A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English

Arwen Blackwood Ximenes,[a] Jason A. Shaw,[b] and Christopher Carignan
MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith, New South Wales 2751, Australia

(Received 30 June 2016; revised 6 May 2017; accepted 11 May 2017; published online 31 July 2017)

In studies of dialect variation, the articulatory nature of vowels is sometimes inferred from formant values using the following heuristic: $F1$ is inversely correlated with tongue height and $F2$ is inversely correlated with tongue backness. This study compared vowel formants and corresponding lingual articulation in two dialects of English, standard North American English, and Australian English. Five speakers of North American English and four speakers of Australian English were recorded producing multiple repetitions of ten monophthongs embedded in the /sVd/ context. Simultaneous articulatory data were collected using electromagnetic articulography. Results show that there are significant correlations between tongue position and formants in the direction predicted by the heuristic but also that the relations implied by the heuristic break down under specific conditions. Articulatory vowel spaces, based on tongue dorsum position, and acoustic vowel spaces, based on formants, show systematic misalignment due in part to the influence of other articulatory factors, including lip rounding and tongue curvature on formant values. Incorporating these dimensions into dialect comparison yields a richer description and a more robust understanding of how vowel formant patterns are reproduced within and across dialects.

© *2017 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4991346]

[CGC] Pages: 363–377

I. INTRODUCTION

One of the aims of dialect studies is to characterize differences between dialects. The majority of studies analyzing phonetic variation across dialects have based their conclusions on differences in the acoustic properties of the dialects in question. Inferences about speech articulation made on the basis of acoustic analyses are often useful in explaining patterns of variation across dialects and patterns of dialect change over time (Cheshire *et al.*, 2011; Cox, 1999; Harrington *et al.*, 2008). However, cross-dialect studies seldom compare dialects on the basis of both acoustic data and corresponding articulatory data directly (although for work reporting on both ultrasound and acoustic data of the GOOSE vowel in Scottish English see Scobbie *et al.*, 2012; and for a large electromagnetic articulography study investigating tongue position in Dutch dialects see Wieling *et al.*, 2016). Instead, dialect researchers rely heavily on phonetic theory—in particular, how acoustics relate to articulation—to bridge between readily available acoustic descriptions of dialect variation and speech articulation. One common assumption is that the first formant ($F1$) of a vowel is inversely correlated with tongue height; another is that the second formant ($F2$) of a vowel is inversely correlated with tongue backness. This paper assesses these oft assumed correspondences across two dialects of English: North American English (NAmE) and Australian English (AusE), reporting the tongue position of vowels and corresponding formant values for both dialects. Acoustic and articulatory descriptions reveal unique perspectives on how these dialects differ and offer examples of where typically assumed correspondences between formant values and tongue position break down.

The acoustics of NAmE vowels have been extensively reported (e.g., Hillenbrand *et al.*, 1995; Peterson and Barney, 1952), and there are both studies focusing on vowel articulation only (e.g., Johnson *et al.*, 1993) and some that report both acoustic and articulatory data for a subset of vowels (e.g., Noiray *et al.*, 2014). Similar to NAmE, the acoustics of AusE vowels are well-studied, but comparative articulatory data are lacking. Some recent studies on AusE vowel articulation focus on a small subset of AusE vowels. Tabain (2008) investigates the articulatory and acoustic properties of one vowel in different prosodic contexts. Watson *et al.* (1998) compared the acoustic and articulatory vowel spaces of AusE and New Zealand English. Their analysis covered four vowels, those in the words *hid*, *head*, *had*, and *herd*. Lin *et al.* (2012) looked at a larger number of AusE vowels in the /CVl/ context, although they focused on how vowel height influences lateral production (/CVl/) rather than on the phonetic properties of the vowels themselves. The most comprehensive articulatory study of AusE vowels was undertaken over 4 decades ago (Bernard, 1970). Bernard reports on the results of an x-ray study investigating all the AusE vowels but does not report any quantitative measurements of the data. Bernard's qualitative description of x-ray data still constitutes the most comprehensive analysis of Australian vowel articulation to date in that it covers the

[a] Electronic mail: a.blackwoodximenes@westernsydney.edu.au
[b] Present address: Department of Linguistics, Yale University, 370 Temple St., New Haven, CT 06511, USA.

entire vowel space, but due to technical limitations in synchronizing acoustic and articulatory recording, the study does not report corresponding formant values. Thus, to date, dialect differences between Australian and American English are limited to those that can be inferred on the basis of acoustic measurements.

In keeping with this special issue's topic on new ways of investigating dialect variation, we report articulatory data collected using electromagnetic articulography (EMA) and simultaneously recorded acoustic data on AusE and NAmE. EMA captures movements of tongue, lips, and jaw with high spatio-temporal resolution and allows for synchronized audio recording.

The known acoustic differences between NAmE and AusE dialects make for an intriguing test case of how reliably formant values reflect differences in articulation across dialects. To illustrate, consider the vowel referred to as the "GOOSE" vowel in Wells' (1982) lexical sets. The encroachment of GOOSE on front vowels, aka "GOOSE-fronting," has occurred in several dialects of English (Harrington *et al.*, 2008; Watt and Tillotson, 2001; Scobbie *et al.*, 2012; Cox, 1999). Increases in $F2$ may correspond to a more anterior tongue position, decreases in lip rounding (Harrington *et al.*, 2011), changes in tongue curvature or pharyngeal cavity size, or some combination of these articulations. Comparison of dialects that differ in $F2$ values for GOOSE allows us to investigate the articulatory basis of a well-known acoustic difference between dialects. If the higher $F2$ observed in acoustic studies for the GOOSE vowel in AusE (see Cox, 1999 for AusE, cf. Hillenbrand *et al.*, 1995 for NAmE) is due to tongue configuration, we would expect the tongue to be more anterior for GOOSE in AusE speakers compared to NAmE speakers. Another notable difference is the NURSE vowel. Reported formant values across dialects are substantially different for NURSE, which is rhotic in NAmE and non-rhotic in AusE. As with GOOSE, $F2$ for NURSE is higher in AusE than NAmE and, on the basis of $F2$ differences, is said to be more "front" in AusE (Cox, 1999). Thus, both NURSE and GOOSE vowels have a higher $F2$ in AusE than in NAmE, but the articulatory basis of this formant difference, whether common or disparate for these two vowels, is not yet known. In the remainder of this paper, we present articulatory and acoustic analyses of ten monophthongs shared in AusE and NAmE, a discussion of how the dialects differ both in terms of acoustics and articulation, and where in the data assumed correspondences between formant values and tongue positions break down.

II. METHODS

Articulatory and acoustic data were collected as part of a larger EMA study at the MARCS Institute, Western Sydney University.

A. Subjects

Data were analyzed from five NAmE speakers (three females) and four AusE speakers (two females). The former group of speakers range in age at time of recording from 31 to 60 and the latter range in age from 20 to 42. All participants were recruited from the Western Sydney University community and were all residents of the Greater Sydney region at the time of recording. Three of the North American speakers had lived in Australia for less than 2 years. The other two speakers had lived in Australia for 8 and 14 years, respectively, at the time of recording. The North American speakers originated from diverse regions of North America as follows: F04 (California), F10 (Chicago), F11 (New England), M01 (Nova Scotia), and M02 (Washington State), and the AusE speakers all originated from New South Wales.

B. Materials

Stimuli comprised a list of lexical items and nonce words containing 15 vowels, including 10 monophthongs, in the sVd context. This paper focuses on analysis of the monophthongs. The stimulus items are provided in Table I. Alongside the orthographic stimuli (column 1), we provide the IPA symbol corresponding to the vowel in North American and AusE and the reference word, or "lexical set," for the vowel devised by Wells (1982). The reference words disambiguate the spelling, which is particularly useful for nonce words and were used as a guide for participants to produce nonce stimuli with the intended target vowel. This set of monophthongs covers the whole of the NAmE and AusE acoustic vowel spaces. The only monophthong missing is START from AusE, which according to Cox (2006), does not differ in its formant structure from the AusE STRUT vowel. As indicated by the NAmE IPA symbols in Table I, a merger between THOUGHT and LOT was expected for some NAmE speakers given the diverse regions of origin.

C. Procedure

The movements of speech articulators were tracked using a Northern Digital Inc. Wave EMA system (Northern Digital Inc., Ontario, Canada) at a sampling rate of 100 Hz. This system uses an electromagnetic field to track the movement of small receiver coils or sensors (∼3 mm in size) glued or taped to the articulators. The electromagnetic field induces an alternating current in the sensors, and the strength of this current is used to determine the position of the sensors in relation to the transmitter. Articulatory movements are captured in the

TABLE I. List of materials for North American English (NAmE) and Australian English (AusE).

sVd stimuli	Australian English vowels (IPA symbols)	North American English vowels (IPA symbols)	Lexical set
sad	æ	æ	TRAP
said	e	ɛ	DRESS
sawed	oː	ɔ (ɑ)	THOUGHT
seed	iː	i	FLEECE
sid	ɪ	ɪ	KIT
sod	ɔ	ɑ	LOT
sood	ʊ	ʊ	FOOT
sud	ɐ	ʌ	STRUT
sued	ʉː	u	GOOSE
surd	ɜː	ɝ	NURSE

vertical, horizontal, and lateral dimensions with high spatial resolution (<0.5 mm root-mean-square error; Berry, 2011). In this study, we focused on movements in the horizontal and vertical dimension, since these are the dimensions typically assumed to correspond to values of the first two formants. The sensor trajectories were synchronized to the audio signal during recording by the NDI system. EMA sensors were glued to the following articulators along the midsagittal plane: jaw, below the lower left incisor; lips, at the vermillion edge of the upper lip (UL) and lower lip (LL); tongue tip (TT), tongue blade (TB) and tongue dorsum (TD). The TD sensor was placed as far back as comfortable for the participant. The TT sensor was placed approximately 5 mm back from the TT and the TB sensor was placed midway between the TT and TD sensors. The three lingual sensors and the UL sensor (with tape) can be seen in Fig. 1, with connecting wires. The wires attached to the LL and jaw sensors can also be seen below the tongue. Speech acoustics were recorded using a shotgun microphone at a sampling rate of 22 050 Hz.

The target stimulus words were displayed on a computer monitor placed outside of the magnetic field (a volume of 300 mm^3). One word was presented per trial. There were 15 trials (one per vowel) per block and eight blocks in the experiment. The eight blocks were divided into two sets of four blocks. Another experimental task intervened between the presentation of the first four vowel blocks and the last four vowel blocks. Within a block, the order of presentation of items was fixed and was the same for all blocks. Altogether, there were 15 (vowels) × 8 (repetitions) = 120 vowel tokens per participant. Of the recorded data, the monophthongs consist of 10 (vowels) × 8 (repetitions) = 80 tokens per participant, 320 monophthong tokens in total for AusE (four speakers). For the NAmE speakers, half of one male speaker's session was not recorded successfully (due to problems with sensor adhesion). Accordingly, only 360 tokens were recorded in total for NAmE. There was an error in the audio for the first 20 tokens of one female AusE speaker, so only 300 tokens were recorded for AusE. Other technical problems due to data acquisition, analysis, and mispronunciation resulted in seven more tokens out of 660 total across dialects (∼4% of the data) being excluded from the analysis: four tokens of NAmE and three tokens of AusE.

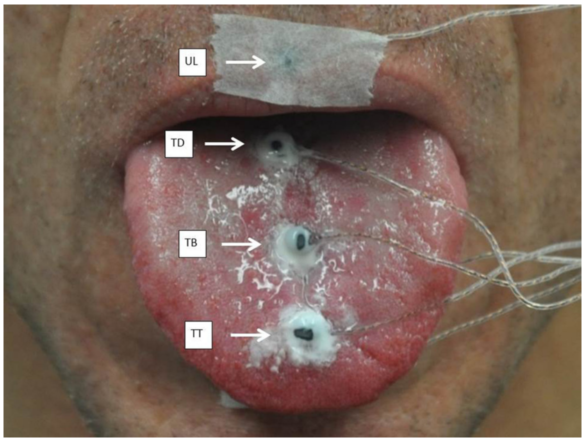

FIG. 1. (Color online) Image of protruded tongue with labeled sensors. UL = upper lip, TT = tongue tip, TB = tongue body, and TD = tongue dorsum. The wires for LL and jaw sensors are also visible.

Head movements were corrected using custom-written MATLAB functions developed by Mark Tiede and revised by Donald Derrick. Sensors taped to the nasion and left/right mastoid processes were used as stable reference points for the head correction procedure. The articulatory data were rotated relative to the occlusal plane so that the origin of the coordinate system corresponds to a point immediately posterior to the incisors. The occlusal plane was established by having the participant bite down on a protractor with three sensors affixed in a triangular formation. For two of the NAmE speakers there were technical complications with head correction and data rotation that motivated another stage of normalization described in Sec. II F.[1]

D. Articulatory measurements

Figure 2 shows a token of the word *seed*, which is used here to illustrate the measurement procedure. The topmost pane shows the tangential velocity of the TD sensor, based on movements in vertical and horizontal dimensions, the middle pane shows the TD trajectory in the vertical dimension, and the lower pane shows the speech waveform. The three panes are synchronized in time. Vertical dashed lines indicate the velocity peaks associated with movement toward and movement away from the vowel target. The solid vertical line indicates the velocity minimum which occurs for this vowel at the highest point reached by the TD. We determined the vowel target based on this velocity minimum. Measurements were extracted from sensor trajectories based on timestamps labeled using *findgest*, an algorithm developed for the MATLAB-based software package, "Multichannel visualization application for displaying dynamic sensor movement" (MVIEW), by Mark Tiede at Haskin

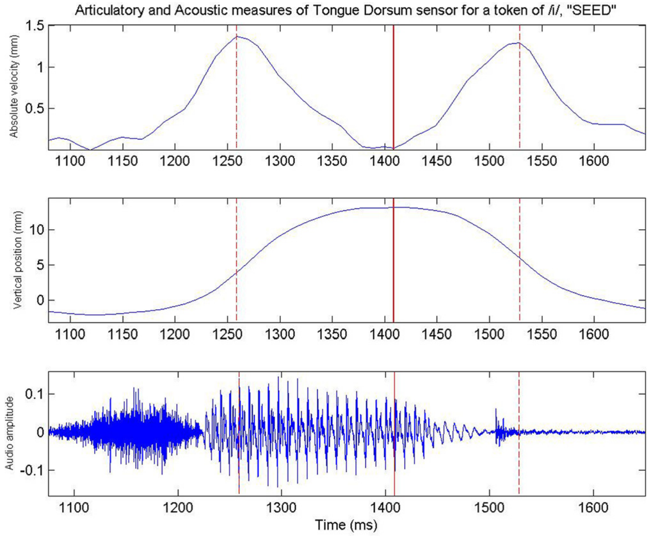

FIG. 2. (Color online) Labeling procedure for a "seed" (FLEECE) token. The lower pane contains the speech waveform. The middle pane represents the trajectory of the TD sensor in the vertical dimension (the occlusal plane was set to 0 mm). The upper pane represents the velocity of the TD sensor. Vowel target in all three panes is indicated by a solid vertical line. Velocity peaks in movements toward and away from target are indicated by dashed lines.

Laboratories. This program was used to detect the nearest tangential velocity minimum of the TD sensor during the interval corresponding to the vowel. We then extracted positional coordinates from all the lingual sensors and from the LL and UL sensors at this vowel target landmark.

For some tokens, the point of minimum velocity in the TD trajectory did not give a reliable indication of the vowel target. This was the case, in particular, for vowels with a long period of little or no movement, i.e., a quasi-steady state. In these tokens, since velocity remains relatively constant, selecting the vowel target based on an absolute velocity minimum is somewhat arbitrary. When the time point of minimum velocity in the TB sensor trajectory provided a clearer indication of the vowel target than the TD sensor, we extracted articulatory coordinates from the minimum velocity of the TB sensor instead of the TD sensor.

E. Acoustic measurements

Formant listings ($F1$ and $F2$) were extracted using LPC analysis in PRAAT (Burg method with a 25 ms window length and a 6 dB per octave pre-emphasis from 50 Hz) at the point determined by the minimum velocity of the TD (see, e.g., Shaw *et al.*, 2013: pp. 166–167). Results were then inspected visually, and outliers were hand corrected as needed. Using the time points extracted from the articulatory measures for the acoustic analysis enables a direct comparison between articulation and acoustics. Parsing vowel targets using the point of minimum velocity in the articulatory data follows similar general principles used to identify formants in Cox (2006) and Harrington *et al.* (1997), where vowel targets were identified based on formant displacement patterns, e.g., max/min $F1$/$F2$, depending on vowel. Max/min formant values relate closely to the minimum velocity of articulator movement in our data. Other acoustic studies have used the acoustic midpoint of the vowel, which did not correspond as consistently to the velocity minimum of the TD or TB sensors in this data, as can be seen, for example, in Fig. 2, where the velocity minimum occurs well after the midpoint of periodic energy in the acoustic signal.

F. Analysis

One of the challenges of analyzing speech production across speakers is that anatomical differences influence both the formant values and EMA positional coordinates. In the case of formants, differences in vocal tract length influence the average formant values. In articulatory data, differences in tongue shape, volume, and sensor placement lead to different average values across speakers. For example, a retraction of the TD to a point 30 mm behind the front teeth would have a different meaning between speakers due to variation in tongue size. In both cases, because of differences in anatomy, between-speaker differences for the same vowel can be larger than within-speaker differences across vowels. In order to facilitate comparison across our speakers, we normalized both the formant values and the lip and tongue positional coordinates by calculating z-scores of sensor positions and formant values, a method established by Lobanov (1971) for vowel formants and extended to EMA sensor positions (e.g., Shaw *et al.*, 2016). Sensor positions were normalized (1) across the three lingual sensors, TD, TB, TT and (2) across the two labial sensors, UL and LL. The horizontal and vertical dimensions were normalized separately. To provide a measure of lip rounding, we calculated the mean horizontal position of the UL and LL sensors. Normalization preserves the within-speaker structure of the data, but allows for a direct comparison across speakers, serving the goal of dialect comparison.

Due to the issue with the data rotation for two male NAmE speakers mentioned above, we applied another step of data normalization to the articulatory data. For the lingual sensors, normalized values for each sensor were projected onto a common millimeter space. This was done by multiplying the z-scores by the mean standard deviation (SD) across all sensors and then the overall mean was added. This allows us to present values in millimeters that retain the structure of the data. The same process was followed for the labial sensors. Thus, the millimeter values discussed in the context of individual differences are values that have been normalized in a manner comparable to our treatment of formants.

III. RESULTS

We report the acoustic results first followed by the articulatory results, including both TD position and lip rounding, for both NAmE and AusE. Following the acoustic and articulatory overviews for each dialect we report correlations between acoustic and articulatory measurements of each vowel and dialect differences found in each type of data.

A. NAmE acoustics and articulation

1. Acoustic data overview

The distribution of normalized $F1$ and $F2$ values across the acoustic vowel space for NAmE is presented in Fig. 3(a). Ellipses represent 95% confidence intervals for each vowel, and are centered on the mean of each vowel category. Normalized $F2$ values are shown on the x axis, and normalized $F1$ values are shown on the y axis.

In the following discussion of the acoustic data, we refer to three groups of vowels—“front,” “central” and “back”—based upon how the vowels are differentiated by relative $F2$ values. In grouping vowels based on $F2$, we consider the covariation of $F2$ and $F1$. Since $F2$ decreases with increases in $F1$, our groupings of front, central, and back follow diagonals from the top left to the bottom right of the formant space. There are four vowels with comparatively high $F2$ that are differentiated by $F1$. In order of low to high $F1$, these vowels are: FLEECE, KIT, DRESS, and TRAP. We refer to these as front vowels. The front vowel with the lowest average $F1$, FLEECE, has an $F2$ that is nearly two SDs above the mean $F2$ while the front vowel with the highest average $F1$, TRAP, is near the mean value of $F2$ in the data. There are five vowels that have relatively low $F2$ values. We refer to these as back vowels and list them here in order from low to high $F1$: GOOSE, NURSE, FOOT, LOT, and THOUGHT. NURSE and FOOT are heavily overlapped in

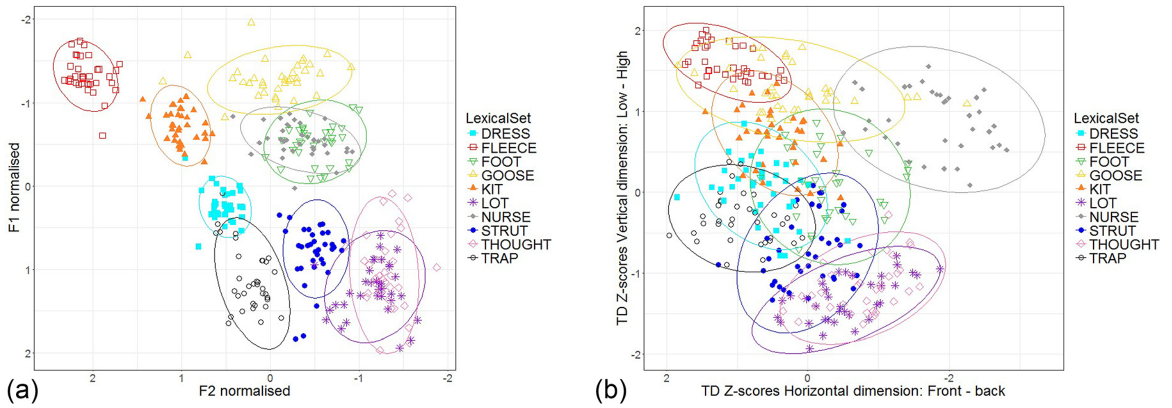

FIG. 3. (Color online) Normalized formants (a) and TD sensor positional coordinates (b) for NAmE vowels.

$F1$ and $F2$, but they are differentiated in $F3$. NURSE, which is rhotic for these speakers, has a lower mean $F3$ (z-score) than FOOT: mean $F3$ for NURSE $= -2.181$ (SD $= 0.718$), cf. mean $F3$ for FOOT $= -0.117$ (SD $= 0.371$), a difference which is significant based on a linear mixed effects model ($\beta_{\text{vowel}} = -2.12$, SE $= 0.39$, $t(4) = -5.41$, $p = 0.005$), where SE is the Standard Error. The remaining vowel, STRUT, has an intermediate $F2$, which is lower than the front vowels, TRAP and DRESS, and higher than the back vowels, THOUGHT and LOT, of comparable $F1$. We refer to this vowel as central. We now turn to the articulatory data to observe how the differences in formant values correspond to tongue position in NAmE.

2. Articulatory data overview

a. TD position. In order to assess whether the vowels we have termed front, central, and back on the basis of formant measurements indeed correspond to front, central, and back lingual articulatory positions, we first present data on the position of the TD sensor. The mapping from articulation to acoustics is of course impacted by differences in vocal tract area function across the entire length of the vocal tract. Focusing on a single fleshpoint necessarily has limitations but has frequently been used as a heuristic for tongue position in vowels (e.g., Noiray *et al.*, 2014; Georgeton *et al.*, 2014), and allows us to maintain comparability to past research. Besides the TD sensor we also explored the TB sensor and the point of inflection of a polynomial curve fit to the three lingual sensors. Of these measures, we found the TD position to be the measure that best differentiated vowels within and across speakers.

Figure 3(b) shows the normalized values (z-scores) of the TD sensor for all five subjects. The y axis shows the vertical position, and the x axis shows the horizontal position from front (positive z-scores on the left side of the figure) to back (negative z-scores on the right side of the figure). As with the formant data, ellipses contain 95% confidence intervals for each vowel distribution and are centered on the mean. The distribution of the TD sensor follows the range of motion with which that fleshpoint on the tongue varies across vowels. Although there are some notable exceptions, by and large, vowels that are differentiated by $F1$ in the acoustics are differentiated by TD height. This is particularly clear for the front vowels. FLEECE, KIT, DRESS, and TRAP are all differentiated by tongue height, and the TD height differences are inversely related to $F1$. While we noted covariation between $F1$ and $F2$ in acoustic space, we do not see corresponding covariation between the horizontal and vertical position of TD. For example, within the front vowels, FLEECE and TRAP are slightly more fronted than KIT and DRESS, cf. the diagonal patterning of these vowels in the acoustic vowel space. On the basis of the formant data, we described NAmE as having one central vowel, STRUT. The TD data indicate that, in addition to STRUT, GOOSE and FOOT also have an intermediate degree of backness. The TD data indicate that GOOSE is more back than FLEECE, FOOT is more back than DRESS, and STRUT is more back than TRAP. The remaining vowels—NURSE, LOT, and THOUGHT—are even more back than GOOSE, FOOT, and STRUT. Of particular note is the fact that NURSE is produced with a considerably more retracted tongue position than FOOT, despite a similar $F2$ value.

b. Lip rounding. Lip rounding involves protrusion of both the UP and the LL. Our metric of lip rounding is the average horizontal position of the UL and LL sensors. Figure 4 shows boxplots indicating the mean position in the x-dimension (horizontal) of the UL and LL sensors across vowels, normalized across speakers. These plots show that in our NAmE data, the most rounded vowel is GOOSE, followed by FOOT and NURSE. The notches in the boxplots indicate 95% confidence intervals around the median values. The confidence intervals for GOOSE, FOOT, and NURSE do not overlap the other vowels, indicating statistically significant differences (at $\alpha = 0.05$), i.e., these three vowels are significantly more rounded than the other vowels. All else equal, an elongated vocal tract resulting from lip rounding is expected to lower formant values, particularly $F2$ (Stevens, 1989). As previously described, GOOSE, FOOT, and NURSE are in the back vowel space based on acoustic measures, which can be a consequence of different degrees of rounding and tongue backness.

In summary, most of the NAmE vowels in this study can be clearly differentiated on the basis of $F1$ and $F2$. Exceptions to this are LOT and THOUGHT, which are overlapping, as well as NURSE and FOOT, which are distinguished by $F3$. The relative tongue positions for the vowels

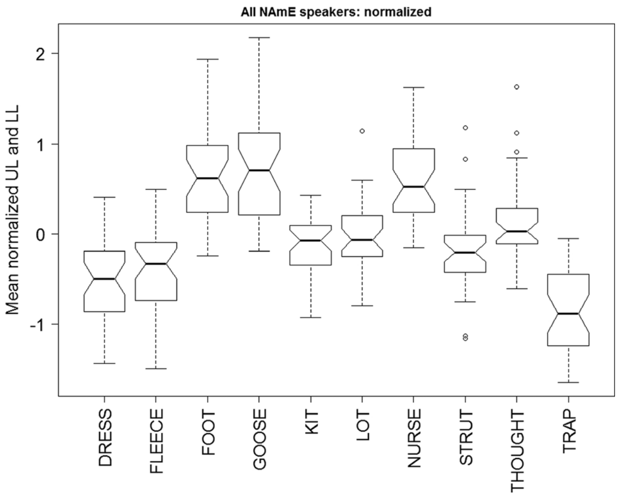

FIG. 4. Box plots of the mean of the UL and LL position in the longitudinal dimension (used as an index of lip rounding): NAmE. Notches indicate 95% Confidence Intervals around median values.

in NAmE show similar but not identical patterns. FLEECE, KIT, DRESS, and TRAP all have front TD positions. NURSE is the farthest back, probably because of its rhotic quality in NAmE. Of the non-rhotic vowels, GOOSE, FOOT, LOT, and THOUGHT are produced with a more retracted tongue position than the other vowels; however, note that GOOSE has variable backness measurements. GOOSE, NURSE, and FOOT were the most rounded of the vowels. Rounding for GOOSE and FOOT may contribute to an explanation of why these vowels show greater separation from front vowels FLEECE and KIT in $F2$ than they do in TD backness. LOT and THOUGHT are realized with the same TD position, while both are further back than STRUT, which is central. Therefore, the acoustic vowel space for NAmE could be considered to display a 4:1:4/5 configuration, whereby there are four front vowels differing in height, one central vowel, and four or five back vowels, depending on whether LOT and THOUGHT are merged. However, based on tongue position alone, it appears the following might be a better description: 4:3:2/3 with FLEECE, KIT, DRESS, and TRAP being front, NURSE, LOT, and THOUGHT being back, and GOOSE, FOOT, and STRUT being central. Thus the descriptions of the vowel space in terms of acoustics ($F1$ and $F2$) vs articulation (quantified as TD height and backness) lead to slightly different conclusions for the central and back vowels, which are not as clearly differentiated by TD position as are the front vowels.

B. AusE acoustics and articulation

1. Acoustic data overview

The distribution of normalized formant values ($F1$ and $F2$) across the acoustic vowel space is presented in Fig. 5(a). The ellipses show 95% confidence intervals for each vowel, and are centered on the mean of each vowel category [as for the NAmE data in Fig. 3(a)]. $F1$ and $F2$ are plotted on the y axis and x axis, respectively. In line with previous acoustic studies of AusE (e.g., Cox, 2006), the vowels are fairly evenly distributed across the vowel space and can be classified as front, central, and back on the basis of $F2$. There are four vowels with high $F2$, i.e., front vowels that differ in $F1$: FLEECE, KIT, DRESS, and TRAP. There are three central vowels that have intermediate $F2$ values, GOOSE, NURSE, and STRUT, and also differ with respect to $F1$. The remaining back vowels have low $F2$: FOOT, THOUGHT, and LOT. We again turn to the articulatory data to observe how the differences in formant values correspond to tongue position in AusE.

2. Articulatory data overview

a. TD position. Figure 5(b) shows the normalized values (z-scores) of the TD sensor for all four AusE subjects. The structure of the figure follows Fig. 3(b). The y axis shows the vertical position, and the x axis shows horizontal position from front (positive z-scores on the left side of the figure) to back (negative z-scores on the right side of the figure). Ellipses represent 95% confidence intervals and are centered on the mean position of each vowel. The distribution of vowels in the articulatory data generally follows the distribution of vowels in formant space, perhaps even more so than NAmE. More specifically, $F1$ tends to be inversely correlated with tongue height, and $F2$ tends to be inversely correlated with tongue backness. Of the front, central, and back vowels determined on the basis of the formants, the

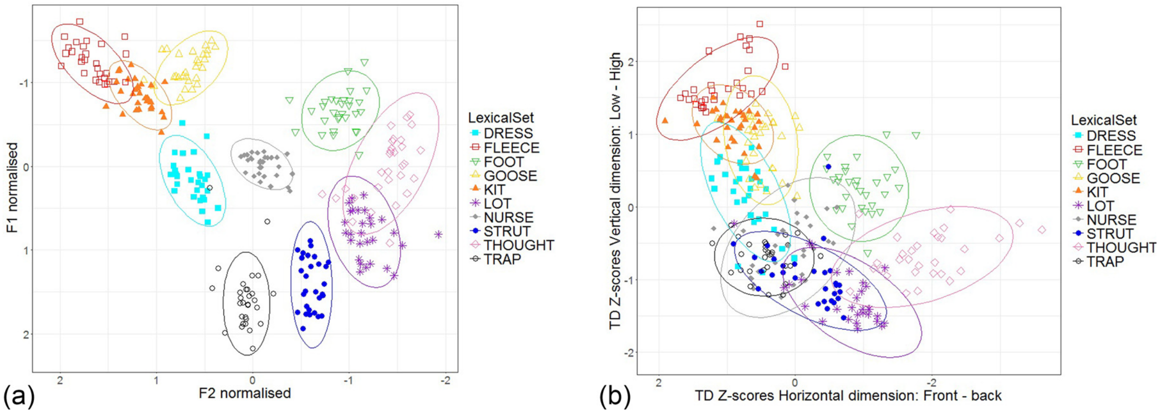

FIG. 5. (Color online) Normalized formants (a) and TD sensor positional coordinates (b) for AusE vowels.

back vowels show the least overlap at the TD sensor. The back vowels—FOOT, THOUGHT, and LOT—are realized with a TD position that is more posterior than the other vowels. THOUGHT is the most back and FOOT and LOT are both realized with a similar longitudinal position, with FOOT higher than LOT, as expected from the formant values. The center of the ellipses for STRUT and NURSE are closest to zero on the x axis, indicating that they are at the average level of backness in the data. We characterized these vowels, in addition to GOOSE, as central vowels by virtue of having intermediate $F2$ values. Of these three central vowels, GOOSE has the most front TD position. The front vowels FLEECE, KIT, DRESS, and TRAP have positions that are indeed more anterior than the other vowels.

3. Lip rounding

Figure 6 shows boxplots of the average horizontal position of the UL and LL sensors across vowels (cf. NAmE in Fig. 4). These data show that the most rounded vowels are FOOT, GOOSE, and THOUGHT, possibly also NURSE, which is slightly different from our NAmE data. GOOSE is more rounded than the other central vowels NURSE and STRUT. Speaker M03 is the only speaker who deviates from this pattern. For him, THOUGHT is less rounded than for the other speakers such that THOUGHT shows a similar degree of rounding as LOT. We note that it is this speaker's tokens of THOUGHT that contributed to the overlap between THOUGHT and LOT ellipses in the vowel space in Fig. 5(a).

In summary, the relative tongue positions for the vowels in AusE are generally as expected from the formant values, given the common heuristics deployed in dialect comparison. By and large, the AusE vowels in this study can be differentiated on the basis of $F1$ and $F2$, and a similar partitioning of the vowel space can be observed in the position of the TD sensor in vertical and longitudinal dimensions, although with greater overlap. The AusE vowel space can be considered to display a 4:3:3 configuration, whereby there are four front vowels differing in height, three central vowels differing in height, and "three" back vowels also differing in height. In Sec. III C, we examine the correlations between acoustic and articulatory data.

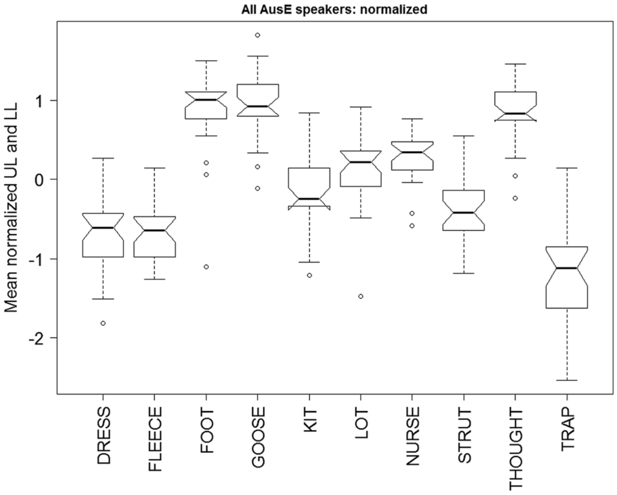

FIG. 6. Box plots of the mean of the UL and LL data in the longitudinal dimension (used as an index of lip rounding): AusE. Notches indicate 95% Confidence Intervals around median values.

C. Acoustic-articulatory relations

In Secs. III A and III B, we provided a general overview of the acoustics and articulation of both NAmE and AusE based on nine speakers in total. In this section we examine the acoustic-articulatory relation more directly. We evaluate linear relations between formant values and TD position across dialects and within dialects and also uncover cases in which the linear relation breaks down.

Pearson product-moment correlations were computed to quantify the relationship between formants and tongue position. Across dialects there was a strong negative correlation between $F1$ and TD height, $r = -0.78$, $p < 0.001$, and a positive correlation between $F2$ and TD backness, $r = 0.69$, $p < 0.001$. These correlations are in the expected directions, since, in our data, the vertical coordinate increases with TD height while the longitudinal coordinate decreases with TD backness. Correlations within dialect produce similar results, slightly stronger negative $F1$/TD height correlations and moderate to strong positive correlations for $F2$/TD backness: NAmE, $F1$/TD height, $r = -0.817$, $p < 0.001$ and for $F2$/TD backness, $r = 0.563$, $p < 0.001$; AusE, $F1$/TD height, $r = -0.741$, $p < 0.001$ and $F2$/TD backness, $r = 0.811$, $p < 0.001$.

Although there are reasonably strong correlations both within and across dialects, we also noticed that linear correlations were stronger for some speakers than for others. For one AusE male speaker [M03, see Figs. 7(a) and 7(b)], we observed an acoustic-articulatory mismatch in backness for vowels LOT and THOUGHT. The TD is further back for THOUGHT than for LOT [Fig. 7(b)] but THOUGHT has a higher $F2$ than LOT [Fig. 7(a)]. In this case, $F2$ provides a poor diagnostic for tongue backness. The lip data revealed that this speaker produced a smaller difference in rounding for this vowel pair, which offers a likely explanation for why this speaker showed a similar lingual articulatory pattern but a different $F2$ pattern from the other AusE speakers. It is possible that for these vowels lip rounding has more of an impact on $F2$ than tongue backness.

Another mismatch in the data is less easy to explain. For one of the AusE male speakers, M05, we observed an inconsistency in the acoustic-articulatory relation in the central part of the vowel space [Figs. 7(c) and 7(d)]. Although this speaker shows the same degree of acoustic-articulatory correspondence as other speakers in the front and back section of the vowel space, the central vowels NURSE, GOOSE, and STRUT all have a similar level of backness, i.e., as determined by the horizontal position of TD, while displaying large differences in $F2$.

Unlike the case of M03's THOUGHT and LOT vowels described above, it is unlikely that the difference in M05's $F2$ across GOOSE, NURSE, and STRUT is due to a degree difference in rounding. This speaker follows the AusE group trend for lip rounding; GOOSE is the most rounded vowel

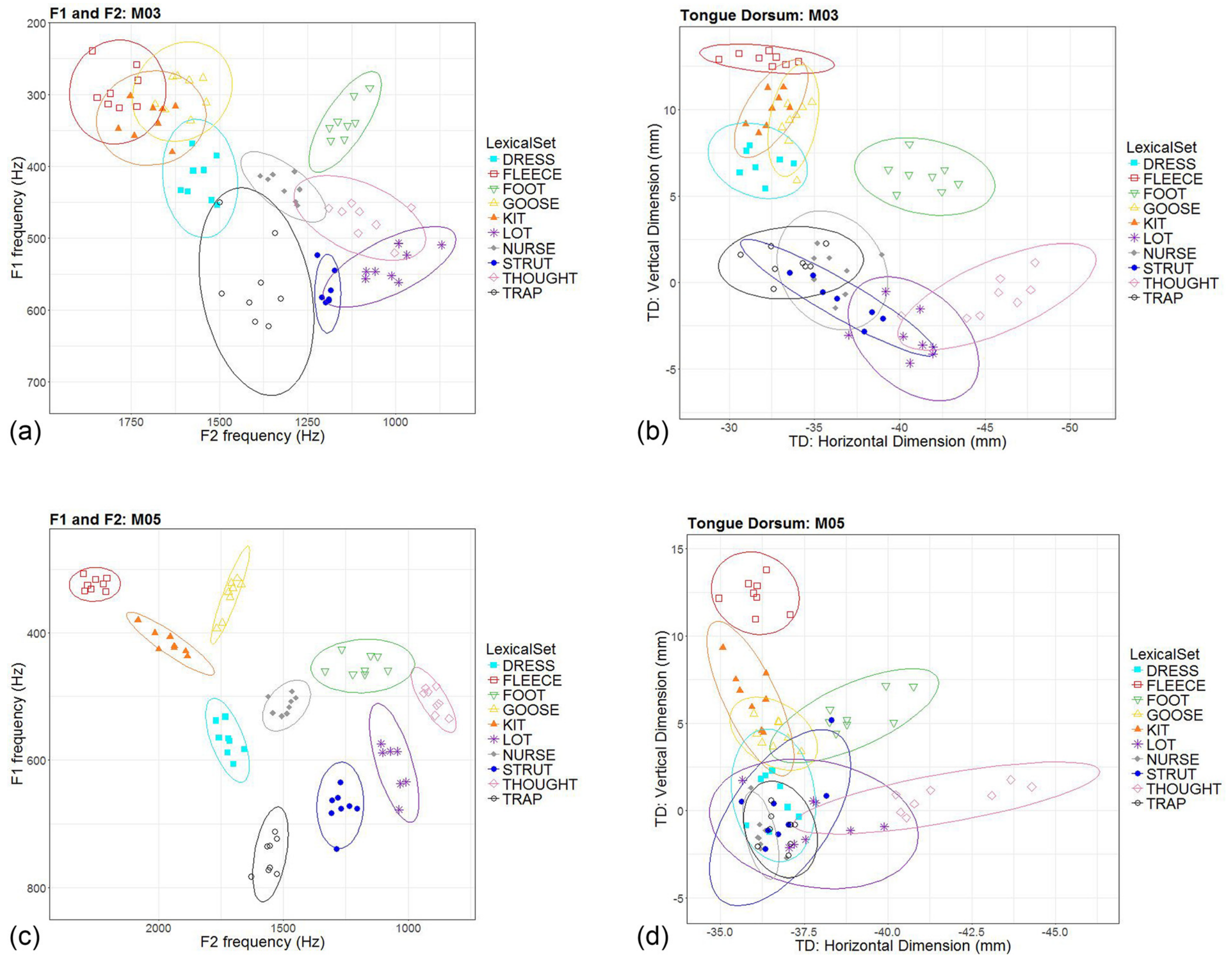

FIG. 7. (Color online) (a) and (c) on the left display individual speaker $F1$ and $F2$ values. (b) and (d) on the right display individual speaker TD sensor positional coordinates for the speaker on the same row. The vowels are differentiated by the same symbols and colors in both plots.

followed by NURSE. However, for this speaker, it is GOOSE that shows unexpectedly high $F2$ values given the TD position. Rounding would be expected to lower $F2$, the opposite pattern of what we observed. The relationship between $F2$ and tongue backness for M05 can be seen in Fig. 8(a). For reference, we have plotted the same relation for another AusE speaker in Fig. 8(b). Vowels are differentiated by color and symbol, with a regression line fit to the data points. For speaker M05, tokens of GOOSE appear above the regression line (for $F2$-TD backness) while tokens of NURSE fall below it. For F07, the relationship between $F2$ and TD backness is more linear, and is in line with expectations based on the acoustic-articulatory relations data presented above. However for M05 there appears to be a non-linear relationship for these aspects of acoustics and articulation. Thus, although we find strong correlations between formant values and TD position, there are also corners of the data in which such correspondences break down.

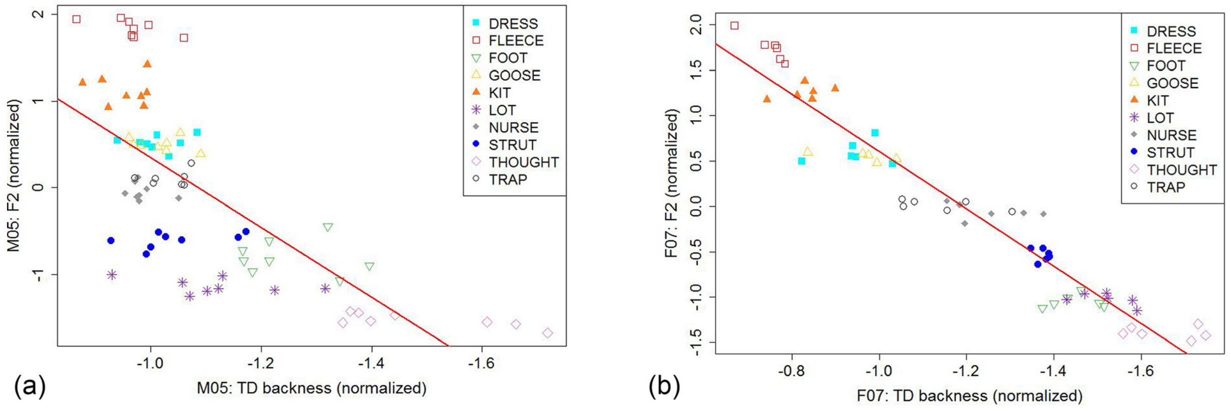

FIG. 8. (Color online) (a) and (b) display normalized $F2$ plotted against TD backness for AusE speakers F07 and M05, respectively.

D. Dialect comparison

To quantify the difference between the two dialects we fit linear mixed-effects models to $F1$, TD height, $F2$, and TD backness. The fixed factors in the models were lexical set (i.e., vowel) and dialect as well as the interaction between these factors. Random variables were speaker and block presentation order, whether the vowels were produced in one of the blocks early in the experiment or in one of the blocks later in the experiment. Tables summarizing the models can be found in the Appendix. The residuals of all models were checked for normality and heteroscedasticity. Dialect was not a significant predictor but the interactions between lexical set and dialect were significant for all dependent measures. To visualize these results, we plotted model predictions for the interaction term. Figure 9 shows the estimated marginal means, or predicted means, with 95% confidence intervals for each dialect by lexical set (Fig. 9). Figure 9(a) shows the results for $F1$ (where $F1$ increases from left to right) and Fig. 9(b) shows $F2$ (where $F2$ increases from left to right). Figure 9(c) shows the results for TD height (where tongue height increases from left to right) and TD backness is represented in Fig. 9(d) (where tongue backness decreases from left to right, i.e., the tongue is in its most anterior position at the far right).

All vowels except for DRESS differ significantly across dialects in either $F1$ or $F2$ (or both $F1$ and $F2$). TD position differentiates fewer vowels. In general, a difference in TD position across dialects implies a difference in formants—there is just one exception. However, a significant difference in formants does not imply a significant difference in TD position. Five vowels differ across dialects in $F1$ (TRAP, THOUGHT, STRUT, NURSE, LOT), but only one of these differs in TD height (NURSE). $F1$ was higher for NAmE THOUGHT and LOT than for AusE THOUGHT and LOT. These differences in $F1$ do not correspond to significant differences in TD height. Particularly for LOT, the dialects are very similar in TD height despite the significant $F1$ differences. The only vowel with a significant difference across dialects in TD height was NURSE. This difference has the expected corresponding difference in $F1$, i.e., NAmE differs from AusE in having both lower $F1$ and higher TD position. Six vowels differ across dialects in $F2$ (TRAP, NURSE, KIT, GOOSE, FOOT, FLEECE). Three of these show corresponding significant differences in TD backness (NURSE, KIT, FOOT). Finally, there was one significant difference in

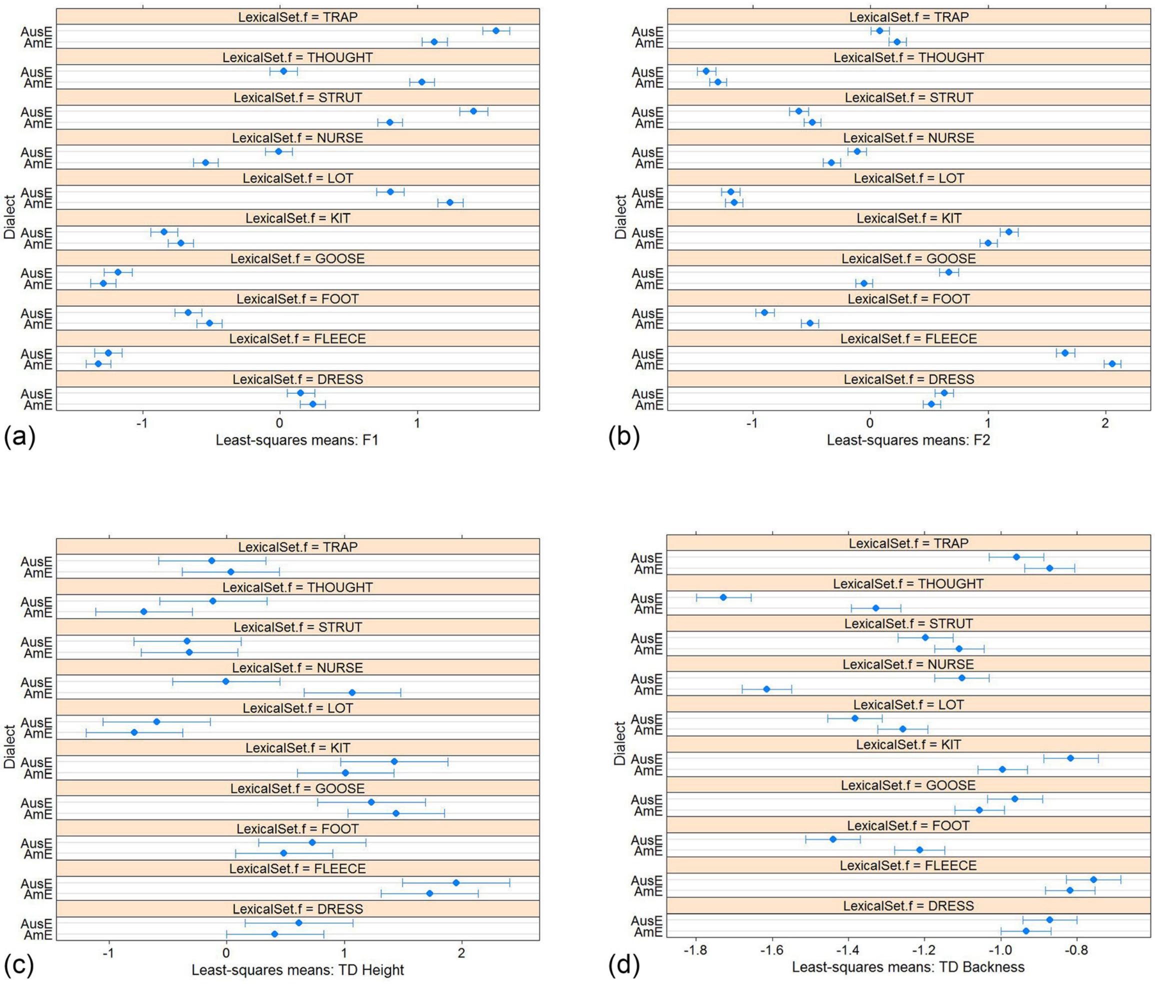

FIG. 9. (Color online) (a) and (b) display least square means (estimated marginal means) for dialect by vowel with 95% Confidence Intervals for $F1$ and $F2$, respectively. (c) and (d) display least square means as above for TDz and TDx, respectively (the x axis from left to right represents TDz positions increasing in height; the x axis from left to right represents TDx positions decreasing in backness). All measurements normalized across both dialects.

TD backness that did not have a corresponding significant difference in $F2$—this was for the vowel THOUGHT, which has a considerably more anterior TD position in NAmE than in AusE.

IV. DISCUSSION

A. Correspondences between acoustics and articulation

Parallel acoustic and articulatory data on monophthongs from two English dialects, NAmE and AusE, have allowed us to evaluate the correspondence between articulation and acoustics that is frequently used to reason about how variation in formants across dialects relates to articulation. Variation in $F1$ is assumed to correlate inversely with tongue height; while variation in $F2$ is assumed to correlate with tongue backness.

These heuristics relating $F1$ and $F2$ to TD position have theoretical bases in tube models of the vocal tract which predict this correlation only within certain ranges of articulatory variation (e.g., Chiba and Kajiyama, 1941; Fant, 1960; Stevens, 1989). For example, Stevens (1989) suggests a two tube model for low vowels, e.g., STRUT, LOT, THOUGHT, and a three-tube model with a Helmholtz resonator for vowels with a narrow constriction, e.g., GOOSE. Given the boundary conditions for the Helmholtz resonance defined for the three-tube model, $F1$ is predicted to increase as the cross-sectional area of the constriction widens. This should give rise to a linear (or quasi-linear) correlation between $F1$ and TD height. Moreover, when the area of the constriction widens beyond the range of values that can support Helmholtz resonance, a further increase in $F1$ is expected in the transition from a three-tube to a two-tube model. This would also contribute to the correlation between $F1$ and TD height. These two mechanisms, increasing the cross-sectional area of the constriction supporting Helmholtz resonance and transitions from vocal tract shapes that support Helmholtz resonance to those that do not, both conspire to yield correlations between $F1$ and TD height. There are also conditions expected to give rise to correlations between $F2$ and TD backness. The nomograms of Stevens (1989) show that advancement of a vocal tract constriction will raise $F2$, if the constriction is in the posterior part of the vocal tract for a three-tube model (high vowels) or the anterior portion of the vocal tract for a two-tube model (low vowels). Outside of these regions, advancement of the TD position can have a minimal effect on $F2$ or even lower $F2$, e.g., anterior constrictions in a three-tube model or posterior constrictions in a two-tube model. From this theoretical standpoint, the stronger correlations between $F1$ and TD height than for $F2$ and TD backness in our study are not particularly surprising. More importantly, we can predict the conditions under which the simple heuristic will break down.

By and large, data from a single fleshpoint on the TD displayed articulatory patterning across vowels that correspond to those in the formants. In particular, the relative lingual height and backness of vowels at the TD sensor correlates with $F1$ and $F2$ values. Wieling *et al.* (2016) is one of the few studies that reports correlations between EMA data and formant values which can serve as a basis for comparing the strength of the correlations in our data. In their corpus of Dutch speakers from Ter Appel and Ubbergen, they report correlations between $F1$ and tongue height of $r = -0.22$ for one set of words and $r = -0.43$ for another. These correlations are much weaker than the $r = -0.78$ correlation found in our data. It is not clear what causes this discrepancy across studies, but there are several methodological differences that may play a role. Our correlations include just one pair of articulatory and acoustic data points per token. The measurements were made at the point of minimum velocity in the movement, a proxy for the target of controlled movement. Wieling *et al.* (2016) included multiple such pairings per token in their correlations. Accordingly the correlations represent both within-token and across-token variability. Another difference is that our vowels were produced in a consistent phonetic frame whereas the vowels in Wieling *et al.* (2016) came from a diverse range of phonetic environments. These methodological differences could have reduced the consistency with which a single fleshpoint is predictive of $F1$. There may also be real linguistic differences across Dutch and English that contribute to how well TD height corresponds to $F1$. The correlations between $F2$ and tongue backness across studies were more comparable. Wieling *et al.* (2016) report $r = -0.44$ for one set of words and $r = -0.63$ for another (cf. $r = 0.69$ in the current study). Due in part to the weak correlations, Wieling *et al.* cautions about interpreting $F1$ and $F2$ in terms of tongue position. We concur with this precaution, but we also seek to understand the conditions under which formant values are more or less predictive of TD position.

For starters, we found correlations between $F1$ and $F2$ in the front part of the vowel space that do not correspond closely to the vertical and longitudinal displacement of the TD. Differences in $F2$ amongst the front vowels within both English dialects investigated in the current study were a result of general properties of formant spaces: as $F1$ increases, $F2$ of front vowels also tends to decrease, leading to a diagonal distribution on the vowel quadrilateral. We assume that these differences in $F2$ amongst the front vowels are at least in part attributable to an intrinsic relationship between tongue height and pharyngeal area due to the conservation of tongue volume and, thus, may not be under speaker control to the same degree as $F2$ in other locations of the vowel space. Decreases in $F2$ as a function of $F1$ in the front part of the vowel space were consistent across dialects and speakers, regardless of TD backness.

As another general observation, the vowel space expressed in terms of TD position is more compact than the vowel space expressed in formants in that there was more overlap in the TD positional coordinates between some vowels than we observed in the formant plots.[2] From this we surmise that other aspects of vowel articulation function to modulate the impact that TD position has on vocal tract resonance. In some cases, we observed that vowels with similar tongue positions, e.g., KIT and GOOSE in AusE, are differentiated by lip rounding. Incorporating other aspects of articulation, e.g., jaw height (Stone and Vatikiotis-Bateson, 1995; Erickson, 2002), tongue shape (Dawson *et al.*, 2016), or additional data points on the tongue may provide a more dispersed view of the articulatory vowel space, and a closer correspondence to the acoustics. In what follows we consider some of these factors in the context of a broader discussion of the degree to which tongue position can be inferred from vowel formants.

Although linear correlations between TD position and formant values were stronger in our data than other similar studies, we observed individual differences in the strength of correlations. In particular, we described some discrepancies in the expected acoustic-articulatory relations for two of the male AusE speakers, M03 and M05. In the case of M03, lingual articulation for THOUGHT and LOT was similar to the other AusE speakers with THOUGHT more retracted than LOT, but his formant values showed a higher $F2$ for THOUGHT than LOT. This difference is consistent with the differences in lip rounding that we also observed. Unlike other AusE speakers, this speaker did not differentiate THOUGHT and LOT in rounding.

Another AusE speaker, M05, showed particularly weak linear correlations between $F2$ and TD backness. The $F2$ values for M05 tend to be above the regression line at high and low values of tongue backness and below the regression line at intermediate values, indicating a non-linear trend, which contrasts to the largely linear trend observed for other speakers (e.g., F07 in Fig. 9). Given the particular anatomy of M05, central vowels may fall into an area of stability such that variation in TD backness has little effect on $F2$. An alternative hypothesis is that something else is influencing $F2$ other than TD backness. One possibility may be tongue curvature. Tongue shape has been shown to differentiate the vowels of English (Dawson *et al.*, 2016). A more curved tongue would lead to a larger pharyngeal cavity which in turn would result in an increase in $F2$, due to an increase in the cross-sectional area of the back cavity (while holding constriction location constant). Further investigation would be needed to discover why $F2$ varies despite similar degrees of TD backness for these central vowels, and why this is the case in this part of the vowel space and for this speaker in particular.

Cases such as these underscore the indeterminacy of interpreting formant values in terms of lingual articulation, or at least with regard to a single fleshpoint. Because they are shaped by multiple articulatory constrictions in the vocal tract, it is not always possible to map changes in formants to changes in TD position.

B. Differences between dialects

Turning now to a comparison between dialects, we discover that whether acoustic or articulatory data are examined changes our conclusions about how vowels differ across NAmE and AusE. Differences between the dialects uncovered in this study using both methods are discussed below.

The acoustic results we reported for NAmE and AusE largely replicated past acoustic studies on these dialects. In terms of formants, all vowels except DRESS differ significantly across dialects in either $F1$ or $F2$. The front vowels were similar across dialects (although note that TRAP differs significantly on $F1$ and FLEECE differs significantly on $F2$). There were several differences in the central and back vowels between dialects. NAmE can be characterized acoustically (on the basis of $F2$) as having just one central vowel, STRUT, while AusE has three, STRUT, NURSE, and GOOSE. In the back vowels, our NAmE speakers tended to merge THOUGHT and LOT. Three speakers made no distinction and two showed overlapping distributions. All AusE speakers, on the other hand, maintained a clear distinction at least in $F1$. Three out of four AusE speakers also differentiated THOUGHT and LOT in $F2$, with THOUGHT having a lower $F2$ than LOT (we discussed lip rounding as the basis for this exception above). Thus, on the basis of the acoustic results, we partitioned the AusE vowel space into four front, three central, and three back vowels, or 4:3:3 and the NAmE vowel space into 4:1:4/5. Both dialects share STRUT as the central vowel but differ in whether the other non-front vowels are central or back.

Vowel spaces based on formants are sometimes assumed to have clear lingual articulatory correlates. We have found that the articulatory data reveals a different partition of the NAmE vowel space, which has implications for how the differences between dialects are characterized. The key difference between the acoustic and articulatory characterization of NAmE vowels involved the position of GOOSE and FOOT, which have low $F2$ values but central TD position. These are both rounded vowels (Fig. 4), and the rounding no doubt lowers $F2$ beyond what would be expected from TD position alone. In the absence of the comparison with AusE, we might even be tempted to conclude that lip rounding is the source of the discrepancy between formant values and TD position for these vowels in NAmE. Viewed in light of the comparison with AusE, this conclusion appears incomplete. GOOSE and FOOT are also rounded in AusE, and to a similar degree as in NAmE (Fig. 6). Despite similar degrees of rounding, there are significant differences across dialects. FOOT is a vowel that differs significantly in both $F2$ and TD backness—it is both further back and has a lower $F2$ in AusE than in NAmE. The comparison with AusE makes it clear that the TD position for FOOT in NAmE is more anterior, even though $F2$ is still relatively low. The same goes for GOOSE. It has a central TD position in NAmE. Across dialects, GOOSE showed significant differences in $F2$ (AusE GOOSE is higher in $F2$), without a corresponding difference in TD backness. Like FOOT, $F2$ for NAmE GOOSE is low despite an advanced (central) TD position. Thus, from an articulatory perspective, both dialects have GOOSE as a central vowel. In the formant space, however, GOOSE is back in NAmE and central in AusE. In this case, how to partition the vowels into front, central, and back depends on whether we refer to the vowel space based on TD position or the vowel space based on formants.

The case of GOOSE-fronting in NAmE without a corresponding rise in $F2$ highlights the need to incorporate articulatory parameters besides TD position and rounding into our understanding of formant variation and our description of dialects. In studies of English dialect variation, GOOSE-fronting refers to an increase of $F2$ in the GOOSE vowel such that the GOOSE category encroaches on FLEECE and KIT. In some dialects of English, GOOSE-fronting was initiated by high frequency words in which GOOSE is followed by a coronal stop (e.g., Derby English; Sóskuthy *et al.*, 2015). Given this environment, GOOSE-fronting is thought to be driven by a coarticulatory effect of the tongue being pulled forward during the vowel in anticipation of the coronal articulation (Harrington *et al.*, 2008). Viewing the acoustic data together with the articulatory data presents a more nuanced view. GOOSE is more front in the acoustic data for AusE than NAmE, in that the $F2$ value for GOOSE is closer to the $F2$ values of FLEECE and KIT in AusE than in NAmE. In articulation, NAmE GOOSE is also "fronted," i.e., not significantly different from AusE.

To better understand how NAmE GOOSE could be fronted articulatorily without raising $F2$, we explored tongue curvature, which we computed by fitting a second-order polynomial to the three sensors on the tongue. Figure 10 shows the mean positions of the lingual sensors with the fitted polynomial for a subset of vowels: NURSE, FOOT, GOOSE, and FLEECE. Figure 10(a) (left) shows the NAmE data; Fig. 10(b) (right) shows AusE. As illustrated in Fig. 10(a), FLEECE is more curved than GOOSE in NAmE. The mean quadratic term for NAmE FLEECE is -1.54 (SD $= 0.45$) and the mean quadratic term for GOOSE $= -0.46$ (SD $= 0.43$). In AusE [Fig. 10(b)], the difference in curvature between FLEECE and GOOSE is not as large. The mean quadratic term for AusE FLEECE is -1.6 (SD $= 0.35$), cf. mean quadratic term for GOOSE $= -1.01$ (SD $= 0.39$). We confirmed the statistical significance of dialect differences in curvature by fitting a mixed effects model with dialect as a fixed factor (and speaker and order as random effects) to the quadratic term from the polynomial function for GOOSE tokens and for FLEECE tokens. AusE GOOSE was significantly more curved than NAmE GOOSE [$\beta = 0.965$, SE $= 0.224$, $t(7) = 4.303$, $p = 0.0035$] but the effect of dialect on FLEECE was not significant (i.e., FLEECE was not significantly more curved in one dialect than the other). Based on this result, it is tempting to conclude that the differences in $F2$ for GOOSE across dialects are attributable (at least in part) to differences in tongue curvature, at least for the subjects reported here. A more curved tongue may be indicative of a larger pharyngeal cavity, which would have the effect of increasing $F2$. We note in this context that other factors, including palate shape, may also influence tongue curvature (Lammert *et al.*, 2013). Although we cannot provide conclusive evidence, the larger $F2$ difference between GOOSE and FLEECE in NAmE than in AusE may be a consequence of tongue shape, rather than tongue position.

The vowels NURSE and FOOT offer additional cases in which curvature appears to be modulating the relation between formant values and TD position. These vowels are not distinguished acoustically in $F1$ and $F2$ for NAmE. Figure 3(a) shows the degree of overlap for NURSE and FOOT in NAmE in the formant space and Figs. 10(a) and 10(b) show the tongue shape for NURSE to be more curved than for FOOT, with a higher tongue that is also more back, which is the general pattern for NAmE. In contrast, these two vowels are clearly differentiated acoustically in AusE. From the formant plots alone, we might conclude that these vowels are similar in NAmE but different in AusE. The articulatory data reveal clear differences between the vowels in both dialects. The reason why they are merged in NAmE $F1/F2$ formant space is likely that being high and curved (NURSE) offsets the effect of TD backness on $F2$. Thus, NURSE is further back than FOOT even though these vowels have a similar $F2$. Given the lowered $F3$ found for NURSE, TD retraction likely corresponds to a constriction at the soft palate and/or pharynx (but see Espy-Wilson *et al.*, 2000 for claims that the pharyngeal constriction is not responsible for lowering $F3$ in American English /ɹ/).

More broadly, we can see that curvature is an articulatory parameter on which vowels and dialects differ. In NAmE, FLEECE and NURSE are curved; GOOSE and FOOT are not. In AusE, FLEECE, GOOSE, and NURSE pattern together as curved to the exclusion of FOOT. Incorporating curvature into our description allows us to observe a similarity across dialects in the NURSE vowel that we would have missed otherwise. NURSE was the only vowel that was significantly different on $F1$, $F2$, TD height, and TD backness, all four of the dependent variables reported in Fig. 9. Despite these numerous differences as well as a difference in rhoticity, NURSE is curved in both dialects.

As we reported in Sec. I, past research had identified differences between GOOSE and NURSE vowels across dialects. Both GOOSE and NURSE have a higher $F2$ in AusE

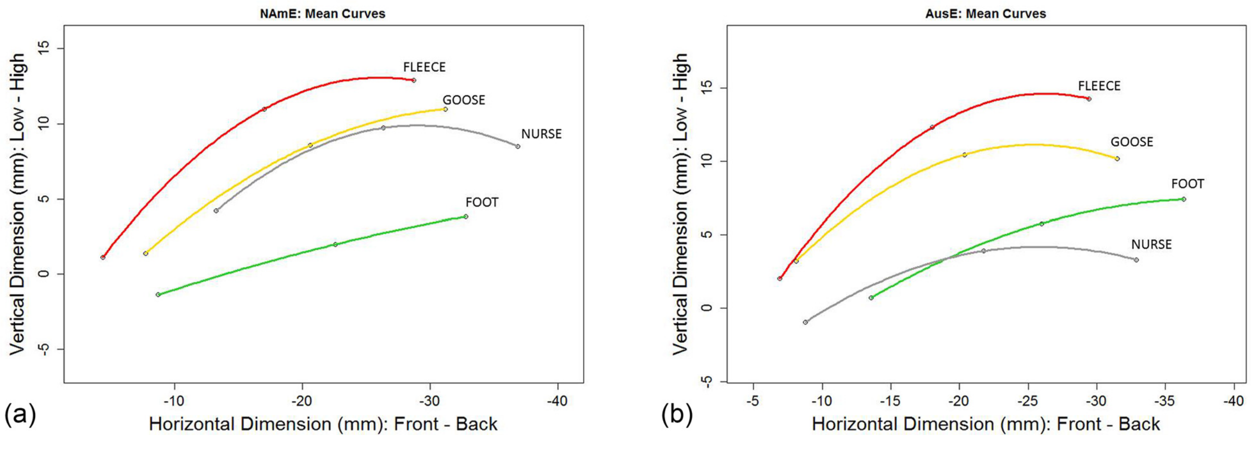

FIG. 10. (Color online) (a) displays mean tongue curves for four NAmE vowels: FLEECE, GOOSE, NURSE, and FOOT. (b) displays tongue curves for the corresponding AusE vowels. Each point indicates a lingual sensor. The three circles on each curve represent the three lingual sensors (from left to right: TT, TB, and TD).

than in NAmE. According to our results, it is clear that the increase in $F2$ in AusE compared to NAmE has a different articulatory basis for the two vowels. The articulatory bases of $F2$ differences for NURSE are tongue height and backness, as this vowel is curved in both dialects. For GOOSE, the difference in $F2$ must be attributed to some other articulatory property, which we have suggested is tongue curvature.

Although there were other vowels besides GOOSE for which significant differences in formants do not correspond to significant differences in TD position, the reverse was rare. A significant difference in TD position typically implied a significant difference in formant values. The one exception to this trend was the THOUGHT vowel. For this vowel, there was a significant difference in TD backness (AusE is further back than NAmE) but no difference in $F2$. Moreover, THOUGHT is rounded in AusE but not NAmE, which should, if anything, further increase the difference in $F2$ expected on the basis of tongue backness. Thus, the THOUGHT vowel is a clear case in which the simple heuristic relating $F2$ to tongue backness breaks down. Assuming that the vocal tract is partitioned into two cavities for THOUGHT with the partition leaving the front cavity larger than the back cavity, the heuristic fails because the theoretical basis for the $F2$ by TD backness correlation is not valid for this configuration. The TD position for THOUGHT in both dialects may fall within a region of stability within which variation in articulatory position exacts little influence on $F2$ (Stevens, 1989). More likely, however, the constriction in AusE is posterior to this quantal region. The gradual advancement of a relatively posterior constriction in a two-tube model is predicted to lower (not raise) both $F2$ and $F1$. We observe the predicted effect on $F1$. $F1$ is lower for AusE, which has the more retracted TD position. We do not observe $F2$ differences for THOUGHT across dialects, but this may be because the influence of TD backness on $F2$ is offset by lip rounding. Overall, then, while THOUGHT defies the simple heuristic relating TD backness to $F2$, the articulatory-acoustic relation for this vowel and the differences across dialects are well-behaved from the theoretical foundations from which the simple heuristic was formulated.

We conclude the discussion by acknowledging some limitations of the present study. The number of speakers ($n = 9$) was one limiting factor which could be remedied in future studies. However, reported multiple tokens for each of the ten monophthongs make this the largest study of parallel acoustic and articulatory data of AusE. Our four AusE speakers were from the same region, but the five NAmE speakers recorded for comparison came from different parts of North America. Overall, the NAmE data showed more variation across speakers than the AusE data, which is likely due at least in part to the regional heterogeneity of the NAmE group. It is also possible that the NAmE group reported here may have been influenced by their time in Australia (Campbell-Kibler *et al.*, 2014). Including speakers residing in North America might be a more reliable way of uncovering dialectal differences. However, we also note that some vowels may remain stable even after moving countries. For example, Nycz (2013) reported stability in low back vowel realization for mobile Canadians. Nevertheless, at least one of our NAmE speakers showed signs of adopting AusE vowels in both acoustic and articulatory data, so including speakers residing in North America with less regional variation might have resulted in clearer differences between the two groups than were reported here. The speaker variation for NAmE may also have strengthened some of the correlations we reported. From the standpoint of assessing the acoustics-articulation relation, variation in articulation provides a way to "sample" the space of possible articulations while observing the acoustic consequences. These limitations notwithstanding, the comparison offered in the current study presents an informative case study both of how articulatory data can enhance dialect description but also of how dialect variation can provide an informative domain for advancing understanding of the acoustics-articulation relation in speech.

V. CONCLUSIONS

How the vowel space is described depends on what type of data are examined, and this in turn influences conclusions made about dialect differences. For two dialects, NAmE and AusE, we reported acoustic vowel spaces based on formants ($F1$ and $F2$) and articulatory vowel spaces based on the position of the TD. Several differences—such as the merger of LOT and THOUGHT in NAmE but not AusE—are reflected clearly in both types of data. Others are not. For example, the GOOSE vowel is central in both dialects if viewed articulatorily, but on the basis of $F2$, it is back in NAmE and central in AusE. In general, significant differences in TD position corresponded to significant differences in formant values, but the reverse was not always found. Differences in formants not reflected in TD position underscore the role that other aspects of articulatory control have on formant values. In particular, we demonstrated several cases in which lip rounding and tongue curvature (as a proxy for pharyngeal cavity size) plausibly perturb correspondence between TD position and formants.

With regard to the relationship between acoustics and articulation often assumed in dialect descriptions—namely, that $F1$ is inversely related to vowel height and $F2$ is inversely related to backness—we confirmed both this general trend as well as some predicted exceptions and individual differences. There were significant linear correlations across all of the data for $F1$ and TD height and weaker correlations for $F2$ and TD backness, but we also found that the strength of the linear relation was stronger for some speakers than for others and that it breaks down in some regions of the vowel space, e.g., low back vowels. We argue that both the general trend and the exceptions follow from the theoretical bases of the common heuristic. Thus, while formants are shaped by too many factors to be predicted by TD position alone and TD position cannot be accurately inferred from formants, the partial correspondence is highly encouraging. Not only does articulatory data enhance the description of dialect variation, but variation across dialects offers an insightful probe into the relation between speech acoustics and articulation.

ACKNOWLEDGMENTS

Research was supported in part by Australian Research Council Grant No. DE120101289 to J.A.S. The researchers would like to thank all the speakers for their participation in this study. We would like to acknowledge colleagues for their assistance with the EMA experiments and post-processing of the data, in particular, Donald Derrick and Christian Kroos. We would also like to thank Toni Cassisi for his technical assistance. We wish to thank attendees of the Australasian Speech Science and Technology conference (December 2016), Cynthia Clopper, and two anonymous reviewers for helpful comments that greatly improved the paper.

APPENDIX

Analysis of variance table of type III with Satterthwaite approximation for degrees of freedom **F1**: F1 ∼ vowel * dialect + (1|speaker) + (1|order).

	Sum Sq	Mean Sq	Num DF	Den DF	F-value	P-value
Dialect	0.00	0.002	1	633	0.03	0.8687
Lexical set	556.84	61.871	9	633	808.24	<0.001
Dialect:Lexical set	34.49	3.833	9	633	50.07	<0.001

Analysis of variance table of type III with Satterthwaite approximation for degrees of freedom **Tongue Dorsum Height**: TD_height ∼ vowel * dialect + (1|speaker) + (1|order)).

	Sum Sq	Mean Sq	Num DF	Den DF	F-value	P-value
Dialect	0.00	0.003	1	6.99	0.03	0.8769
Lexical set	403.34	44.816	9	624.96	395.98	<0.001
Dialect:Lexical set	31.11	3.457	9	624.96	30.54	<0.001

Analysis of variance table of type III with Satterthwaite approximation for degrees of freedom **F2**: F2 ∼ vowel * dialect + (1|speaker) + (1|order).

	Sum Sq	Mean Sq	Num DF	Den DF	F-value	P-value
Dialect	0.002	0.0025	1	633	0.005	0.9434
Lexical set	153.172	17.0191	9	633	34.409	<0.001
Dialect:Lexical set	167.814	18.6460	9	633	37.699	<0.001

Analysis of variance table of type III with Satterthwaite approximation for degrees of freedom **Tongue Dorsum Backness**: TD_height ∼ vowel * dialect + (1|speaker) + (1|order).

	Sum Sq	Mean Sq	Num DF	Den DF	F-value	P-value
Dialect	0.000	0.0001	1	6.96	0.005	0.9467
Lexical set	36.756	4.0840	9	624.67	172.951	<0.001
Dialect:Lexical set	9.026	1.0029	9	624.67	42.473	<0.001

Bernard, J.-B. (**1970**). "A cine-X-ray study of some sounds of Australian English," Phonetica **21**(3), 138–150.

Berry, J. (**2011**). "Accuracy of the NDI wave speech research system," J. Speech, Lang., Hear. Res. **54**(5), 1295–1301.

Campbell-Kibler, K., Walker, A., Elward, S., and Carmichael, K. (**2014**). "Apparent time and network effects on long-term cross-dialect accommodation among college students," Univ. Penn. Work. Papers Linguist. **20**(2), 4, available at http://repository.upenn.edu/pwpl/vol20/iss2/4.

Cheshire, J., Kerswill, P., Fox, S., and Torgersen, E. (**2011**). "Contact, the feature pool and the speech community: The emergence of Multicultural London English," J. Sociolinguist. **15**(2), 151–196.

Chiba, T., and Kajiyama, M. (**1941**). *The Vowel: Its Nature and Structure* (Tokyo-Kaiseikan Publishing Co., Tokyo, Japan).

Cox, F. (**1999**). "Vowel change in Australian English," Phonetica **56**(1–2), 1–27.

Cox, F. (**2006**). "The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers," Austr. J. Linguist. **26**(2), 147–179.

Dawson, K. M., Tiede, M. K., and Whalen, D. (**2016**). "Methods for quantifying tongue shape and complexity using ultrasound imaging," Clin. Linguist. Phonetics **30**(3–5), 328–344.

Erickson, D. (**2002**). "Articulation of extreme formant patterns for emphasized vowels," Phonetica **59**(2–3), 134–149.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., and Alwan, A. (**2000**). "Acoustic modeling of American English /r/," J. Acoust. Soc. Am. **108**(1), 343–356.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Georgeton, L., Kocjančič, T., and Fougeron, C. (**2014**). "Domain initial strengthening and height contrast in French: Acoustic and ultrasound data," in *10th International Seminar on Speech Production*, pp. 142–145.

Harrington, J., Cox, F., and Evans, Z. (**1997**). "An acoustic phonetic study of broad, general, and cultivated Australian English vowels," Austr. J. Linguist. **17**(2), 155–184.

Harrington, J., Kleber, F., and Reubold, U. (**2008**). "Compensation for coarticulation, /u/ fronting, and sound change in standard southern British: An acoustic and perceptual study," J. Acoust. Soc. Am. **123**(5), 2825–2835.

Harrington, J., Kleber, F., and Reubold, U. (**2011**). "The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English," J. Int. Phonetic Assoc. **41**(02), 137–156.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111.

Johnson, K., Ladefoged, P., and Lindau, M. (**1993**). "Individual differences in vowel production," J. Acoust. Soc. Am. **94**(2), 701–714.

Lammert, A., Proctor, M., and Narayanan, S. (**2013**). "Interspeaker variability in hard palate morphology and vowel production," J. Speech, Lang., Hear. Res. **56**(6), S1924–S1933.

Lin, S., Palethorpe, S., and Cox, F. (**2012**). "An ultrasound exploration of Australian English /CVl/ words," in *Proceedings of the 14th Speech Science and Technology Conference*, edited by F. Cox, K. Demuth, S. Lin, K. Miles, S. Palethorpe, J. Shaw, and I. Yuen, Australasian Speech Science and Technology Association, Sydney, Australia, pp. 105–108.

Lobanov, B. (**1971**). "Classification of Russian vowels spoken by different listeners," J. Acoust. Soc. Am. **49**, 606–608.

Noiray, A., Iskarous, K., and Whalen, D. (**2014**). "Variability in English vowels is comparable in articulation and acoustics," Lab. Phonol. **5**(2), 271–288.

Nycz, J. (**2013**). "New contrast acquisition: Methodological issues and theoretical implications," English Lang. Linguist. **17**(02), 325–357.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**(2), 175–184.

Scobbie, J. M., Lawson, E., and Stuart-Smith, J. (**2012**). "Back to front: A socially-stratified ultrasound tongue imaging study of Scottish English /u/,"

[1] The NDI Wave system has an automatic head-correction procedure. This was accidentally applied during recording of two of our NAmE speakers, rendering the data relative to the right mastoid sensor. After rotating this data to the occlusal plan, the location of the sensors within our reconstructed coordinate system differed systematically from the other speakers. The normalization step described in our analysis section rendered all data relative to the center of the articulatory space, correcting for differences introduced in post-processing.

[2] It must be noted here that this does not necessarily mean that there was less variability in the articulation for each vowel as compared to the formants.

Rivista di Linguistica/Italian Journal of Linguistics, Special Issue: Articulatory Techniques for Sociophonetic Research **24**(1), 103–148.

Shaw, J. A., Chen, W. R., Proctor, M. I., and Derrick, D. (**2016**). "Influences of tone on vowel articulation in Mandarin Chinese," J. Speech, Lang., Hear. Res. **59**(6), S1566–S1574.

Shaw, J. A., Tyler, M. D., Kasisopa, B., Ma, Y., Proctor, M. I., Han, C., Derrick, D., and Burnham, D. K. (**2013**). "Vowel identity conditions the time course of tone recognition," in *INTERSPEECH*, pp. 3142–3146.

Sóskuthy, M., Foulkes, P., Haddican, B., Hay, J., and Hughes, V. (**2015**). "Word-level distributions and structural factors codetermine GOOSE fronting," in *Proceedings of the 18th International Congress of Phonetic Sciences*, edited by The Scottish Consortium for ICPhS, The University of Glasgow, Glasgow, United Kingdom, pp. 1001–1006.

Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonet. **17**, 3–45.

Stone, M., and Vatikiotis-Bateson, E. (**1995**). "Trade-offs in tongue, jaw, and palate contributions to speech production," J. Phonet. **23**(1), 81–100.

Tabain, M. (**2008**). "Production of Australian English language-specific variability," Austr. J. Linguist. **28**(2), 195–224.

Watson, C. I., Harrington, J., and Palethorpe, S. (**1998**). "A kinematic analysis of New Zealand and Australian English vowel spaces," paper presented at the *5th International Conference on Spoken Language Processing*, Vol. 6, pp. 2363–2366.

Watt, D., and Tillotson, J. (**2001**). "A spectrographic analysis of vowel fronting in Bradford English," Eng. World-Wide **22**(2), 269–303.

Wells, J. C. (**1982**). *Accents of English*, Vol. 1 (Cambridge University Press, Cambridge), pp. 117–183.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S., and Baayen, R. H. (**2016**). "Investigating dialectal differences using articulography," J. Phonet. **59**, 122–143.