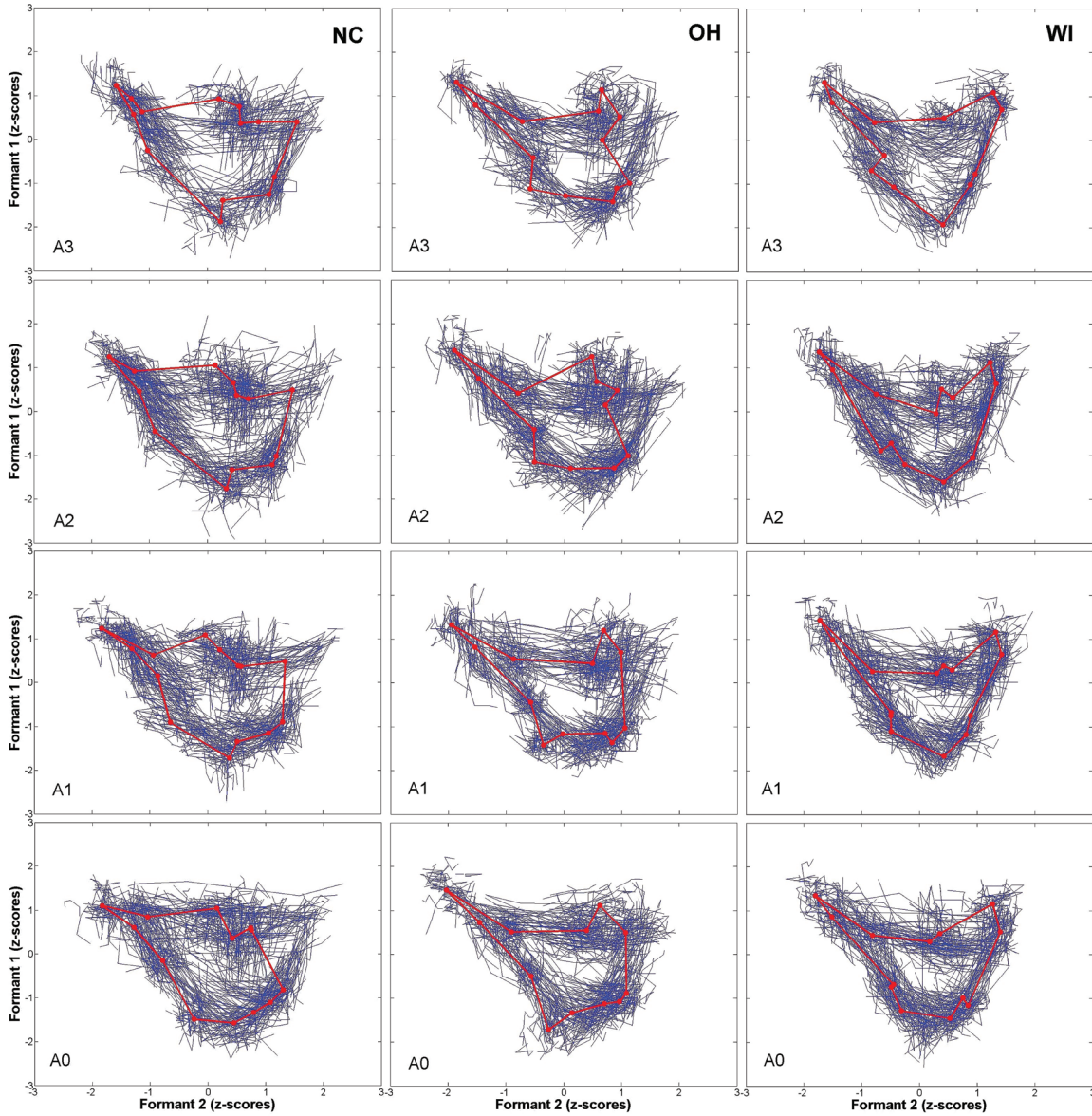


FIG. 5. Average concave hull spaces superimposed on the dynamic formant trajectories from all individual productions. The number of perimeter vowels per space ranges from 12 (OH A0-A2; WI A3) to 14 (NC A1, A3; WI A1). The displays are for three dialects (NC, OH, WI) and four age groups representing four generations of speakers ranging from old adults (A3) to children (A0).

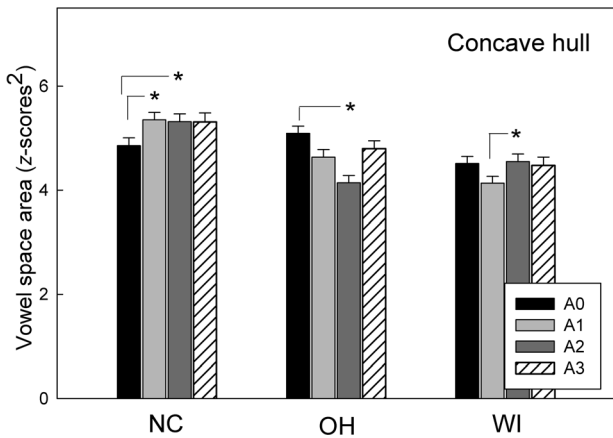


FIG. 6. Pairwise comparisons for concave hull vowel space area as a function of dialect (NC, OH, WI) and generation (A0-A3) are illustrated. The error bars represent one standard error.

Exploring the possibility that the inherently liberal approach of convex geometry in defining the boundaries may obscure the dialectal and generational differences, we used the concave hull metric. The use of concavity to determine the perimeter vowels—which, in turn, redefine the boundaries—generates a more conservative space, holding the promise to improve the dialect- and generation-specific shape of the vowel space. In particular, while a convex hull approximates the shape, a concave hull brings in details, carrying more “local” than “global” information about vowel dispersion patterns. As has been known for years, high-curvature (concave) shapes supply more structural information than low-curvature (convex) shapes (e.g., Resnikoff, 1985). The detailed concave representation of the vowel space produced yet different results. The concavity resulted in an expansion of the NC working space and reduction of both OH and WI, an outcome opposite to that of the quadrilateral and, to some extent, of the convex hull.

The first aim of the current study was to revisit the assumption that the size of the VSA is geometrically constrained by a specific number of distinct vowel categories. The inconsistent results for the polygon areas explored here invite the question of whether a calculated VSA is capable of characterizing not only the dialectal differences in the size of a working vowel space but also of illuminating dialect-specific patterns of variation within a system. In particular, is there a meaningful relationship between the size of VSA and a dialect-specific vowel dispersion pattern related to vowel shifts (rotations) reflecting a specific sound change?

The three regional varieties selected for this study exemplify three different configurations within each vowel system. According to sociolinguistic studies, the southern NC system is affected by the Southern Shift, whose hallmark features are monophthongization of /aɪ/, acoustic reversals of /e-ɛ/ and /i-ɪ/, and raising of /æ/ (Labov *et al.*, 2006). An extensive fronting of /u/ is another common feature of southern dialects. The northern WI system exhibits an operation of a different chain shift known as the Northern Cities Shift, whose stages involve raising of /æ/, fronting of /ɑ/, lowering and fronting of /ɔ/, backing of /ʌ/ and /ɛ/, and lowering of /ɪ/ (Gordon, 2001). Finally, the vowel system in central OH has traditionally been regarded as not participating in any systematic shift (Labov *et al.*, 2006) other than being affected by low back merger of /ɑ/ and /ɔ/ (i.e., a suspension of phonemic contrast between “taught” and “tot”). Recent reports suggest that the system may be in fact influenced by a form of the Canadian Shift with a systematic retraction of /æ, ɛ, ɪ/ (Durian *et al.*, 2010).

Importantly, these dialect-specific sound changes may take generations to complete. An incremental change in the vowel system of each successive generation results from reorganization of the initially acquired system by children. According to the Labov’s model, once the linguistic system stabilizes in young adulthood, no structural changes to its vowel configuration are expected if these adults stay in the same speech community (Labov, 2007). How much of the cross-generational changes are reflected in the calculated dialect-inherent polygon areas in Figs. 2, 3, and 5?

The quadrilateral area, with all its limitations, certainly reflects the fact that the vowel /æ/ has significantly lowered in the acoustic space in children in each dialect (and, to some extent, in the younger adults) and the remaining “corner” vowels have remained relatively stable throughout generations. This recent change in the pronunciation of young speakers (Jacewicz *et al.*, 2011a) has not yet been widely observed in the literature due to a general lack of cross-dialectal data from children. But this is not to say that the lowered /æ/ is a common pronunciation pattern in NC and WI across all speakers. Indisputably, older speakers still produce more conservative raised forms of /æ/ (compare A3 and A2 groups in NC and WI). Indeed, it is well known that speakers in northwestern WI raise their /æ/ even more in pre-velar contexts (such as in “bag”) although the raising in other consonant environments appears to cease among younger speakers, particularly in women (Benson *et al.*, 2011). In terms of the southern variant of /æ/ in NC, its lowering in

the A1 and A0 groups reflects the fact that the Southern Shift has been receding in younger generations across the South although older adults still produce the most conservative patterns (Fridland, 2012). It needs to be pointed out the current sample does not include individuals in their early twenties. Productions from the “college-years generation” could be informative as to the possible earlier appearance of the /æ/-lowering in these speech communities.

As already discussed, the convex hull area, despite including a greater number of peripheral vowels, still did not represent well the outer boundaries of a dialect-specific working space. In terms of the cross-generational sound change, a fine-tuning of the space was only apparent in OH. In particular, we could observe a retraction of /æ, ɛ/ in OH A0 group along with a low back merger of /ɑ/ and /ɔ/. These changes, along with the lowering of /æ/ as already revealed in the quadrilateral, contributed to a significant increase of the calculated VSA relative to the older generations. No relation between sound change and VSA could be found for the two other dialects. The configurations of vowels affected by the Northern Cities Shift in WI and by the Southern Shift in NC did not contribute to changes in the shape of outer boundaries.

The concave hull approach resulted in reductions of dialect-specific VSAs relative to the areas defined by the convex hull. With respect to sound change, this approach was of some benefit only to NC dialect. In particular, the concave hull areas revealed further details in NC space related to the operation of the Southern Shift in the oldest speakers (A3) as evident in the proximity of the midpoints of the /i, ɪ, e, ɛ/ cluster. Reorganization of the upper front cluster could also be detected across subsequent generations, tracking how the Shift receded with each younger speaker group. No additional sound change-related vowel configurations in OH and WI could be captured by the VSA defined by the concave hull.

It needs to be pointed out that the detailed shapes of the concave hull spaces came about, in part, because some diphthongs were also included in the calculations as long as the concavity criterion was met. Relatively little is known about the amount and nature of spectral changes in diphthongs in American English dialects, which is considered a weakness of previous analyses of dialect variation and sound change (Thomas, 2016). Therefore, the current choice of midpoints in diphthongs as corresponding to a particular perimeter vowel was an arbitrary decision in the absence of any other compelling method. A consequence of this decision was that portions of the VSAs in the high back region were excluded from VSA calculations, predominantly in NC and OH dialects. These omitted areas corresponded to later diphthongal portions of /au/ and /o/ and earlier portions of /oi/ in these two dialects. Exclusions of these diphthongal portions in concave hull (and, to a lesser extent, in convex hull) areas were triggered by the progressively more fronted positions of /u/ in OH and NC, respectively. On the contrary, the configuration of monophthongs and diphthongs in the WI space, particularly the far back positions of /u/ and /o/, created an ideal condition for inclusion of the high back region in

VSA calculations. However, the inclusion of the high back region in the convex hull occurred at the expense of also including an unused area in the upper part of WI space. These empty areas were then (rightly) excluded when the concave hull approach was used.

The foregoing discussion of gains and losses for each type of the polygon in relation to dialectal differences and cross-generational sound change leads to the conclusion that each approach is problematic in characterizing a working VSA. The tradeoffs between the convex and concave geometry such as those noted for NC and WI indicate that, ideally, a combination of the two approaches instead of selecting one over the other could possibly eliminate the inconsistencies and define the periphery of a dialect-specific space. We will return to this point in Sec. IV. At present, we conclude that neither of the three polygon-based metrics is able to adequately characterize the complete vowel space and thus none can reliably depict dialectal and generational variations. These obvious limitations motivated the second goal of the paper, that of redefining

the vowel space by utilizing the dialect-inherent pattern of spectral dynamics.

E. Formant density regions

The approach advanced in this study moves away from the view of the vowel space as a polygon area defined by distinct vowel categories. The acoustic space examined here is called the formant space (rather than vowel space). This is because the focus is on the absolute frequency of occurrence (count) of formant points in the F1/F2 plane and frequency distribution of these points (bins) without considering the particular vowel category to which they belong.

Figure 7 shows the areas in the formant space that were actually utilized in each dialect and age group. These spatial areas were derived from frequency distribution of F1/F2 points, whose density determined their size and shape (see Fig. 1 for the consecutive steps). The emphasized contour lines in Fig. 7 represent the low-density boundaries derived from 20 occurrences of a particular F1/F2 point. These low-

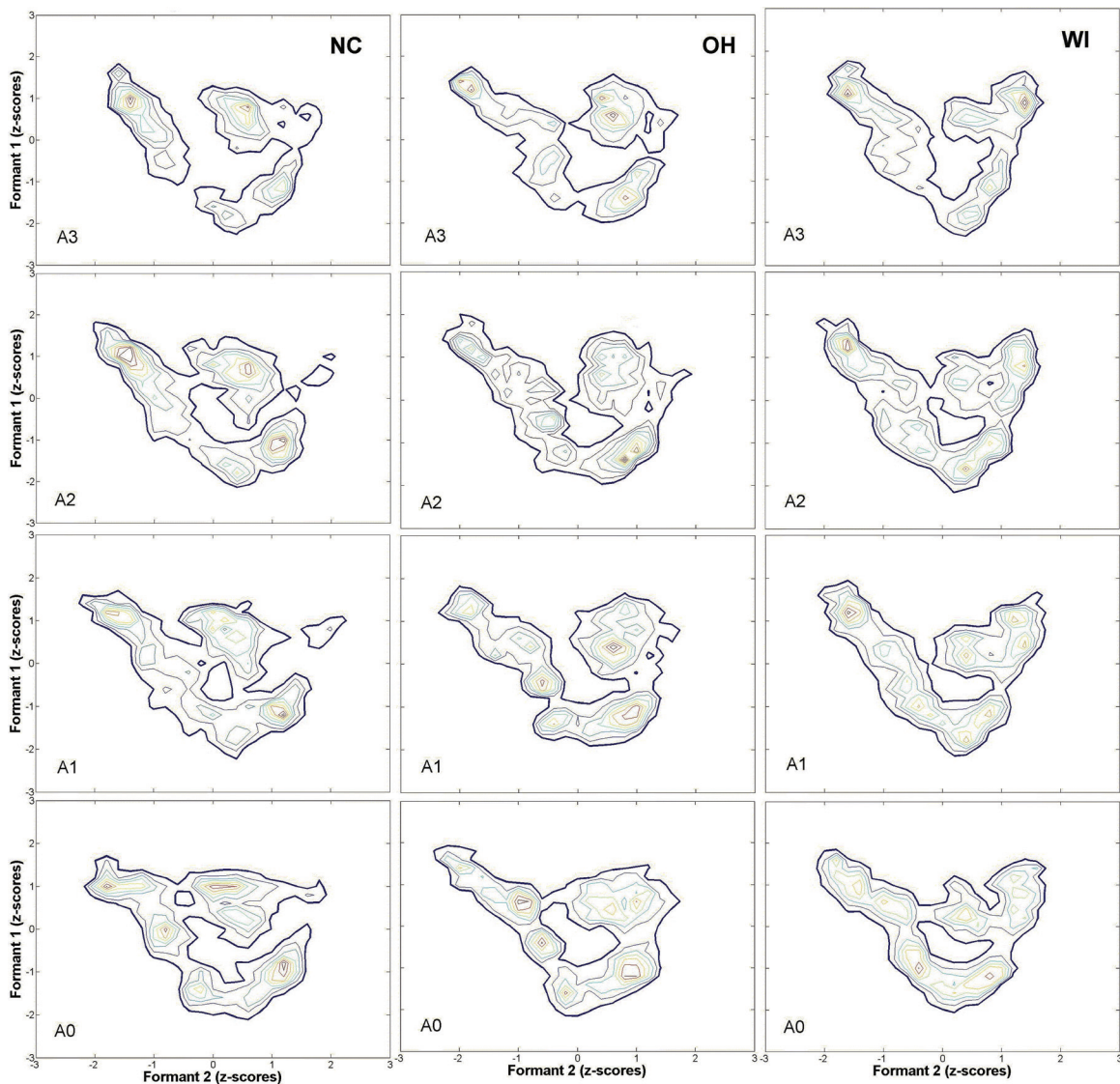


FIG. 7. Formant density contours emphasizing low-density boundaries computed from 20 occurrences of each F1/F2 point. The displays are for three dialects (NC, OH, WI) and four age groups representing four generations of speakers ranging from old adults (A3) to children (A0).

density boundaries approximate the outer boundaries of the utilized formant space as they are located at the functional periphery of the formant space. The boundaries separate the low-density regions from the outermost areas that were only sparsely utilized (frequency of less than 20) and were excluded. The low-density criterion was arbitrarily set to 20 to enable dialectal and generational comparisons using a common benchmark. Some of the inner low-density boundaries, such as those in WI A2 or NC A1 groups, arose because the central vowel /ʌ/ was not included in the stimulus set. With respect to the low-density boundaries, there is an obvious improvement in defining the working space particularly in high central and back regions, which were the most cumbersome parts of the vowel space for application of the polygon area metrics. The lighter lines in Fig. 7 showing the two-dimensional F1 × F2 contours represent a progressively higher frequency count of F1/F2 points, corresponding to increased density of each sub-area. A three-dimensional representation of these two-dimensional graphs in the form of an interpolated mesh grid is shown in Fig. 8.

The density plots in Fig. 8 visualize the concentration of formant points within the spatial areas defined by the low-density boundaries. It is apparent that some parts of the formant space are used more heavily (have greater density of points) than others and the distribution of the high-density regions is dialect- and generation-specific. The densest regions, generated from high frequency of F1/F2 points, have a frequency of more than 80 (again, a common criterion for visual cross-group comparisons) and may correspond to vowel clusters or even to individual vowels. We will now inspect these patterns separately for each dialect.

There are three high-density regions in older NC speakers (A3 and A2). The high front region contains instances of /i, ɪ, e, ε / whose dynamic formant patterns created a substantial spectral overlap. The spectral proximity of this vowel cluster is a hallmark of the Southern Shift. The high back region includes /ʊ, u, o/ and a portion of the diphthong /oi/. The low back region includes /ɑ, ɔ/ and portions of /au/. A reorganization of these density regions takes place in younger speakers (A1) and culminates in children (A0), whose formant space exhibits emergence of six high-density regions owing to the retreating Southern Shift. The six regions reflect the following set of changes: a separation of /ε/ from the high front overlap, lowering of /æ/ which now overlaps with portions of /aɪ, au/, and fronting of /ʊ/ which creates a spectral overlap with portions of /oi/, separating this new region from the high-density region containing the vowel /u/.

The density patterns for the OH groups also evidence cross-generational sound change. Four high-density regions in the older speakers (A3 and A2) increase to six in the younger speakers (A1), reflecting a separation of /i/ and /ε/ from the high front region, lowering of the /æ/, overlap of /ɑ, ɔ/ and portions of /au/, and raising of the diphthong /oi/ to create an overlap with /ʊ, o/. The number of high-density regions increases to seven in children (A0), reflecting a further set of changes: monophthongization of /ɪ/ which results in a high concentration of formant points overlapping with portions of /aɪ, e/, monophthongization and lowering of both

/ε/ and /æ/, merger of /ɑ/ and /ɔ/, and backing of /o/ which creates two high-density regions in a close proximity to one other, the first resulting from an overlap of portions of /o, oi, au/ and the second from the overlap of portions of /ʊ/ and /oi/.

The cross-generational sound change in WI also increased the number of high-density regions to as many as nine in the two younger groups, (A1) and (A0). The increase is driven by both monophthongization and lowering of /ɪ/ and /ε/, as well as /ɑ/-raising, /oi/-raising and lowering of /ʊ/. The emergence of a greater number of density regions in young speakers, predominantly in children, is a new trend common to all three dialects. This pattern of change reflects a dialect-specific reorganization of the acoustic space, which alters the distribution of spectral dynamics. For example, increased monophthongization of selected vowels, including the /æ/, becomes a more wide-spread new development in the vowel systems of young populations as discussed elsewhere (Jacewicz *et al.*, 2011a).

The displays in Fig. 8 are helpful in recognizing that both the positional vowel changes which historically define sound change and systematic changes in their dynamic structure alter the number and the nature of formant density regions in the space. The emergence of new high-density areas may result from category overlap and changes in the amount of formant movement, as some vowels become more monophthongal and some become more diphthongized (such as the increasingly diphthongal realizations of /ai/ in the South) in the process of cross-generational sound transmission. The differential use of the formant space by successive generations reflects the spectral complexity of sound change, which certainly involves more than positional rotations of vowels in a vowel space.

Density plots also offer a useful way to visualize the differences between dialects. By subtracting the histogram matrix of one dialect group from another dialect group, density difference plots can be created (following interpolation and smoothing). Figure 9 displays dialect density differences for the three groups of children (A0), showing positive and negative density contours that represent density regions for each of the two dialects being compared. Comparing the distribution of NC contours with the WI contours, Fig. 9(a) indicates that the two dialects differ not as much in the overall size of the utilized formant space but primarily in the distribution of the high density regions, which tend to complement the two dialects rather than overlap in a common space. We observe that dialectal differences persist in the children's spaces even if the number of density regions increased in these age groups as discussed above. For example, the fronted /u-ʊ/ overlap corresponds to the fronting of these two vowels in the southern dialect (NC) whereas the /u-o/ overlap occupies a far back position in the North (WI). We also find a complementary distribution of density regions in the low back area of the space reflecting the fact that the /ɑ/ is not raised as much in the North as it is in the South, where it is also located further back (note that the negative WI region at the very bottom of the space corresponds to the onsets of /aɪ/ and /au/ in that dialect). Furthermore, there is a clear difference between the dialects in the distribution of

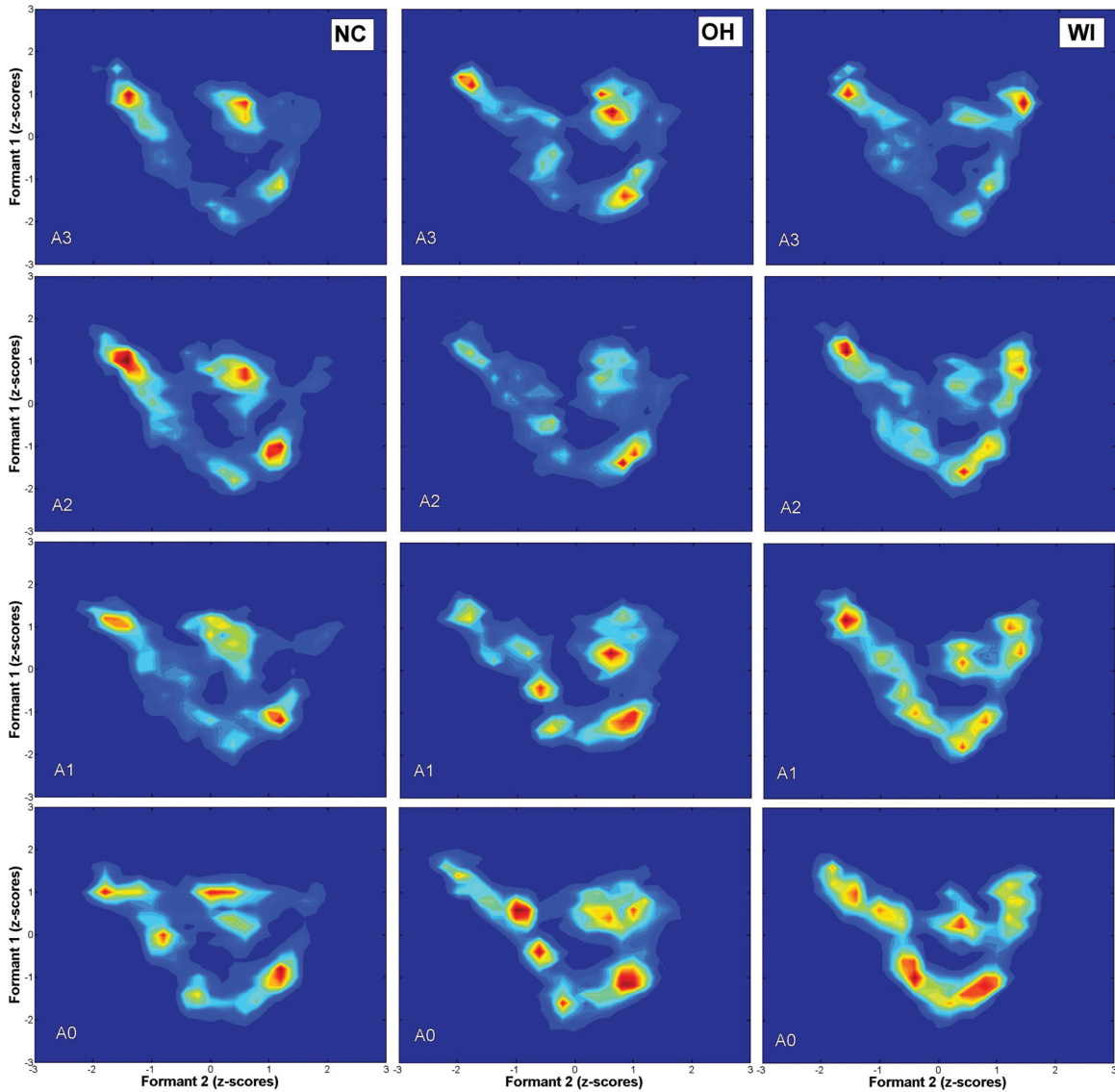


FIG. 8. High-density regions representing concentration of formant points (F1/F2) in the formant space. The areas in light green correspond to the frequency (count) of 80 and the areas on the color spectrum from yellow to dark red reflect increasingly higher frequency of F1/F2 points utilized in each dialect (NC, OH, WI) and in each age group, ranging from old adults (A3) to children (A0).

spectral overlap in the mid-low front region, corresponding to the raising of /æ/ and its proximity to /ɛ/ in the North and to the lowering of /æ/ and its separation from /ɛ/ in the South.

Similar observations can be made when comparing NC with OH in Fig. 9(b). Of note is a complementary distribution of front vowels, including the higher position of /i/, retraction of /ɪ, ɛ/ and lowering of /æ/ in OH relative to NC. The overlapping /u-u/ region in NC is more fronted than that in OH, reflecting the different degrees of u-fronting in the two dialects. Also, the /ɔ-a/ merger in OH occupies a more interior region in the acoustic space relative to the spectral overlap between /ɔ, a, au/ in NC. Last, the comparison between WI and OH in Fig. 9(c) also reveals differences between the two dialects. We find a distinct high density region in the high front part of the WI space, which is created by overlapping instances of /e/ and the offsets of the diphthongs /oɪ/ and /aɪ/. We can also infer that the /ɪ/ in WI

is retracted relative to the OH variant. It is also the case that the OH /æ/ occupies a lower region in the space than the WI /æ/, whose onset frequencies also overlap with the WI variant of /ɛ/. Also, the /ɔ-a/ merger in OH is raised relative to the lower positions of /a/ and /ɔ/ in WI, in accord with reports in the literature that /a/ is fronted out of the low back area as part of the Northern Cities Shift (Labov *et al.*, 2006). Furthermore, the /u-u/ overlap in WI is distinctly far back in the acoustic space relative to both the overlapping /u-u/ region in OH [Fig. 9(c)] and the southern fronting in NC [Fig. 9(a)].

Together, the density-difference plots emphasize dialect-specific use of formant space by children. The dialectal differences in the distribution of density regions, resulting from inherent dispersion of vowels, their inherent dynamic structure, and sound change-related variations imply that American English dialects continue to diverge rather than become more homogeneous.

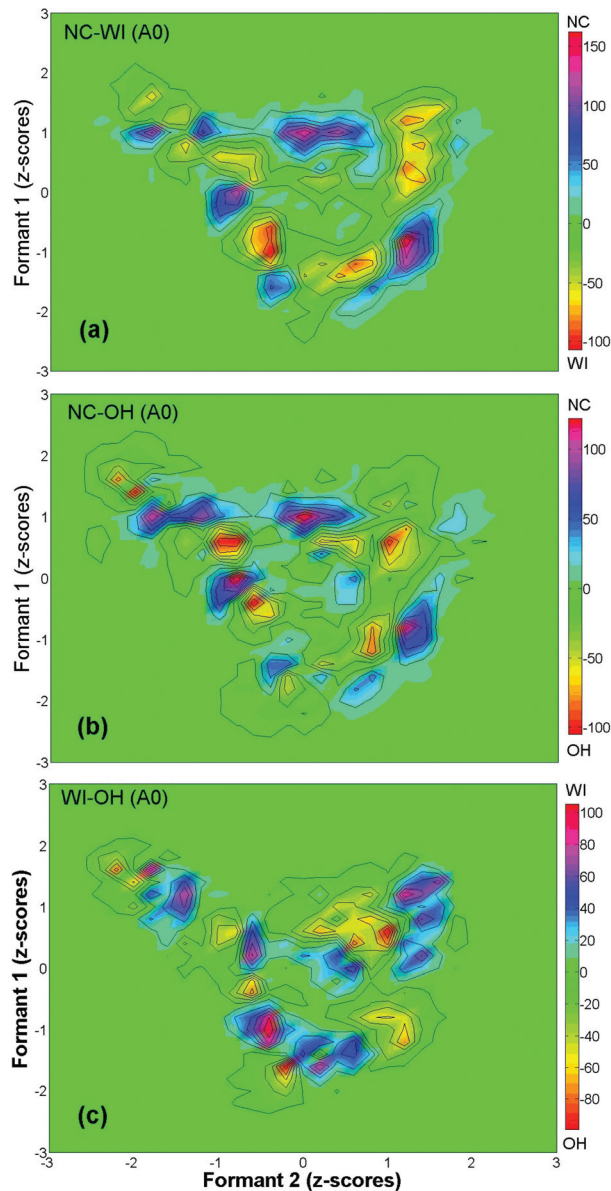


FIG. 9. Density difference plots comparing the distribution of density regions in children (A0) in three dialects: (a) NC with WI, (b) NC with OH, and (c) WI with OH. The positive density contours are color coded on the scale from green to violet red and the negative contours are coded from green to red. The zero-difference is represented by a common background color.

IV. GENERAL DISCUSSION

The current study was motivated by mixed results of our previous work which signaled the limited utility of the vowel quadrilateral in the context of dialect variation (Fox and Jacewicz, 2008; Jacewicz *et al.*, 2007). As a follow up, this paper revisited the concept of the vowel space and reexamined the associated metrics assessing vowel space area. In this pursuit, we first evaluated the effectiveness of polygon area metrics with a larger number of speakers representing three dialects and four age groups. We obtained inconsistent estimates of VSA across the metrics, with the quadrilateral severely underestimating the actual working space, the convex hull minimizing the differences between dialects by including unused areas at the periphery of the space, and the concave hull still excluding areas that have been actually

used. The inconsistencies in estimates of VSA resulting from these approaches were particularly evident in the assessment of cross-generational differences within a dialect. Although statistically significant differences were obtained for a few isolated group comparisons, these differences were only weakly related to dialect-specific sets of vowel changes such as the Southern Shift or the Northern Cities Shift. The few sporadic correspondences between the VSA and the manifestations of a particular shift cannot lead to reliable predictions about a possible relationship between the size of the area and dialect-specific sound change.

The problems with applying polygon geometry to delineate the areas utilized by different dialects lie primarily in defining the outer boundaries of the space and in selecting the most appropriate vertices (or corners) to characterize its shape. We showed that neither convex nor concave geometry could adequately represent a complete “surface” of the space. The main issue is how conservative or how liberal a researcher wants to be in designating particular vowels as polygon vertices. The vowel quadrilateral is, of course, a drastic example of how much of the acoustic space can be excluded from calculations. However, a more liberal approach to the selection of perimeter vowels such as convex hull can still lead to serious misrepresentations of a dialect-inherent working space.

Relatedly, the size of VSA may be influenced by measurement location, that is, by the temporal point in a vowel at which formant frequencies are extracted. In this study, we used vowel midpoints as polygon vertices in part because of a long phonetic tradition, and in part because our previous work showed that the size of VSA calculated on the basis of midpoints tends to be intermediate between an expanded VSA when measured closer to vowel onset (the 35%-point), and a reduced VSA when measured closer to vowel offset (the 80%-point) (Fox and Jacewicz, 2008). But this approach raises serious concerns related to the very concept of the vowel space. In particular, the choice of the midpoints as corners (or “targets” obtained as an average of frequency values at a vowel’s “steady state”) implicitly implies that vowels in the space are either stationary or that VSA should not include the parts of the acoustic space that can be calculated on the basis of formant frequencies obtained at earlier temporal locations in a vowel, such as those that might correspond to a syllable peak (Jacewicz and Fox, 2008).

The matter of measurement location naturally leads to the question of whether an optimal polygonal VSA can be obtained if greater measurement flexibility is allowed. To explore this possibility, we worked with the current dataset for NC dialect and utilized the within-dialect means of all measurements obtained at 20%-35%-50%-65%-80%-time points. In other words, the means for each of these five time-points were seen as potential “candidate points” for delineation of the VSA perimeter (14 vowels \times 5 time-point means = 60 points) that would maximize the outer boundaries. In these calculations of the VSA, no restrictions were given as to the choice of a uniform location in a vowel such as midpoint, which would limit the number of possible perimeter vowels. Using the liberal convex hull approach, our goal was to maximize the VSA having to our

disposal an unrestricted number of vowels and measurement locations. Again, the perimeter points were selected using a custom MATLAB program. The resulting vowel spaces for NC A0 and NC A3 groups are displayed in Fig. 10 (the maximized spaces for A1 and A2 groups are not shown). As can be seen, this approach refined the outer boundaries and successfully expanded the area of the cumbersome upper back corner of the space for each age group. A between-subject ANOVA of VSA with age group (A0, A1, A2, A3) as the only factor returned no significant effect of age group [$F(3, 41) = 1.85, p = 0.153$], indicating that the maximized spaces did not differ among generations.

Thus, utilizing vowel dynamics and flexible measurement conventions, it is possible to obtain a geometrically reasonable characterization of the actual space utilized by each generation, which implies that a working VSA is common to all. Yet this approach carries a different set of problems. For example, it allows for defining the outer boundaries on the basis of several measurements obtained from a single vowel category (compare the positions of /i/ in A0 and /æ/ in A3 groups in Fig. 10), thus challenging the conceptualization of vowel space as an area bounded by phonologically distinct vowels as its corners (see, for example,

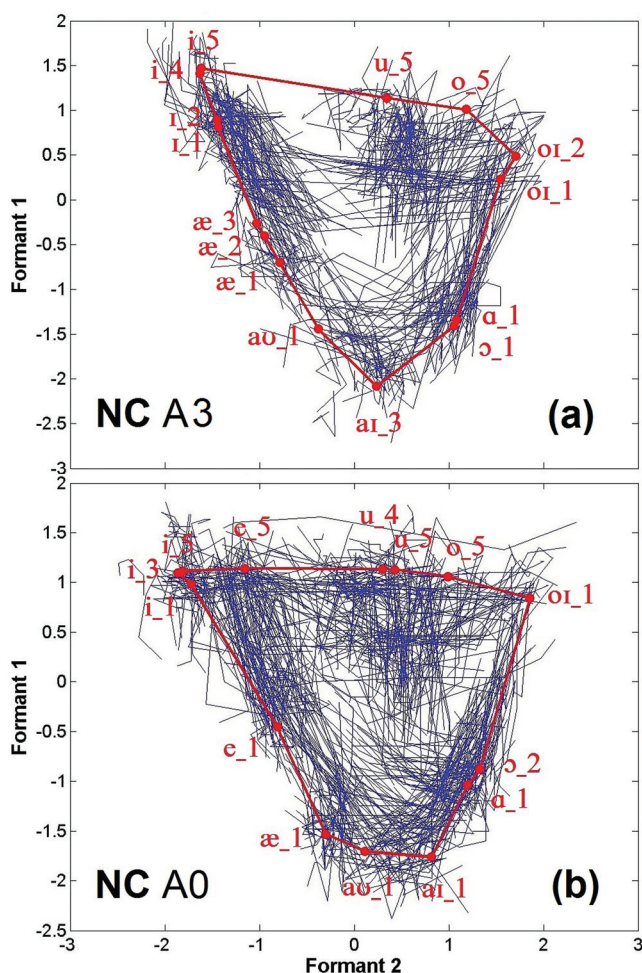


FIG. 10. Average maximized convex hull spaces obtained by utilizing flexible temporal measurement locations in a vowel: 1 = 20%, 2 = 35%, 3 = 50%, 4 = 65%, and 5 = 80%. The displays are for one dialect (NC) and two age groups, old adults (A3) and children (A0).

the convex hull approach in Sandoval *et al.*, 2013). Moreover, it allows for utilizing the same vowel category twice to define perimeters for two levels of height (compare the two positions of /e/ in A0 group in Fig. 10 along the high-low dimension). Also, it can lead to false interpretations related to sound change such as that there is a low back merger of /a/ and /ɔ/ in A3 group, whereas only the onsets of these two vowels are in proximity while both are acoustically very distinct. On the positive side, we can detect traces of the Southern Shift in A3 group in that the onset portions of the vowel /i/ are positioned at the periphery, even if there is no true acoustic /i-/ reversal. Similarly, the peripheral position of /e/ in A0 group along with its absence in A3 group suggest that the Southern Shift is receding in children. While these observations are useful, the maximized polygons cannot tell us more about cross-generational sound change in this dialect because some of the changes corresponding to particular stages of the Southern Shift are interior to the outer boundaries of the utilized space. Whether dialectal differences also become neutralized when their maximized acoustic spaces are compared will be determined in the future.

In our search for a better characterization of the working vowel space in American English, we explored configurations of spectral overlap in an acoustic space rather than the dispersion patterns of distinct vowel categories. Spectral overlap of several vowels in common regions in the acoustic space is natural and expected as most formant trajectories in English are inherently time-varying. Our understanding of the dynamic structure of vowels and dialect-specific use of formant dynamics in the three dialects studied here led us to hypothesize that dialects also differ in the distribution of the regions of spectral overlap created by these spectral variations. The density contours provided ample evidence that both density of spectral overlap and distribution of these overlapping regions indeed vary with dialects and generations. This finding implies that the working space may be, in fact, common to all dialects (as the foregoing analysis of the maximized convex hulls seems to suggest) and it is the internal distribution of spectral density regions that defines dialect-specific “usage” of the acoustic space. Importantly, the dialect-specific distribution of high and low density regions is largely shaped by sound change, which also includes systematic variation in formant dynamics.

We wish to emphasize that it was the phonemically balanced speech material and common speaking style (e.g., citation form vowels in a tightly controlled /hVd/ context produced in isolation by all 135 speakers), that enabled us to detect how the internal distribution of density regions varied across dialects and generations. However, a concern with the use of citation-form material is that the high level of control over speaking style and consonant environment may obscure more natural patterns found in typical conversational speech. Notwithstanding the virtues of spontaneous speech, we do not find this highly controlled material to be particularly problematic for explorations of spectral and temporal patterns. In fact, we have recently addressed this issue in the context of dialectal variation in vowel duration and found that the citation-form vowels approximated the durations of

emphatic vowels in connected speech and, most importantly, preserved dialect- and gender-specific patterns (Jacewicz and Fox, 2015). Those results along with the results of the current study indicate that speakers have knowledge of spectro-temporal relations holding among vowels in their dialect and employ this knowledge in a variety of tasks and social settings. While more extreme formant values in citation form vowels relative to those in spontaneous speech might be expected, it is also the case that relationships among the vowels and their formant dynamics are maintained in spontaneous speech even as the formant values change with vowel duration and speaking style (DiCanio *et al.*, 2015; Fox and Jacewicz, 2012).

It is of course important to note that, in reality, materials from spontaneous speech are not as phonemically and phonetically balanced as those used in the current study. The frequency of occurrence of individual vowels and diphthongs in word initial, medial, and final positions in conversational speech is highly variable (Mines *et al.*, 1978). The actual occurrences of particular vowels in a spontaneous speech sample will likely produce variable density patterns and these patterns will be highly dependent upon social aspects of discourse and the specific speaking situation. What we have shown in this paper is that, all other things being equal, dialect- and sound change-related variations produce considerable differences in the amount and nature of spectral overlap. Before we turn to formal models of formant density patterns, we need to establish how such density patterns in conversational speech differ from those patterns obtained in the balanced and highly controlled speech material in this study. But this is a direction for future research. The current work serves as a starting point toward improving our understanding of how a working acoustic space is utilized by speakers of different dialects and generations.

As a final point, we emphasize that our efforts toward reconceptualizing the vowel space are entirely in the context of dialect variation and sound change. Polygon geometry may meet the needs of speech intelligibility work and may also work well to answer questions pertaining to a within-talker variation. Perhaps the vowel quadrilateral can still be of service to other branches of speech communication, with some caveats pertaining to calling it the “vowel space.”

ACKNOWLEDGMENTS

Research reported in this paper was supported by National Institute on Deafness and Other Communication Disorders under Award No. R01DC006871. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank Joseph Salmons for his contributions to this research.

Al-Tamimi, J.-E., and Ferragne, E. (2005). “Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French,” in *Proceedings of the 9th European Conference on Speech Communication and Technology INTERSPEECH-EUROSPEECH 2005*, Lisbon, Portugal, pp. 2465–2468.

Aylett, M. (1998). “Building a statistical model of the vowel space for phoneticians,” in *Proceedings of the Vth International Conference on*

Spoken Language Processing, edited by R. G. Munnell and J. Robert-Ribes (ASSTA, Sydney, Australia), pp. 85–90.

Aylett, M., and Turk, A. (2006). “Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei,” *J. Acoust. Soc. Am.* **119**, 3048–3058.

Benson, E. J., Fox, M. J., and Balkman, J. (2011). “The bag that Scott bought: The low vowels in northwest Wisconsin,” *Am. Speech* **86**, 271–311.

Bradlow, A. R., Toretta, G., and Pisoni, D. (1996). “Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics,” *Speech Commun.* **20**, 255–272.

Bunton, K., and Leddy, M. (2011). “An evaluation of articulatory working space area in vowel production of adults with Down syndrome,” *Clinical Ling. Phonetics* **25**, 321–334.

Chung, H., Kong, E., Edwards, J., Weismer, G., Fourakis, M., and Hwang, M. (2012). “Cross-linguistic studies of children’s and adults’ vowel spaces,” *J. Acoust. Soc. Am.* **131**, 442–454.

de Berg, M., van Kreveld, M., Overmars, M., and Schwartzkopf, O. (2000). *Computational Geometry*, 2nd ed. (Springer-Verlag, Berlin, Germany), pp. 2–8.

DiCanio, C., Nam, H., Amith, J. D., Castillo Garcia, R., and Whalen, D. (2015). “Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec,” *J. Phonetics* **48**, 45–59.

Durian, D., Dodsworth, R., and Schumacher, J. (2010). “Convergence in blue-collar Columbus, Ohio, African American and White vowel systems?,” in *AAE Speakers and Their Participation in Local Sound Changes: A Comparative Study*, Publication of the American Dialect Society, edited by M. Yaeger-Dror and E. Thomas (Duke University Press, Durham, NC), Vol. 94, pp. 161–190.

Eckert, P. (1997). “Age as a sociolinguistic variable,” in *Handbook of Sociolinguistics*, edited by F. Coulmas (Oxford University Press, Oxford, UK), pp. 151–167.

Ferguson, S. H., and Quené, H. (2014). “Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners,” *J. Acoust. Soc. Am.* **135**, 3570–3584.

Flipsen, P., and Lee, S. (2012). “Reference data for the American English acoustic vowel space,” *Clinical Ling. Phonetics* **26**, 926–933.

Fourakis, M. (1991). “Tempo, stress, and vowel reduction in American English,” *J. Acoust. Soc. Am.* **90**, 1816–1827.

Fox, R. A., and Jacewicz, E. (2008). “Analysis of total vowel space areas in three regional dialects of American English,” in *Proceedings of Acoustics’08* (SFA, Paris, France), pp. 495–500.

Fox, R. A., and Jacewicz, E. (2009). “Cross-dialectal variation in formant dynamics of American English vowels,” *J. Acoust. Soc. Am.* **126**, 2603–2618.

Fox, R. A., and Jacewicz, E. (2012). “Dialectal and generational variations in vowels in spontaneous speech,” in *Proceedings of INTERSPEECH 2011* (ISCA, Florence, Italy), pp. 2921–2924.

Fridland, V. (2012). “Rebel vowels: Southern vowel shift and the N/S speech divide,” *Lang. Ling. Compass* **6**, 183–192.

Gordon, M. (2001). *Small-Town Values and Big-City Vowels: A Study of the Northern Cities Shift in Michigan*, Publication of the American Dialect Society (Duke University Press, Durham, NC), Vol. 84, 229 p.

Higgins, C. M., and Hodge, M. M. (2002). “Vowel area and intelligibility in children with and without dysarthria,” *J. Med. Speech Lang. Pa.* **10**, 271–277.

Jacewicz, E., and Fox, R. A. (2008). “The temporal location of rms peak in coarticulated vowels,” in *Proceedings of Acoustics’08* (SFA, Paris, France), pp. 627–632.

Jacewicz, E., and Fox, R. A. (2015). “Eliciting sociophonetic variation in vowel duration,” in *Proceedings of the 18th International Congress of Phonetic Sciences*, edited by M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, and J. Scobbie (University of Glasgow, Glasgow, UK), paper ICPHS0016, pp. 1–5.

Jacewicz, E., Fox, R. A., and Salmons, J. (2007). “Vowel space areas across dialects and gender,” in *Proceedings of the 16th International Congress of Phonetic Sciences*, edited by J. Trouvain and W. J. Barry (University of Saarland, Saarbrücken, Germany), pp. 1465–1468.

Jacewicz, E., Fox, R. A., and Salmons, J. (2011a). “Cross-generational vowel change in American English,” *Lang. Var. Change* **23**, 45–86.

Jacewicz, E., Fox, R. A., and Salmons, J. (2011b). “Vowel change across three age groups of speakers in three regional varieties of American English,” *J. Phonetics* **39**, 683–693.

Kondaurova, M. V., Bergeson, T. R., and Dilley, L. C. (2012). “Effects of deafness on acoustic characteristics of American English tens/lax vowels in maternal speech to infants,” *J. Acoust. Soc. Am.* **132**, 1039–1049.

- Labov, W. (2001). *Principles of Linguistic Change. II: Social Factors* (Blackwell, Oxford, UK), pp. 415–581.
- Labov, W. (2007). “Transmission and diffusion,” *Language* **83**, 344–387.
- Labov, W., Ash, S., and Boberg, C. (2006). *Atlas of North American English: Phonetics, Phonology, and Sound Change* (Mouton de Gruyter, Berlin, Germany), 318 p.
- Lam, J., and Tjaden, K. (2016). “Clear speech variants: An acoustic study in Parkinson’s Disease,” *J. Speech Lang. Hear. Res.* **59**, 631–646.
- Lee, J., Shaiman, S., and Weismer, G. (2016). “Relationship between tongue positions and formant frequencies in female speakers,” *J. Acoust. Soc. Am.* **139**, 426–440.
- Liu, H.-M., Kuhl, P., and Tsao, F.-M. (2003). “An association between mothers’ speech clarity and infants’ speech discrimination skills,” *Dev. Sci.* **6**(3), F1–F10.
- Lobanov, B. (1971). “Classification of Russian vowels spoken by different speakers,” *J. Acoust. Soc. Am.* **49**, 606–608.
- McGowan, R. W., McGowan, R. S., Denny, M., and Nittrouer, S. (2014). “A longitudinal study of very young children’s vowel production,” *J. Speech Lang. Hear. Res.* **57**, 1–15.
- Milenkovic, P. (2003). “TF32 [software program],” University of Wisconsin, Madison, WI.
- Mines, M., Hanson, B., and Shoup, J. (1978). “Frequency of occurrence of phonemes in conversational English,” *Lang. Speech* **21**, 221–241.
- Morrison, G. S., and Assmann, P. F. (2013). *Vowel Inherent Spectral Change* (Springer-Verlag, Berlin, Germany), 286 pp.
- Nearey, T. M. (2013). “Vowel inherent spectral change in the vowels of North American English,” in *Vowel Inherent Spectral Change*, edited by G. S. Morrison and P. F. Assmann (Springer-Verlag, Berlin, Germany), pp. 49–85.
- Nearey, T. M., and Assmann, P. F. (1986). “Modeling the role of vowel inherent spectral change in vowel identification,” *J. Acoust. Soc. Am.* **80**, 1297–1308.
- Neel, A. T. (2008). “Vowel space characteristics and vowel identification accuracy,” *J. Speech Lang. Hear. Res.* **51**, 574–585.
- Pettinato, M., Tuomainen, O., Granlund, S., and Hazan, V. (2016). “Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication,” *J. Phonetics* **54**, 1–14.
- Pierce, C. A., Block, R. A., and Aguinis, H. (2004). “Cautionary note on reporting eta-squared values from multifactor ANOVA designs,” *Educ. Psychol. Meas.* **64**, 916–924.
- Resnikoff, H. L. (1985). *The Illusion of Reality: Topics in Information Science* (Springer Verlag, New York), 339 pp.
- Sandoval, S., Berisha, V., Utianski, R. L., Liss, J. M., and Spanias, A. (2013). “Automatic assessment of vowel space area,” *J. Acoust. Soc. Am.* **134**, EL477–EL483.
- Stevens, K. N., and House, A. S. (1955). “Development of quantitative description of vowel articulation,” *J. Acoust. Soc. Am.* **27**, 484–493.
- Thomas, E. R. (2001). *An Acoustic Analysis of Vowel Variation in New World English*, Publication of the American Dialect Society (Duke University Press, Durham, NC), Vol. 84, pp. 15–58.
- Thomas, E. R. (2016). “*The Atlas of North American English* and its impacts on approaches to dialect geography,” *J. Socioling.* **20**, 489–497.
- Tjaden, K., Lam, J., and Wilding, G. (2013). “Vowel acoustics in Parkinson’s disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions,” *J. Speech Lang. Hear. Res.* **56**, 1485–1502.
- Vorperian, H. K., and Kent, R. D. (2007). “Vowel acoustic space development in children: A synthesis of acoustic and anatomic data,” *J. Speech Lang. Hear. Res.* **50**, 1510–1545.
- Weismer, G., Jeng, J. Y., Laures, J. S., Kent, R. D., and Kent, J. F. (2001). “Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders,” *Folia Phoniatr. Logop.* **53**, 1–18.
- Yang, J., Fox, R. A., and Jacewicz, E. (2015). “Vowel development in an emergent Mandarin-English bilingual child: A longitudinal study,” *J. Child Lang.* **42**, 1125–1145.