Positional targets for lingual consonants defined using electromagnetic articulography

Yunusova, Yana
Department of Speech-Language Pathology, University of Toronto
160 - 500 University Avenue,
Toronto, ON, Canada, M5G 1V7

Rosenthal, Jeffrey S.
Department of Statistics, University of Toronto
100 St. George Street, Room 6018
Toronto, ON, Canada M5S 3G3

Rudy, Krista
Department of Speech-Language Pathology, University of Toronto
160 - 500 University Avenue,
Toronto, ON, Canada, M5G 1V7

Baljko, Melanie
Department of Computer Science and Engineering, York University
4700 Keele Street,
Toronto, Ontario, Canada, M3J 1P3

Daskalogiannakis, John
Department of Orthodontics, University of Toronto
124 Edward Street,
Toronto, ON, Canada M5G 1G6

Correspondence:

Yana Yunusova
Department of Speech and Language Pathology
Rehabilitation Sciences Building
University of Toronto
160 - 500 University Avenue
Toronto, ON M5G 1V7
Tel: 416 978-6890
Fax: 416 978-1596
yana.yunusova@utoronto.ca

ABSTRACT

The study examined the positional targets for lingual consonants defined using a point-parameterized approach with Wave (NDI). The overall goal was to determine which consonants had unique tongue positions with respect to other consonants. Nineteen talkers repeated VCV syllables that included consonants /t, d, s, z, ʃ, ʧ, k, g/ in symmetrical vowel contexts /i, u, a/, embedded in a carrier phrase. Target regions for each consonant, characterized in terms of x,y,z tongue positions at the point of maximum tongue elevation, were extracted. Distances and overlaps were computed between all consonant pairs and compared to the distances and overlaps of their contextual targets. Cognates and postalveolar homorganics were found to share the location of their target regions. On average, alveolar stops showed distinctively different target regions than alveolar fricatives, which in turn showed different target region locations than the postalveolar consonants. Across talker variability in target locations was partially explained by differences in habitual speaking rate and hard palate characteristics.

I. INTRODUCTION

Mapping positions of the articulators in the vocal tract to different speech sound categories is not trivial. Positional variability is associated with a myriad of factors including those of the linguistic, prosodic, anatomic, physiological and acoustic origin. Apparent extensive positional variability of articulators during speech led to extensive theoretical discussions on the nature of the unit and mechanism of speech motor control. Arguably, the theory development is currently limited by a lack of kinematic data describing speech articulation. The need to understand articulatory behaviours goes beyond advancing theory, however. Emergence of the instrumental approaches for real-time tongue tracking allows development of new applications for speech recognition and clinical intervention purposes, for which normative reference data are crucial.

A traditional linguistic theory views the place of articulation as one of the essential parameters in the description of the patterns of controlled movements for consonants (Ladefoged & Maddieson, 2008). English lingual consonants are divided, from front to back, into alveolar, postalveolar (or palatal) and velar categories, by the location of the tongue to palate constriction. Cognates (i.e., pairs /d-t/, /s-z/, and /k-g/) are presumed to share the place and manner of articulation but differ by voicing. Similarly, homorganic consonants (e.g., alveolars /d, t, s, z/ or postalveolars /ʃ, ʧ/) are assumed to share their place of articulation, despite differences in the manner of production and voicing. When linguagrams and palatograms of coronal consonants (/d, t, n, l, s, z,/) were visually examined, however, the observed contact patterns showed notable variability in consonant place of articulation between talkers of the same language (e.g., English or French) and a significant interaction between the place and manner of articulation, resulting in different pattering of lingua-palatal contact for stops and fricatives (Dart, 1998). The patterns of lingua-palatal contact and place of articulation for consonants have been quantified with electropalatography [EPG], with normative expectations for this method relatively well established (Dagenais & Critz-Crosby, 1991; Fletcher, 1989; Liker & Gibbon, 2008; Gibbon, Yuen, Lee, & Adams, 2007). In broad terms, the EPG studies reported similarities of tongue contact patterns across control talkers producing the same consonants and cognates as well as differences in tongue to palate contact between consonant classes (e.g., stops and fricatives) produced with the same place of articulation.

The goal of this study was to develop a set of expectations regarding tongue positions during lingual non-sonorant consonants using a point-parameterized electromagnetic tracking technology (e.g., Wave, EMMA). Ever since point-parameterized methods for studying speech movements became available, the positional targets that are reached by the tongue during speech have been defined in terms of ranges of acceptable positions (target regions) in 2- or 3-dimensional space rather than as invariant points (see Guenther, 1995; Keating, 1990; Perkell & Klatt, 1986). If these ranges for vowels and approximants (e.g., /r/) have been systematically studied (Hasegawa-Johnson, Pizza, Alwan, Cha, & Haker, 2003; Guenther et al., 1999; Perkell, 1996; Perkell & Nelson, 1982; 1985), "the situation for consonants is less clear at present" (Guenther et al., 1998; p.613). One of the leading models of speech production, the Directions into Velocities of Articulators model (DIVA), currently models consonant target regions similar to those for vowels and semivowels (Guenther, 1995; Guenther,Gosh, & Tourville, 2006). Consonants, however, might be fundamentally different in their control and production than vowels and semi-vowels (Bailly, 1997; Öhman, 1967). [1]

[1]

Variability of the tongue positions during repeated productions of speech sounds (in other words target size) has been influenced by contextual, prosodic, and speaker-related factors, including the shape of the hard palate (Brunner, Fuchs, & Perrier, 2009; Dembowski, Lindstrom, & Westbury, 1998; Houde, 1967; Guenther, 1995; Keating, 1990; Recasens & Espinosa, 2009). The effect of adjacent sounds on positional variability of the tongue in point-parameterized methods has been studied in most detail in the context of coarticulation research, specifically in the degree of articulatory constraints model [DAC] (see Farnetani & Recasens, 2010 for an extensive review). Recasens and Espinosa (2009) used articulography to capture differences in the variability measures between consonants in symmetrical VCVs. Their findings supported the DAC model, which proposed that the extent of the tongue position variability during a consonant varied primarily based on its susceptibility to co-articulatory influences of the adjacent vowels, which in turn was related to the biomechanical constraints imposed on the primary articulator for the consonant and for the vowel (Recasens et a., 1997). The locations of the positional targets relative to other targets for different consonants were not specified in this study.

In an early cinefluographic study of two talkers, Kent and Moll (1972) reported minimal effect of vowels (/i, a, u/) and thus contextual variability of tongue marker positions during selected lingual consonants, linking tongue positions with phonemic identification. In contrast, a later study by Dembowski and colleagues (1998) reported overlapping distributional properties of tongue positions during stops embedded in a variety of sound sequences in a relatively large number of speakers recorded with x-ray microbeam. The authors questioned the categorical nature of differentiation between allophones of the same phoneme or between phonemes based on kinematic data (Dembowski, et al.,1998). However, in their study the phonetic context varied greatly between sounds, preventing a direct comparison of positional variability and target locations between the stops. Recently, locations for the tongue tip sensor were examined by Mooshammer and colleagues in a study of the jaw and tongue interactions during various lingual consonants in a single vowel context (Mooshammer, Hoole & Geumann, 2007). These authors showed relatively different target regions for /t, s, ʃ/ as compared to highly overlapping regions for the /n, d, l/. Beyond these studies, the positional targets for various lingual consonants remain non-established.

The main goal of this study was to compare tongue positions during different lingual consonants as represented by a point-parameterized technique and to identify which consonants occupy distinctively unique locations in the vocal tract by examining their locations with respect to their own contextual targets and other consonants. Based on the traditional linguistic view of consonant place of articulation, we hypothesize that cognates and homorganic consonants would be characterized by similar positional target regions. In contrast, alveolars and postalveolars would be characterized by unique target region locations.

II. METHOD

A. Participants

Nineteen speakers (F=9; M=10) provided data for this study. The average age was 28.5 (SD=6.1) for the female participants and 32.8 (SD=8.5) for the male participants. All participants were native speakers of Canadian English. Fourteen participants spoke with a dialect typical for Southern Ontario. The remaining five talkers were from British Columbia. Participants reported no history of speech, language, or hearing abnormalities. Examination of the oral cavity revealed no gross abnormalities of the mouth, including normal dentition, without missing teeth or dental appliances. All participants passed a standard hearing screening.

As part of an earlier study (Yunusova et al., 2012), palatal measures were obtained for each talker. Briefly, the palate morphology was characterized by obtaining the coordinates of specific landmarks identified on the palate cast (Brunner et al., 2005; 2009; Ferrario, Sforza, Schmitz, and Colombo, 1998). Specific landmarks included: (1) incisive papilla (IP), the point between the maxillary incisors; (2) molar right (MR), the point on the gingival margin at the first permanent molar on the participants right side; (3) molar left (ML), the point on the gingival margin at the first permanent molar on the participant's left side; and (4) midpoint (M), the center point between MR and ML on the palate surface. Palate morphology (size, shape) was represented by the following measurements: (1) palate height, defined as the shortest distance between M and the line between ML and MR; (2) palate width, defined as the distance between points MR and ML; (3) palate length, defined as the horizontal distance between IP and M; (4) palate slope, defined as the inclination of the straight line between IP and M; and (5) palate curvature (α), defined as the coronal shape of the hard palate and measured by tracing the palate cast between points MR and ML and fitting a parabola to represent the shape. The coefficient *a* from the line equation was used to calculate α using the previously published formula (see Perrier, Boe, and Sock, 1992; see also Brunner et al., 2005; 2009).

Table 1 provides a summary of these palatal measures for the male and female groups, separately. Because we cannot predict which of the palate measures would affect the consonant target location variability across talkers, we used all of these measures as potential covariates in the current study.

B. Speaking Task

Talkers were asked to read symmetrical vowel-consonant-vowel (VCV) syllables embedded into a carrier phrase "It's VCV game" with stress placed on the second CV syllable. The syllables contained three corner vowels (/i, u, a/) and nine consonants associated with different places and manners of production (/t, d, s, z, ʃ, ʧ, k, g/). The consonants represented all places of articulation for English lingual consonants. The vowels were chosen to expand the space where consonants are produced in the vocal tract, due to co-articulatory influences. The phrase was selected to simplify acoustic boundary identification. Each utterance was repeated ten times, producing 30 measurable events for each consonant (Consonant * 3 Vowels * 10 repetitions). Simple VCV utterances were chosen as an initial step in this analysis as a typical speech treatment begins with simple syllables (see Dagenais & Critz-Crosby, 1991).

The carrier phrases were produced at the talker's habitual comfortable speaking rate. Habitual speaking rate is known to vary between individuals and may affect the variability of speech movement (Tasko & McClean, 2004; Tsao &Weismer, 1997). We measured the speaking rates (R) in our sample of talkers by determining the average duration of the carrier phrases for each talker and dividing it by the number of syllables in the phrase. The average speaking rate was 271.37 ms/syl (SD=36.64), with minimum at 219.71 (for W18) and maximum at 360.77 (for W02). Speaking rate was used as a covariate in the analyses of consonant target locations.

C. Data Collection and Post-processing

Kinematic signals were collected with the Wave articulography system [NDI, Canada] (Berry, 2011). This electromagnetic system sampled movements of sensors in the electromagnetic field of the device at 100 Hz in three dimensions (3D). Two small 5 Degree Of Freedom (DOF) sensors (3 positional + 2 rotational coordinates) were attached. Sensor TF (tongue front) was placed on th e tongue body at midline, approximately 1 cm away from the tongue tip (*M* = 1.3 cm, *SD* = 0.3 cm). Sensor TB (tongue back) was placed 2 cm away from sensor TF (*M* = 2.1 cm, *SD* = 0.4 cm). The TF sensor was associated with the production of alveolar and palatal consonants; the TB sensor was associated with the production of velar consonants.

One 6 DOF sensor (3 positional + 3 rotational coordinates) was attached to the bridge of the nose and used to record head movements. The movements of the tongue were collected relative to the head during acquisition using NDI WaveFront software. The tongue movements were re-expressed relative to an anatomically-based Cartesian coordinate system with abscissa located along the plane between the maxilla and the mandible and the plane normal to it, with the origin at the central maxillary incisors (Westbury, 1994). This was accomplished using a bite plate recording (see Wesbury, 1994). Post-acquisition, acoustic and kinematic signals were examined for perceptual errors and movement mistrackings. Only syllables with perceptually correct productions and those without the tracking errors were used in the analyses (less than 1% of data contained such errors). Kinematic signals were low-pass filtered at 15Hz using a zero-phase digital filter (8-pole Butterworth). Relevant kinematic channels were parsed based on acoustic landmarks (mid V1 to mid V2).

D. Measurements

The point of maximum elevation of the tongue in the vertical dimension during the consonant was identified during the parsed segment and X, Y, and Z coordinates of this point were extracted algorithmically. X, Y, and Z coordinates for each repetition formed a point cloud referred to also as a “consonant target region”. An illustration of these targets for selected consonants for a single talker is given in Figure 1. *Consonant targets* were defined as a 3D space enclosed by 2SD ellipsoids fit around all of the repetitions of the same sound across all three contexts (see consonants /s/. /t/, and /ʃ/). The insert on the right side of Figure 1 shows similar 2SD ellipses fit around *contextual targets* for /ʃ/ (i.e., /iʃi/, /aʃa/, and /uʃu/). Note that these contextual targets are context-specific (non-overlapping) as they are under the influence of the surrounding vowels (Kent & Moll, 1972), yet closely adjacent.

Insert Figure 1 about here

The following measures were derived for each consonant and contextual target region:

1. *Distance (D)* was the measure of the Euclidian distance between the centers of two target regions. The measure represented relative location of one target with respect to another. Two types of distances were computed:
    a. *D1* was the Euclidian distance between the mean of each contextual target and the mean of the consonant target region computed across all three contexts (see Figure 1, insert). This measure showed how far the contextual targets were located from the center of the overall consonant distribution. This measure was used by others to represent the degree of coarticulatory resistance for consonants (Recasens & Espinosa, 2009).
    b. *D2* was defined as the Euclidian distance between the means of two different consonant targets (see Figure 1). This distance was used to show how far the consonant targets were located from each other.

2. *Overlap (O)* was the measure of the extent of similarity between the probability distributions of the X, Y, Z data for pairs of target regions (see Appendix 1 for overlap calculations).
   a. *O1* represented overlaps between the densities of the contextual targets. These overlaps provided a measure of the amount of influence of context on the target's density function.
   b. *O2* represented an overlap between pairs of consonants. These overlaps provided a measure of similarity between consonant pairs.

Intuitively, this overlap is a measure of the amount of similarity between the two probability distributions. More formally, this overlap is equal to one minus *"the total variation distance"* between the probability distributions having densities *f* and *g* respectively, and thus has many equivalences (see e.g. Proposition 3 of Roberts and Rosenthal, 2004). For example, the overlap between *f* and *g* is equal to the largest possible value of Prob[X=Y], over all random variables X and Y such that X has probability density *f*, and Y has probability density *g*. So, if *f* = *g*, then the overlap always equals one. By contrast, if *f* and *g* are completely disjoint, then the overlap equals zero. The maximum degree of overlap for our data sample, composed of discrete points in 3D space, was estimated by computing the overlap between randomly split data points for each consonant. The maximum overlap was 0.64 (0.12) for the front consonants and 0.70 (0.10) for the back (velar) consonants.

E. Statistical Analysis

We tested whether or not different pairs of consonants could be considered to come from distinct point clouds. Intuitively, they are distinct if the distance between a pair of consonant targets (D2) is significantly larger than the distance between the mean locations of the contextual targets with respect to the center of their combined distribution (D1). Additionally, the consonant targets are distinct if the overlaps between the probability densities for the pairs of consonants (O2) are significantly smaller than the overlaps between the densities for individual consonants in different contexts (O1). In other words, if the distances D2 are significantly larger than those of D1, or if the overlaps O2 are significantly smaller than those of O1, then we conclude that the corresponding pairs of consonants do indeed have distinct target regions. If they are not, then we conclude that the corresponding pairs of consonants have target regions, which can be regarded as the same.

For each analysis, the measures were grouped by a consonant category (e.g., cognate pairs, homorganic pairs). For example, when a pair of cognates was tested, D1 and O1 included all distances and overlaps, respectively, between contextual targets for each of the consonants (N1: 2 consonants x 3 contexts x 19 talkers=114). D2 and O2 included N2=19 talkers. Once a consonant pair was judged to share a common location (e.g., alveolar stops in the cognate analysis), then the pair could be "collapsed" or regarded as a unit in further comparisons. For example, when homorganic pairs were tested, we combined /d/ and /t/, and /z/ and /s/, considering all four pairs d-z, d-s, t-s, and t-z as a unit. Thus, D1 & O1 included all alveolar pairs (N1: 4 pairs x 3 contexts x 19 talkers=228), and D2 & O2 included 19 talkers x 4 alveolar pairs (N2=76).

The D1, D2, O1, and O2 distributions were first examined visually; the data appeared not to be normally distributed. Therefore, a non-parametric statistical test, specifically a one-sided Mann-Whitney-Wilcoxon rank-sum test, was used for comparisons (Wilcoxon, 1945; Mann and Whitney, 1947). This test ranks all of the distances in the two samples in numerical order, and then computes a rank-sum statistic U equal to the sum of the ordinal positions of values from the first list. This statistic is then used to produce a *p*-value for the null hypothesis that the two lists' underlying distributions have equal means, versus the alternative hypothesis that the mean of D2 is larger than that of D1 (or, that the mean of O2 is smaller than that of O1). If this p-value is less than 0.05, then it indicates a statistically significant distinction between the consonants; if it is more than 0.05, then no such distinction can be concluded.

Between-talker variability in D2 and O2 measures was tested with respect to a number of possible covariates. They included the individual's age, sex, dialect, speaking rate, and palatal size and shape (i.e., height, width, length, slope and curvature). Linear multiple regressions and correlations were used in these analyses.

III. RESULTS

A. Analysis of Cognates

Figure 2 shows two representative examples of cognate data. Note the overlapping target regions for cognate pairs /s-z/, /d-t/ and /k-g/ in both talkers. Table 2a and 2b reports summary statistics computed across talkers for D1 & D2 and O1 & O2 measures, respectively. The Mann-Whitney-Wilcoxon rank-sum test revealed that D2 distances between cognates were not significantly larger than the D1 distances of their contextual targets (D2 <= D1), and O2 overlaps were not significantly smaller than the O1 overlaps of their contextual targets (O2 >= O1). This was most pronounced for the /k-g/ pair, for which D2 was notably smaller than D1, and O2 was larger than O1. Based on the results of the statistical analysis, we concluded that cognates have identical positional targets, and were regarded as a single unit in future comparisons.

Insert Figure 2 about here

B. Analysis of Homorganic Consonants

Figure 3 shows representative examples of the target locations for the alveolar and postalveolar consonants for four talkers. Note that velar stops were excluded from any further analyses as the locations of their targets differed from other consonants due to the differences in T1 and T2 sensor placement. Talkers W12 and W24 (top two plots) demonstrated distinctively different target locations between alveolar pairs (i.e., d-z, d-s, t-s, t-z), when talkers W25 and W15 (bottom two plots) showed great similarity in these targets. The postalveolar homorganic pair (/ʧ/-/ʃ/) showed similar target locations in these talkers.

Insert Figure 3 about here

The summaries of statistical results computed across all talkers are shown in Table 3a and 3b. Statistical comparisons of the complete data set revealed significant differences between D1 and D2 for the alveolar pairs, but not for the postalveolar pair. Their overlaps did not differ from the overlaps observed across their respective contextual pairs. We thus concluded that the consonants /d/ and /t/ had distinct locations from /s/ and /z/ and could not be collapsed for further analyses. On the other hand, /ʃ/ and /ʧ/ were not distinct and were regarded as a single unit in the subsequent analyses.

Since the distances and overlaps between /d/ and /t/ on the one hand, and /s/ and /z/ on the other, varied greatly from individual to individual (see plots in Figure 3 and SDs in Table 3a), we next investigated whether some covariate(s) of the individual subjects might partially explain these effects. We found that speaking rate was the only covariate with statistically significant correlation with D2 computed for alveolar pairs (r= 0.498, p-value 0.029). The data showed that talkers with slower habitual speaking rate produced larger distances between the consonant target regions for the alveolar pairs. Speaking rate (r= -0.508, p-value 0.026) as well as the palate curvature (r= -0.562, p-value 0.012) were associated with across-talker variation in O2. Talkers who spoke slowly and had flatter palates showed less overlap between consonant targets. The measure of palate width was "nearly" significantly correlated with O2, with estimated correlation -0.455 and corresponding p-value 0.051 (see Figure 4, column of plots on the left). Regressing the average overlap against the two covariates gives the formula:

$$O2 = 1.28 - 0.0017*R - 0.339* \alpha \qquad (1)$$

with R^2 equal to 0.4562, and with curvature significant but rate not significant at the 0.05 level (p-values 0.028 and 0.059, respectively).

Insert Figure 4 about here

C. Analysis of alveolars and postalveolars

Figure 3 shows the most typical patterns of the alveolar versus postalveolar target location across talkers. W12 and W25 demonstrated distinctive locations of the postalveolar targets from the alveolar stops and fricatives alike (see two plots on the left). This pattern was typical for 11 talkers in our sample. W24 showed overlapping locations of the postalveolars and the alveolar stops but not alveolar fricatives; a total of four talkers showed this pattern. Talkers W15 demonstrated an overlap between all alveolars and postalveolars; this pattern was exhibited by four different talkers in our sample.

Tables 4a and 4b summarize the statistical results of D2-D1 and O2-O1 comparisons for the alveolar – postalveolar pairs, respectively, where alveolar stops and fricatives were treated separately. The Mann-Whitney-Wilcoxon rank-sum test for both D1-D2 and O1-O2 comparisons revealed significant differences between consonant targets for alveolar stops and postalveolar consonants and alveolar fricatives and postalveolar consonants.

The between-talker variability was examined graphically and through correlations for these pairs of consonants as well (see Figure 4, the column of plots on the right). The results revealed that for D2, the only significant correlation was in the alveolar fricatives and postalveolar comparison with palate width (r=-0.549, p-value=0.006). Wider palates were associated with greater distances between these consonants. Variation in O2 in the alveolar fricative versus postalveolar consonant pairs was associated with palate width (r=-0.606, p-value=0.006) and palate curvature (r=-0.518, p-value=0.0232). In the alveolar stops versus postalveolar consonant comparisons, the variation in O2 was explained by palate width (r=-0.528, p-value=0.0202) and sex (r=0.47, p-value=0.041). Males who also have significantly wider palates showed smaller overlaps between consonant targets.

IV. DISCUSSION

The main goal of this study was to compare the positional target regions for different lingual consonants as represented by a point-parameterized motion tracking technique and to identify which lingual consonants occupy distinctively different target locations in the vocal tract with respect to other consonants. The notion of articulatory targets that were labelled "target regions" representing repetition variability in tongue positions across contexts was evoked (see Guenther, 1995; Kent & Moll, 1972). The proximity of different consonant targets was assessed relative to the proximity of their contextual targets. The rationale for this type of comparison was based on the idea that there would be no predetermined differences in distances between consonant targets across talkers. More specifically, the consonant regions may be 2 mm apart for one talker and 20 mm apart for another talker. Consonants would be differentiated into separate categories in both of these talkers as long as their target regions were further from each other than their contextual targets and if their overlaps were less than those of their contextual targets.

We found that cognate pairs and homorganic postalveolars shared the locations of their positional targets across a group of talkers. A number of consonants showed unique positions of their target regions. On average, alveolar stops occupied locations that were distinctively different from those occupied by alveolar fricatives. Additionally, both alveolar stops and fricatives showed positional targets that were different from the postalveolar fricative and the affricate. Substantial across talker variability in the location of the positional targets was observed in the pairs that showed differences in target locations. As a preliminary finding, these differences were attributable to speaking rate and palate characteristics. We discuss these findings in more detail below.

A. Cognates

Cognates, by their linguistic definition, share the place (and manner) of their articulation. Voiced and voiceless cognates, as would have been predicted from their linguistic description, showed indistinguishable target regions in our study. This finding agreed with a number of existing studies using EPG and MRI (Dagenais & Critz-Crosby, 1991; Fletcher, 1989; Gibbon et al., 2007; Narayanan, Alwan, & Haker, 1995). The overall similarity in place of articulation of the cognates was in contrast with that reported in Mooshammer et al. (2007), who, similarly to us, used electromagnetic articulography and reported tongue tip sensor positions in /d/ and /t/ embedded into an aCa syllable. Their data showed that in their sample of five talkers the voiceless alveolar stop was significantly more fronted than the voiced stop. When our data for the same syllable were examined, the discrepancy between our two studies remained, with the majority of talkers showing overlapping targets for the cognates, and a few talkers that showed one or another of these targets in the forward position. In addition to differences in sample size, linguistic differences between English and German may explain the observed discrepancy between the two studies.

B. Alveolar stops versus alveolar fricatives

Electropalatographic studies showed characteristically different tongue-palate contact patterns between alveolar stops and fricatives (see Goozée, Murdoch, & Theodoros, 2003; Fletcher, 1989; McAuliffe, Ward, & Murdoch, 2002). In this study, which employed an articulographic method, their positional targets revealed unique locations as well. On average, our data showed that the place of articulation for fricatives was somewhat more anterior and lower than that of stops. This might be due to tongue surface adjustments that create tongue grooving, an essential element of fricative articulation (Fletcher, 1989; Narayanan et al., 1995). On average, /s/ had a lower tongue position as compared to /t/ in Mooshammer et al as well (see their Figure 3). Note, however, that in our sample this was true for some talkers but not others. This variability may be due to the between-talker differences in fricative articulation. It is known that some talkers use an apical articulation while shaping alveolar fricatives, when others use the laminal articulation. Narayanan et al. (1995) observed that apically-produced fricatives showed deeper grooving in the anterior tongue body behind the constriction than the laminally-produced fricatives. Potentially, those talkers that use the apical articulation (i.e., more grooving and front and low target locations) might be the ones that showed the most contrast in target locations between the alveolar stops and fricatives (see W12 and W24 in Figure 3). It might be beneficial to determine the individual tendency for this type of patterning in advance in future studies.

The point of attachment of the T1 sensor may be an additional source of variability in these data. Although, it is typical to place the T1 sensor at approximately 1 cm from the tongue tip and use it to represent alveolar stops and fricatives (e.g., see Iskarous, Shadle, & Proctor, 2011; Mooshammer et al., 2007), this approach disregards the various tongue shaping strategies as well as potential between-subject differences in the location of the consonant constrictions. The role of sensor placement on target region locations and overlaps has to be examined in detail in the future.

There was a great degree of variability in the locations of target regions for the alveolar consonants across talkers. A number of potential sources of variability were identified and included difference in speaking rate, sex, and palate characteristics. Speaking rate was related to between talker variability in the location of the alveolar fricatives and stops. Individuals who adopted a slower (perhaps clear, citation form) speech style of pronunciation during the experiment showed more differentiated locations of the target regions for /s, z/ versus /t, d/. This is not surprising based on what we know about the effect of speaking rate/style on articulation (Lindblom, 1990). In further studies, the tongue's positional variability should be tested by experimentally varying speaking rate and style.

C. Postalveolars and their comparison to the alveolars

In a traditional linguistic description, the two postalveolar consonants studied are specified as homorganic. In our study the location of the target region of /ʧ/ was not different from that of /ʃ/ across talkers, supporting this description. Similarities in tongue-palate contact patterns between these two consonants have been described in the past using EPG. Fletcher (1989) reported that both the stop and fricative portions of the affricate are produced at the location coinciding with that of the postalveolar fricative. Liker and colleagues reported a single relatively posterior location of the tongue for the affricate in their electropalatographic analysis (Liker, Gibbon, Wrench, & Horga, 2007). These authors also reported that the occlusion phase during the affricate was relatively stable, without a notable transition between the stop portion of the sound and its fricative portion. As pointed by Fletcher (1989), articulatory data generally supports the notion of affricates as unified monophoneme as supposed to a combination of two independent phonemes joined through a transition. Overall, articulatory movement data for affricates are extremely limited but very much needed in order to better describe the time-varying nature of articulation during these sounds.

Since alveolar stops and fricatives showed different target locations, they were compared individually to the postalveolar consonants; postalveolar data were combined, however. The tongue locations for alveolar fricatives and postalveolars as well as alveolar stops and postalveolars were, on average, different in our study. Postalveolars, including /ʃ/, were located on average 4 mm behind /s-z/. This observation agrees with previously reported electropalatographic data (Fletcher, 1989; Hoole, Ziegler, Hartmann, & Hardcastle, 1989) and

MRI data (Narayanan et al., 1995). Narayanan and co-workers reported that the primary constriction for the postalveolar fricative was 5 to 10 mm behind that of the alveolar fricative. The difference between our findings is most likely due to the difference in the two methods. MRI in contrast to articulography has a great advantage in identifying the constriction location and shaping patterns of the tongue regions as well as the tongue as a whole. Our data compares closely to those reported by Mooshammer et al. (2007) for a different articulography system, however, which showed that tongue sensor was more retracted and higher for /ʃ/ than for /s/ in three out of their five talkers.

Talkers differed in the distances and overlaps of their target regions. The between-talker variability in the production of alveolar and postalveolar fricatives and the affricate is not surprising (see Shadle, Proctor, & Iskarous, 2008). These between-talker differences were explained to some extent by associations with palate width and curvature.

D. The effect of palate shape on the location of the consonant target regions

In our previous study based on the same data set, we evaluated the effect of palate characteristics (width, length, height, slope and curvature) on consonant token-to-token variability or, in other words, the size of the target regions (Rudy & Yunusova, 2010). We found that individuals with relatively coronally flat palates showed smaller target regions during consonants as compared to the talkers with coronally domed palates. Similar results were reported for vowels (Brunner, et al., 2005 & 2009; Mooshammer et al., 2004). In this study, we examined the effect of the palate on the locations of consonant targets relative to each other. Here we also found that the palate characteristics mattered. Individuals with flatter palates tended to produce certain consonants more distinctively in terms of the place of articulation (i.e., with less overlap between their target regions) than talkers with relatively domed palates. A recent study by Weirich & Fuchs (2011) showed similar relationship between the palate curvature and articulatory distance between tongue positions in /s/-/ʃ/ contrast.

One of the potential explanations offered in the previous studies of the association between positional variability and palate shape was due to the amount of tongue-palate contact (see Brunner et al., 2005). Presumably, flatter palates provide a larger surface for such contact to occur as compared to the domed palates. Brunner et al. (2005) discovered, however, that although talkers with flatter palates produced less variability in tongue positions during vowels than talkers with domed palates, flatter palates did not show more tongue-palate contact than the domed palates. Since this report, the explanatory focus of the work on the articulatory variability and palate shape has been on the control of the acoustic variability and its relation to palate morphology (see Brunner et al., 2009). However, the consonant data reported in this study might suggest that the question of the amount of the palate contact during various consonants and palate shape needs to be revisited. An EPG technology paired with a method of estimating the extent of palate contact at tongue's maximum elevation is needed to elucidate this relationship for a relatively large number of talkers with various palate morphologies.

Palate width emerged as a fairly consistent correlate of the distance and overlap between alveolar stops and postalveolars, and alveolar fricatives and postalveolars. Wider palates showed more distinction in consonant place articulation than the narrower palates. This finding contradicts the report by Weirich & Fuchs (2011), who showed that the anterior width of the palate is negatively correlated with the distance between places of articulation for /s/ and/ʃ/. We measured palate relatively posteriorly, between the first permanent molars, and this difference might be at the root of the discrepancy between our findings. Overall, palate shape remains important to examine with respect to identification of the positional targets for consonants across talkers.

E. Implications

The data reported in this study have implications for a number of theoretical questions. For example, the tongue position variability is examined between different consonants within the DAC model of coarticulation (Recasens et al., 1997). In fact Recasens and Espinosa (2009) used the measure of distance between the center of the target region for a consonant and its contextual targets (the same as D1 in this study) to represent the coarticulatory resistance of consonants in Catalan. Only three consonants were similar between the two studies (/s, ʃ, k/), for which Recasens and Espinosa collected positions of three tongue sensors. Recasens and Espinosa compared this measure between consonants, finding significant differences and supporting the notion of articulatory resistance. Our D1 served as a statistical reference to the D2 measure that quantified the distance between consonant targets. We did not compare the consonants with respect to their D1 differences.

In our previous study (Rudy & Yunusova, 2010), we calculated the size of the target regions, a measure that represented positional tongue variability during the same consonants reported here, and found no difference in variability estimates between the fricatives /s/ and /ʃ/ in our sample. Note, however, that Recasens and Espinosa reported more notable between-sound differences in the tongue blade and dorsum sensors for these sounds; their tongue tip sensor differences were within a margin of error of the articulograph. The notable variability between talkers that is clearly seen our study emphasizes, however, the need to record larger samples of talkers, to test the DAC model (e.g., only three talkers provided data in Recasens & Espinosa, 2009).

The main goal of this study was to identify which consonants had distinctively different target regions and which did not. This study clearly found that tongue positions (at least as measured by a single sensor) are not completely unique within a talker, and alone they would be unable to differentiate between phonemes. A number of talkers do not seem to distinguish the tongue sensor locations between consonants (for example see talker W15 in Figure 3). Yet, certain talkers and certain sounds show unique positional characteristics (talker W12 in Figure 3). These data emphasize the fact that consonant are multidimensional in nature. A single characteristic (e.g., tongue position) is not sufficient to identify a consonant (Schwartz, Boë, & Abry, 2007). The data also suggest that talkers might be on a continuum of using certain articulatory characteristics to distinguish consonants. Some may relay on tongue position information, when others (e.g., those with particular palate shapes) may have to relay more heavily on acoustic and aerodynamic factors.

Practical implications of these data have to do with the development of technologies for speech recognition and visual augmented articulatory feedback. The tongue trajectories containing location information are beginning to be successfully used in speech recognition algorithms based on kinematic signals, which aim to enhance existing acoustic-based recognition models (Wang et al., 2010). Additionally, positional target similar to ones shown here have been used to treat place of articulation in motor speech disorders (see Katz, Bharadwji, & Carstens, 1999; McNeil et al., 2010). Our results suggest that certain sounds are characterised by positional distinctiveness, with can be used effectively for enhancement of speech recognition and feedback-based treatments. Yet, there is a significant interaction between sounds class/ type and individual talker characteristics, which needs to be consider in such applications.

F. *Conclusions*

The findings of this study largely confirmed general expectation for consonant tongue positions observed with point-parameterized method. Yet they also emphasized the extent of the variability between talkers in consonant target locations, which was missing from the existing studies with the small number of talkers. These results allow us to empirically establish expectation as to where consonants are formed in relation to each other when tracked with articulography. For example, on average talkers show distinctive locations of the tongue front sensor in /s/ (anterior, low), /t/ (posterior to /s/ and higher along the surface of the palate), and /ʃ/ (further back from /t/ but still along the palate surface). Cognates and the postalveolar pair (/ʧ/-/ʃ/) show similar target locations. Yet, there is notable variability among talkers in the distinctiveness of consonant target locations among alveolars and postalveolars. Importantly, this study identified that individuals' palate characteristics and their habitual speaking rate are the necessary variables to track in order to account for such variation. Since our findings pertain only to the differences in the habitual speaking rate, however, future work should look into the effect of speaking rate/style variation on target region locations within a speaker to identify a mode in which target region specification can be better localized.

APPENDIX 1

To calculate the overlap for each of the two sounds, first we used the available data to compute a *kernel density estimator*, i.e. an estimate of the multidimensional probability density for the tongue positions of that individual when producing that consonant. We computed kernel density estimators using Gaussian kernel functions together with Scott's Rule (Scott, 1992, p. 152), as follows. If the observed data were $w_1, w_2, \ldots, w_n$, where each w_i had three spatial components $w_i = (w_{i,1}, w_{i,2}, w_{i,3})$, then for any vector $v = (v_1, v_2, v_3)$, we estimated the density function f at v by:

$$f(v) = \frac{1}{n}\sum_{i=1}^{n}\Phi(v; w_i, h)$$

where

$$\Phi(v; w_i, h) = \prod_{j=1}^{3}\frac{1}{\sqrt{2\pi h_j}}\exp[-(v_j - w_{i,j})^2 / 2h_j^2]$$

was the density of a three-dimensional Gaussian distribution having mean w_i and standard deviation $h = \mathrm{diag}(h_1, h_2, h_3)$.

where $h_j = n^{-1/7}\mathrm{sd}(w_{1,j}, w_{2,j}, \ldots, w_{n,j})$

Equivalently,

$$f(v) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{3}\phi(v_j; w_{i,j}, h_j)$$

where

$$\phi(v_j; w_{i,j}, h_j) = \frac{1}{\sqrt{2\pi h_j}}\exp[-(v_j - w_{i,j})^2 / 2h_j^2]$$

is the density of a one-dimensional Gaussian distribution having mean w_i and standard deviation h_j.

Then, once we have estimates, say f and g, of the density functions corresponding to the two different point clouds, we then compute their *overlap* by:

$$overlap = \int \min[f(v), g(v)]dv$$

where the integral is over all points $v = (v_1, v_2, v_3)$ in three-dimensional space. That is, the overlap equals the amount of area which is under *both* of the density functions, i.e. which is *common* to both probability distributions.

ACKNOWLEDGEMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), CFI-LOF, and the American Speech Language Hearing Association (ASHA) Award to New Investigators awarded to Y. Yunusova. Dr. Rosenthal's work on this project was partially supported by the NSERC fund.

FOOTNOTE

[1] Note that the notion of "target" is used in this manuscript without reference to the control mechanism for speech (cf., Guenther, Hampson, & Johnson, 1998). In this work, we use the term "target" simply to describe the locations of the tongue represented by a sensor during tongue's maximum elevation. These targets are defined using point-clouds that are formed through repetitive productions of the sounds.

REFERENCES

Bailly, G. (1997). Learning to speak. Sensori-motor control of speech movements. *Speech Commun., 22,* 251-267.

Berry, J. (2011). Accuracy of the NDI Wave speech research system. *J. Speech. Lang. Hear. Res., 54*, 1295-1301.

Brunner, J., Fuchs, S., & Perrier, P. (2005). The influence of the palate shape on articulatory token-to-token variability. *ZAS Pap. Ling., 42*, 43-67.

Brunner, J., Fuchs, S., & Perrier, P. (2009). On the relationship between palate shape and articulatory behaviour. *J. Acoust. Soc. Am., 125*, 3936-3949.

Dagenais, P. A., & Critz-Crosby, P. (1991). Consonant lingual-palatal contacts produced by normal-hearing and hearing-impaired children. *J. Speech Lang. Hear. Res., 34*, 1423-1435.

Dart, S. N. (1998). Comparing French and English coronal consonant articulation. *J. Phon., 26*, 71–94.

Dembowski, J., Lindstrom, M. J., & Westbury, J. R. (1998). "Articulator point variability in the production of stop consonants," in *Neuromotor Speech Disorders: Nature, Assessment, and Management*, edited by M. P. Cannito, K. M. Yorkston, and D. R. Beukelman (Paul H. Brookes Publishing, Baltimore), pp. 27–46.

Farnetani, E. & Recasens, D. (2010). Coarticulation and connected speech processes, (eds.) Hardcastle, W.J., Laver, J. & Gibbon, F.E. The Handbook of Phonetic Sciences (Second Edition).

Ferrario, V. F., Sforza, C., Schmitz, J. H., & Colombo, A. (1998). Quantitative descriptions of the morphology of the human palate by a mathematical equation. *Cleft Palate-Cran. J., 35*, 396-401.

Fletcher, S. (1989). Palatometric specification of stop, affricate, and sibilant sounds. *J. Speech Lang. Hear. Res., 32*, 736-748.

Gibbon, F. E., Yuen, I., Lee, A., & Adams, L. (2007). Normal adult speakers' tongue palate contact patterns for alveolar oral and nasal stops. *Int. J. Speech Lang. Pathol., 9*, 82-89.

Goozée, J. V., Murdoch, B. E., & Theodoros, D. G. (2003). Tongue-to-palate contacts exhibited in dysarthria following traumatic brain injury: spatial characteristics. *J. Med. Speech Lang. Pathol., 11*, 115-129.

Guenther, F. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychol. Rev., 102*, 594-621.

Guenther. F. H., Gosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang., 96*, 280-301.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychol. Rev., 105*, 611-633.

Guenther. F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *J. Acoust. Soc. Am, 105*, 2854-2865.

Hasegawa-Johnson, M., Pizza, S., Alwan, A., Cha, J. S., & Haker, K. (2003). Vowel category dependence of the relationship between palate height, tongue height, and oral area. *J. Speech Lang. Hear. Res., 46*, 738-753.

Hoole, P., Ziegler, W., Hartmann, E., & Hardcastle, W. (1989). Parallel electropalatographic and acoustic measures of fricatives. *Clin. Linguist. Phon., 3*, 59-69.

Houde, R. A. (1967). A study of tongue body motion during selected speech sounds. Ph. D. thesis, University of Michigan.

Iskarous, K., Shadle, C., & Proctor, M., I. (2011). Articulatory-acoustic kinematics: The production of American English /s/. *J. Acoust. Soc. Am., 129*, 944-954.

Katz, W. F., Bharadwaj, S. V., & Carstens, B. (1999). Electromagnetic Articulography Treatment for an adult with Broca's aphasia and apraxia of speech. *J. Speech Lang. Hear. Res., 42*, 1355-1366.

Keating, P. A. (1990). "The window model of coarticulation: articulatory evidence," in *Papers in Laboratory Phonology I* , edited by J. Kingston & M. Beckman (Cambridge University Press, Cambridge, UK), pp. 451-470.

Kent, R. D., & Moll, K. L. (1972). Cinefluorographic analyses of selected lingual consonants. *J. Speech Lang. Hear. Res., 15*, 453-473.

Ladefoged, P., & Maddieson, I. (2008). *The sounds of the world's languages* (Blackwell Publishing: Oxford), pp. 1-427.

Liker, M., Gibbon, F. E., Wrench, A., & Horga, D. (2007). Articulatory characteristics of the occlusion phase of /tʃ/ compared to /t/ in adult speech. *Advan Speech-Lang Path, 9*, 101-108.

Liker, M., & Gibbon, F. E. (2008). Tongue palate contact patterns of velar stops in normal adult english speakers. *Clin. Linguist. Phon., 22*, 137-148.

Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H & H theory," in *Speech Production and Speech Modeling*, edited by W. Hardcastle & A. Marchal (Kluwer Academic Publishers, Dordrecht), pp. 403-439.

Mann, H. B., & Whitney, (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, *18*, 50-60.

McAuliffe, M. J. Ward, E. C., & Murdoch, B. E. (2002). Tongue-to-palate contact patterns and variability of four English consonants in an /i/ vowel environment. *Logoped. Phoniatr. Vocol., 26*, 165-178.

McNeil, M., Katz, W., Fossett, T., Garst, D., Szuminsky, N., Carter, G., & Lim, K. (2010). Effects of on-line augmented kinematic and perceptual feedback on treatment of speech movement in apraxia of speech*. Folia Phoniatr. Logop., 62,* 127-133.

Mooshammer, C., Hoole, P., & Geumann, A. (2007). Jaw and Order. *Lang Speech, 50*, 145-176.

Mooshammer, C., Perrier, P., Fuchs, S., Geng, C., & Pape, D. (2004). An EMMA and EPG study on token-to-token variability. *AIPUK*, 36, 47-63.

Narayanan, S. S., Alwan, A. A., & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *J. Acoust. Soc. Am., 93*, 1325-1335.

Öhman, S.E.G. (1967). Numerical model of coarticulation. *J. Acoust. Soc. Am., 41*, 310-320.

Perkell, J. S., & Klatt, D. H. (1986). Invariance and variability in speech processes (Lawrence Erlbaum Associates, Hillsdale, NJ, England), pp. 1-604.

Perkell, J. S., (1996). Properties of the tongue help to define vowel categories: Hypotheses based on physiologically oriented modeling. *J. Phon., 24*, 3-22.

Perkell, J. S., & Nelson, W. L. (1985). Variability in production of the vowels /i/ and /a/. *J. Acoust. Soc. Am., 77*, 1889-1895.

Perkell, J. S., & Nelson, W. L. (1982). "Articulatory targets and speech motor control: A study of vowel production in Speech Motor Control," in *Speech Motor Control*, edited by S. Grillner, A Persson, B. Lindblom, and J. Lubker (Pergamon Press: New York), pp. 187-204.

Perrier, P., Boe, L. J., & Sock, R. (1992). Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: Modeling the transitions with two sets of coefficients. *J. Speech Lang. Hear. Res., 35*, 53-67.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan, *J. Acoust. Soc. Am.*, *125*, 2288-2298.

Recasens, D., Pallares, M., and Fontdevila, J. (1997). "A model of lingual coarticulation based on articulatory constraints," J. Acoust. Soc. Am.,, 102, 544–561.

Roberts, G. O., & Rosenthal, J.S. (2004), General state space Markov chains and MCMC algorithms. Probability Surveys, 1, pp. 20-71.

Rudy, K., & Yunusova, Y. (2010, November). The effect of palatal geometry on consonant articulation. Poster session presented at the ASHA Convention, Philadelphia, PA.

Schwartz, J. L., Boë, L. J., & Abry, C. (2007). Linking the dispersion-focalization theory (DFT) and the maximum utilization of the available distinctive features (MUAF) principle in a perception-for-action-control theory (PACT). In M. J. Sole´, P. Beddor, & M. Ohala (Eds.), Experimental approaches to phonology (pp. 104–124). Oxford University Press.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization* (John Wiley & Sons, New York), pp. 1-317.

Shadle, C., Proctor, M. I., & Iskarous, K. (2008). An MRI study of the effect of vowel context on English fricatives. *Acoustics '08 Paris: Joint meeting of the ASA & Acoustics '08 Paris: Joint meeting of the ASA, EAA & Société Française d'Acoustique, Paris, 29 June - 4 July 2008.*

Tasko, S., & McClean, M. D. (2004). Variations in articulatory movement with changes in speech task. *J. Speech Lang. Hear. Res., 47*, 85-100.

Tsao, Y. C., & Weismer, G. (1997). Inter-speaker variation in habitual speaking rate: Evidence for a neuromuscular component. *J. Speech Lang. Hear. Res., 40*, 858-866.

Wang, J., Green, J. R., Samal, A., & Carrell, T. D. (2010). Vowel recognition from continuous articulatory movements for speaker-dependent applications. IEEE Intl. Conf. on Signal Processing and Communication Systems, 1-7.

Weirich, M., & Fuchs, S. (2011). Vocal tract morphology can influence speaker specific realisations of phonemic contrasts. *Proceedings of the ISSP*, 251-258.

Westbury, J. (1994). On coordinate systems and the representation of articulatory movements. *J. Acoust. Soc. Am.*, *95*, 2271-2273.

Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biom, 1*, 80-83.

Yunusova, Y., Baljko, M., Pintilie, G., Rudy, K., Faloutsos, P., & Daskalogiannakis, J. (2012). Acquisition of the 3D surface of the palate by in-vivo digitization. *Speech Commun*. 54, 923-931.

TABLE I. Summary statistics (means and standard deviations) of measures describing palate size and shape by sex.

Group	Palate Curvature (α)	Palate Slope	Palate Length (mm)	Palate Width (mm)	Palate Height (mm)
Males	1.86 (0.19)	0.33 (0.21)	35.21 (2.82)	34.99 (2.35)	13.98 (2.08)
Females	1.85 (0.21)	0.39 (0.28)	30.79 (4.55)	31.99 (3.87)	11.45 (1.68)

TABLE IIa. Summary statistics (means and standard deviations) for the two distance measures (mm) computed for the each cognate pair; N1 and N2 indicate the number of comparisons. The results of the Mann-Whitney-Wilcoxon rank-sum test are also shown.

Pair	D1	D2	N1	N2	U statistic	*p*-value
/d/-/t/	1.52 (0.80)	1.37 (0.77)	114	19	958	0.790
/z/ - /s/	1.36 (0.90)	1.46 (0.86)	114	19	1166	0.300
/g/ - /k/	2.14 (1.03)	1.23 (0.50)	102	17	318	0.999

TABLE IIb. Summary statistics (means and standard deviations) for the two overlap measures computed for the each cognate pair; N1 and N2 indicate the number of comparisons. The results of the Mann-Whitney-Wilcoxon rank-sum test are also shown.

Pair	O1	O2	N1	N2	U statistic	*p*-value
/d/-/t/	0.39 (0.13)	0.35 (0.17)	114	19	878	0.094
/z/ - /s/	0.37 (0.16)	0.31 (0.19)	114	19	848	0.066
/g/ - /k/	0.34 (0.13)	0.47 (0.13)	102	17	1306	0.999

TABLE IIIa. Summary statistics (means and standard deviations) for the two distance measures (mm) computed for the each homorganic pair (alveolar pairs are collapsed); N1 and N2 indicate the number of comparisons. The results of the Mann-Whitney-Wilcoxon rank-sum test are also shown.

Pairs	D1	D2	N1	N2	U statistic	*p*-value
d-z, d-s, t-s, t-z	1.44 (0.85)	2.72 (2.09)	228	76	12646	0.001*
ʃ-tʃ	1.27 (0.74)	1.47 (0.99)	114	19	1174	0.280

TABLE IIIb. Summary statistics (means and standard deviations) for the two overlap measures computed for the each homorganic pair (alveolar pairs are collapsed); N1 and N2 indicate the number of comparisons. The results of the Mann-Whitney-Wilcoxon rank-sum test are also shown.

Pairs	O1	O2	N1	N2	U statistic	*p*-value
d-z, d-s, t-s, t-z	0.38 (0.15)	0.23 (0.19)	228	76	4618	0.001*
ʃ-tʃ	0.43 (0.15)	0.42 (0.18)	114	19	1113	0.577

TABLE IVa. Summary statistics (means and standard deviations) for the two distance measures (mm) computed for the postalveolar-alveolar stops and postalveolar-alveolar fricatives (collapsed); N1 and N2 indicate the number of comparisons. The results of the Mann-Whitney-Wilcoxon rank-sum test are also shown.

Pairs	D1	D2	N1	N2	U statistic	*p*-value
d-ʃ , t-ʃ, d-ʧ , t-ʧ	1.39 (0.78)	2.93 (1.4)	228	76	14446	0.001*
s-ʃ, z-ʃ, s-ʧ , z-ʧ	1.31 (0.82)	4.32 (1.75)	228	76	16606	0.001*

TABLE IVb. Summary statistics (means and standard deviations) for the two overlap measures computed for the postalveolar-alveolar stops and postalveolar-alveolar fricatives (collapsed); N1 and N2 indicate the number of comparisons. The results of the Mann-Whitney-Wilcoxon rank-sum test are also shown.

Pairs	O1	O2	N1	N2	U statistic	*p*-value
d-ʃ , t-ʃ, d-ʧ , t-ʧ	0.41 (0.14)	0.19 (0.15)	228	76	2508	0.001*
s-ʃ, z-ʃ, s-ʧ , z-ʧ	0.40 (0.15)	0.08 (0.09)	228	76	567	0.001*

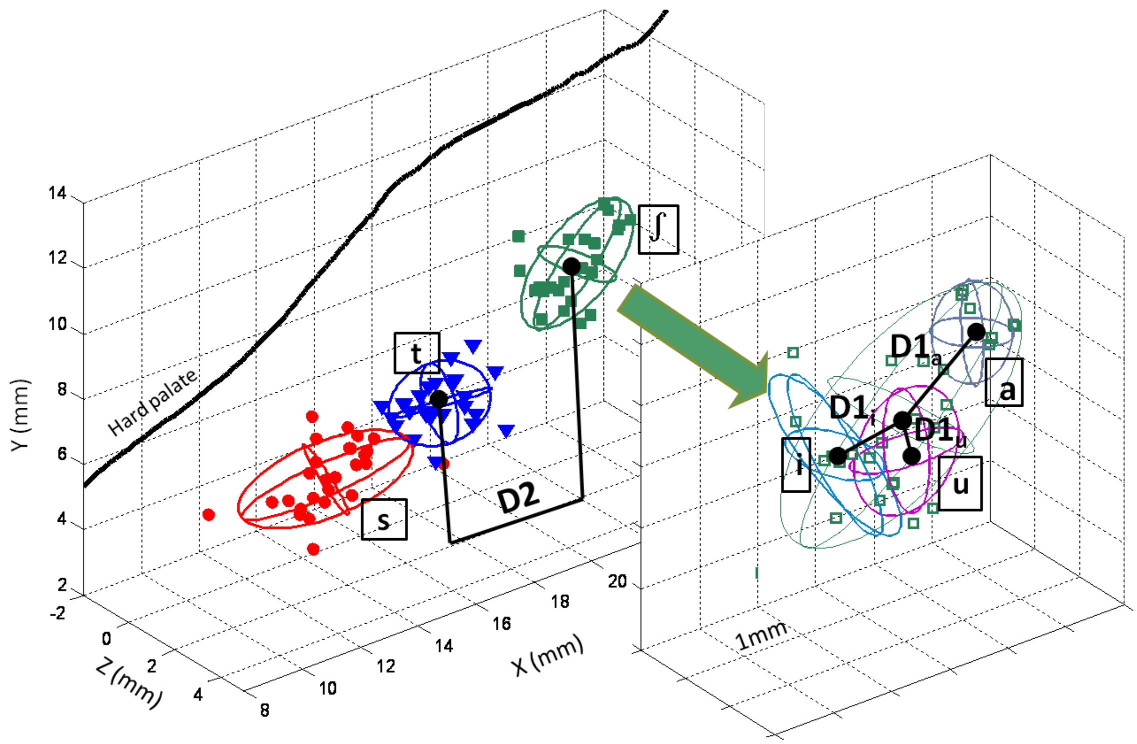
FIGURE CAPTIONS

FIG. 1. Point clouds representing positions of the tongue front sensor for /s/ (circles), /t/ (triangles), and /ʃ/ (squares) produced by a single speaker (W12). Two standard deviation (SD) ellipses are fit around the point clouds. Mid-sagittal trace of the hard palate is also plotted. The insert on the right shows 2SD ellipses fit around/ʃ/ in /i/, /u/ and /a/ contexts. The overall, across-contexts 2SD ellipse for the consonant is also plotted.

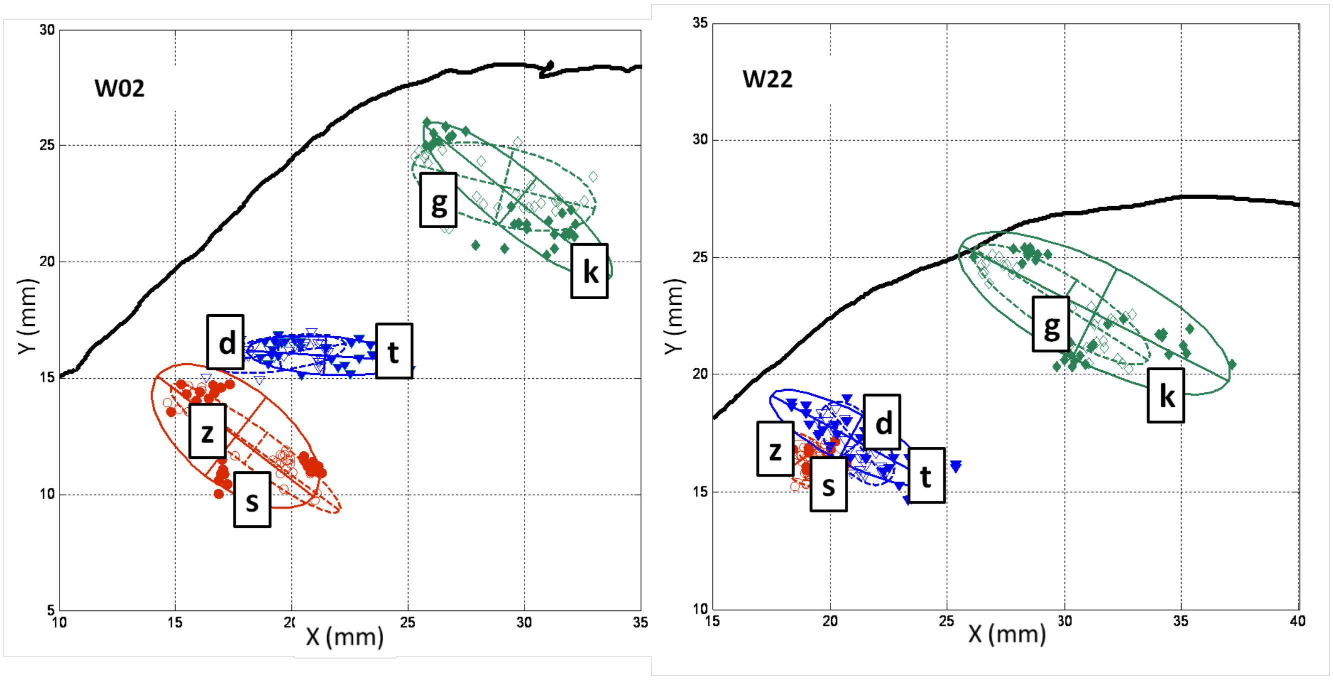
FIG. 2. Cognate pairs are plotted in 2D for two talkers (W02 and W22) midsagittaly. Fricatives are shown in circles (closed = s, open = z); alveolar stops are shown in triangles (closed = t, open = d); and velar stops are shown in rhombuses (closed = k, open = g). The boundaries of the voiceless targets are identified with a solid line; the voiced targets are marked with a dashed line.

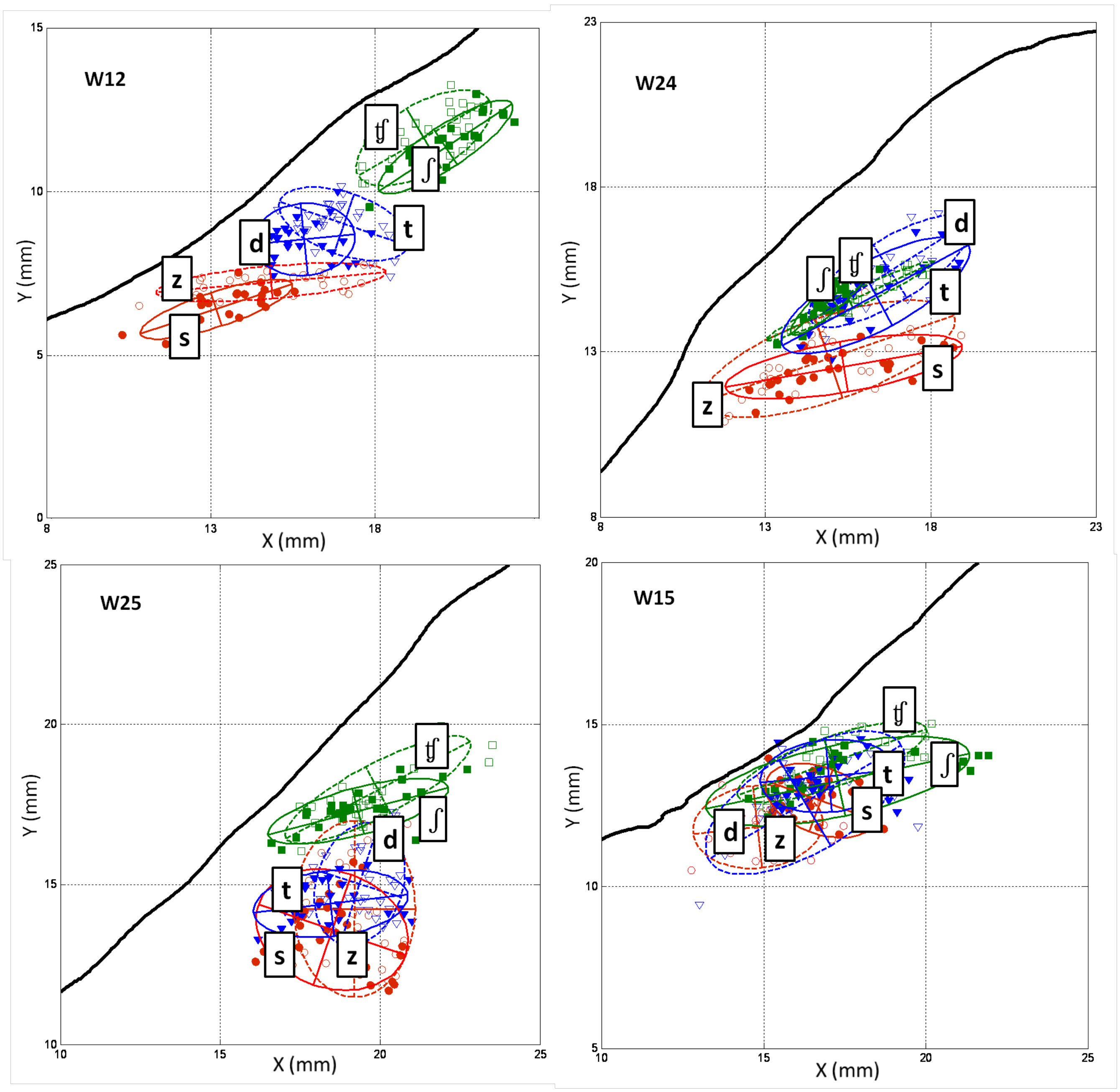
FIG. 3. Six consonants are plotted in 2D for four talkers (W12, W25, W24 and W15). Alveolar fricatives are shown in circles (closed = s, open = z); alveolar stops are shown in triangles (closed = t, open = d); and postalveolar consonants are shown in squares (closed = ʃ, open = ʧ). The boundaries of the voiceless targets are identified with a solid line; the voiced targets are marked with a dashed line.

FIG 4. Scatterplots illustrate the relationships between D2 and O2 measures and their significant correlates.

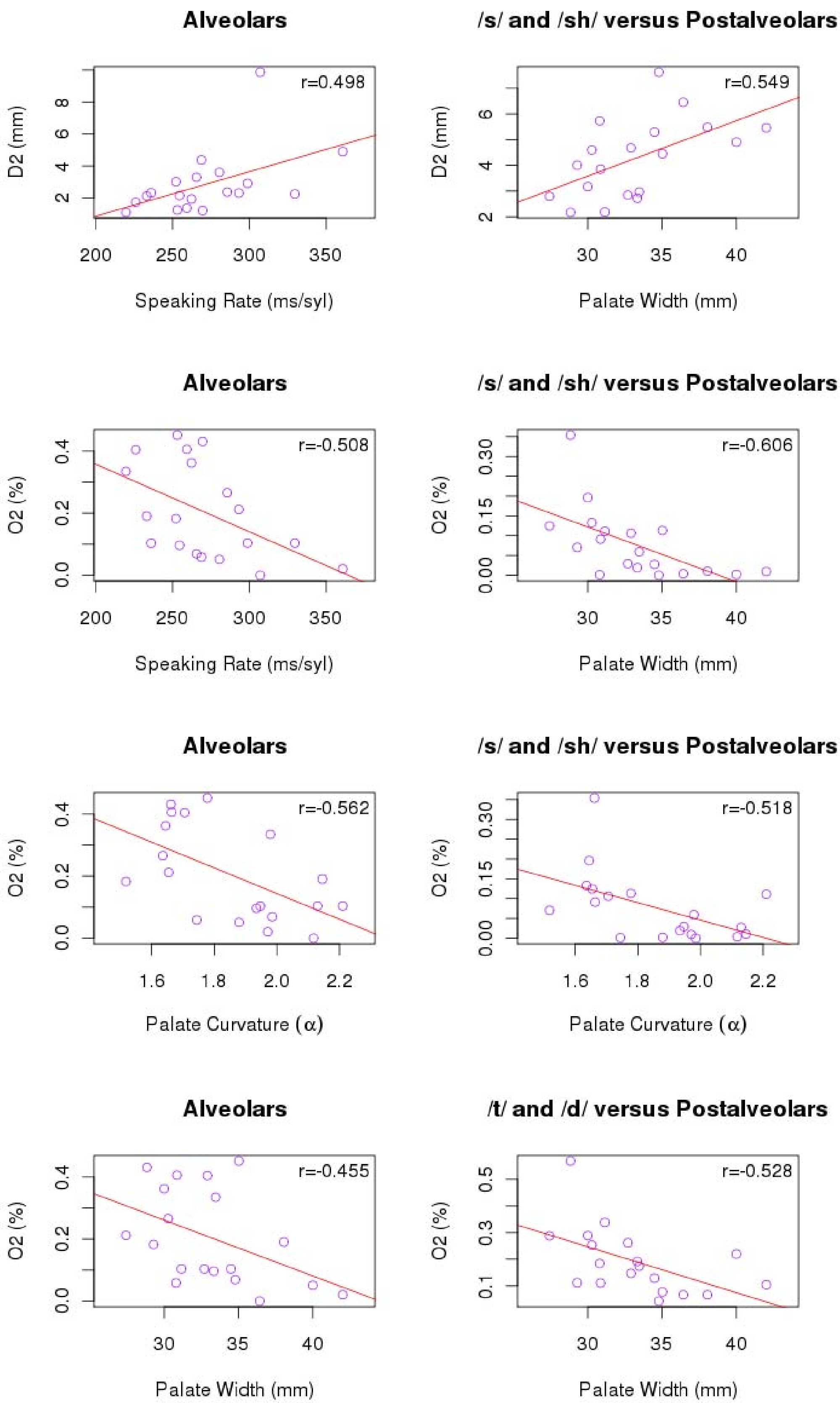


Hard palate
Y (mm)
14
12
10
8
6
4
2
-2
0
2
4
Z (mm)
8
10
12
14
16
18
20
X (mm)
s
t
ʃ
D2
D1a
D1i
D1u
a
i
u
1mm


W02
Y (mm)
X (mm)
30
25
20
15
10
5
10
15
20
25
30
35
d
t
z
s
g
k
W22
Y (mm)
X (mm)
35
30
25
20
15
10
15
20
25
30
35
40
d
t
z
s
g
k



W12
z
s
d
t
tʃ
ʃ
Y (mm)
X (mm)
15
10
5
0
8
13
18
W24
z
s
d
t
tʃ
ʃ
Y (mm)
X (mm)
23
18
13
8
8
13
18
23
W25
z
s
d
t
tʃ
ʃ
Y (mm)
X (mm)
25
20
15
10
10
15
20
25
W15
z
s
d
t
tʃ
ʃ
Y (mm)
X (mm)
20
15
10
5
10
15
20
25



Alveolars
r=0.498
D2 (mm)
Speaking Rate (ms/syl)
/s/ and /sh/ versus Postalveolars
r=0.549
D2 (mm)
Palate Width (mm)
Alveolars
r=-0.508
O2 (%)
Speaking Rate (ms/syl)
/s/ and /sh/ versus Postalveolars
r=-0.606
O2 (%)
Palate Width (mm)
Alveolars
r=-0.562
O2 (%)
Palate Curvature (α)
/s/ and /sh/ versus Postalveolars
r=-0.518
O2 (%)
Palate Curvature (α)
Alveolars
r=-0.455
O2 (%)
Palate Width (mm)
/t/ and /d/ versus Postalveolars
r=-0.528
O2 (%)
Palate Width (mm)