Title

Speech Production Patterns in Producing Linguistic Contrasts are Partly Determined by Individual Differences in Anatomy

Journal

UC Berkeley PhonLab Annual Report, 14(1)

Author

Johnson, Keith

Publication Date

2018

DOI

10.5070/P7141042483

Speech production patterns in producing linguistic contrasts are partly determined by individual differences in anatomy

Keith Johnson
Linguistics, UC Berkeley

Abstract

This study explored correlations between (a) measures of vocal tract anatomy and (b) measures of articulatory/linguistic contrasts in vowels and coronal fricatives. The data for the study come from the Wisconsin X-Ray MicroBeam Database (Westbury, 1994). The anatomical measures included vocal tract length, oral cavity length, palate size and shape, as well as measures of maximal tongue protrusion and jaw wagging amplitude. Measures of the articulatory vowel space included the range of x and y location at vowel midpoints for four pellets on the tongue, the interpolated highest point of the tongue, and the locations of pellets on the upper and lower lips and to the lower incisor. For each of these clouds of vowel midpoint measurements, the orientation of variation was also measured. For fricatives, measures of tongue advancement and tongue tip lowering were taken. The results showed that the articulatory vowel space was related to both the length of the vocal tract, and to the shape of the palate, while fricative variation was related to palate parameters alone. In simple correlations, the percentage of articulatory variance between segments that could be predicted by anatomical characteristics was modest; never more than 36% for vowels and 25% for fricatives. Canonical correlation analysis found two anatomical factors that predict articulatory patterns jointly in vowels and coronal fricatives. The first canonical variable found a relationship between vocal tract length/palate depth and vowel tongue vertical range and jaw motion. Talkers with long vocal tracts and deep palates showed large tongue vertical range and small jaw range. The second canonical variable found a relationship between palate depth and tongue tip raising in coronal fricatives. Talkers with more shallow palate tended to have a tongue-tip up posture in fricatives. Phonetic tagging for the XRMBDB is made publicly available by this project.

I. Introduction

In one conception of speech communication, the medium of communication is vocal tract gestures (Goldstein & Fowler, 2003). In this view, speech production is guided by specifications for the timing, location, and degree of vocal tract constriction actions (gestures), and speech perception is oriented around recovering these key action parameters. It is these discrete and recombinable “public” actions that serve as the “common phonological currency” among talkers and listeners.

An alternative view suggests that talkers and listeners achieve parity in terms of speech sounds. In this view, the units of speech communication are time/frequency acoustic patterns. The talker’s task (Perkell et al., 2000; Houde & Nagarajan, 2011) is to produce sounds that match an expected sound pattern, and the listener's task is to recognize the sounds produced by the talker (Pisoni & Sawusch, 1975).

Between-talker phonetic variability is a bane for both of these approaches.

If we assume that parity between talker and listener is achieved in terms of the sounds of speech (the auditory parity hypothesis) then between-talker variability is a complication because talkers have different vocal tract geometry and this variation leads to complicating acoustic differences. Research on perceptual talker normalization attempts to address this issue.

On the other hand, if we assume that parity between talker and listener is achieved in terms of the actions of speech (the gestural parity hypothesis) then individual differences in speech production pose a potential problem. To the extent that articulatory movement patterns differ from talker to talker, gestures may not be a particularly viable common currency. However, "vocal tract constriction actions", the common currency suggested above, are somewhat abstracted away from the details of articulatory motions and may limit the scope of individual differences that would pose a problem.

Beyond these theoretical considerations, it is important to understand and account for Individual differences in speech production in several domains of research having to do with speech. For example, in the study of **speech motor control** where the response to various forms of altered feedback is studied, it is potentially important to understand the individual articulatory strategies used by talkers. Similarly, in **clinical speech pathology** it is important to appreciate and perhaps be able to predict individual differences in articulatory control (Laine, 1986). At a more **theoretical linguistic level**, the concept that there is a pool of synchronic phonetic variation that feeds linguistic sound change (Ohala, 1993; Garrett & Johnson, 2013) highlights the importance of factors that influence phonetic variation, including individual phonetic differences.

This paper examines between-talker articulatory variability; aiming to contribute to the existing literature documenting individual differences and to our understanding of the degree to which individual differences in articulatory movements can be predicted from individual differences in vocal tract anatomy. After reviews of the research literature on anatomical vocal tract differences (section I.A) and relationships between anatomical features and articulatory strategies (section I.B), the methods of the current study are described in section II. The results in Section III examine anatomical predictors of the articulatory vowel space (III.A), and coronal fricative production (III.B), and a combined multivariate analysis (III.D). The paper concludes with a general discussion of the findings and prospects for future research.

A. Vocal tract differences

Human vocal tracts differ from each other in a number of different ways. Among the most easily measured and frequently studied anatomical differences between talkers are the **dental measurements** - the width of the maxillary arch measured from the left first molar tooth to right first molar; the depth of the palate as the maximal rise above the oral occlusal plane; from these we can measure the doming of the palatal arch in the coronal plane (Brunner et al., 2009). Measures of the upper dentition are mirrored by measures of the lower dentition, and the size of the jaw and tongue (Kuehn & Moll, 1976). In addition to measures of dentition, dimensions of the **vocal tract as a whole** differ between talkers. For example, the overall vocal tract length can be estimated fairly accurately from acoustic vowel formant frequency measurements (Lammert & Narayanan, 2015). The relative sizes of the oral and pharyngeal cavities of the vocal tract also differ. The ratio of the length of the oral cavity to the pharynx cavity length is probably the biggest vocal tract difference between men and women (Goldstein, 1980). Beyond these static measures of dentition and vocal tract, talkers may also differ in aspects of **articulator movement patterns**. For example, the maximum opening of the jaw differs between talkers, as does the maximum extent of tongue or lip protrusion (Westbury, 1994). Finally, the function of the temporo-mandibular joint and the type of occlusion or pattern of jaw translation and rotation may differ between talkers (Edwards & Harris, 1990).

This paper is about the link between anatomical vocal tract parameters such as these and speech production patterns. As will be reviewed in section B, much is known about this. However, much of our knowledge is based on studies with various serious limitations - either in the number of talkers involved in the study, or in the phonetic range of utterances that were studied. In addition, several of the studies in the literature focus on a single anatomical feature without considering a more comprehensive range of features. The study reported here is based on a relatively large number of talkers (n=48) and samples a larger range of phonetic/ articulatory variables, relating them to several different anatomical features of the talkers.

B. Linking Anatomy and Articulation

Table 1 outlines some of the key findings of prior research on how individual differences in vocal tract anatomy relate to differences in speech articulation. Two anatomical parameters have been studied most extensively - the doming of the palate, and the length of the vocal tract. These have been measured in different ways in the research literature. For both palate doming and vocal tract length, there is research looking at token-to-token articulatory variability in producing particular speech sounds (within-category variation), as well as articulatory strategies for producing phonetic contrasts between speech sounds (between-category variation).

Several studies, listed in section (1)(a) in table 1, have found that segments produced with constriction on the palate (/s/, /r/, /ɹ/, /u/, etc.) are pronounced more precisely (with smaller within-category variability) by talkers whose palates are relatively flat, as compared with talkers who have domed palates. Modeling studies (Bakst & Johnson, 2018; Lammert et al., 2011; Winkler et al., 2006) indicate that the articulation-to-acoustics mapping shows greater sensitivity when the shape of the palate is flat, so small movements produce relatively larger acoustic changes. Talkers who have relatively flat palates seem to respond to this by exercising greater control with smaller token-to-token variation in the placement of the tongue in coronal sounds than do talkers who have more deeply domed palates.

TABLE 1. *Key findings in literature on articulatory effects of individual anatomical differences*.

(1) Palate doming	(a) within-category	Perkell et al., 1995	flat palate -> no lip/tongue trading in /u/
		Brunner et al., 2009	flat palate -> less articulatory variability in vowels
		Rudy & Yunusova, 2012	flat palate -> more articulatory coarticulation in front consonants
		Alfwaress, 2015	narrow palate, short oral tract -> /r/ becomes glide /j/
		Bakst & Lin, 2015; Bakst, 2016	flat palate -> less articulatory variability in /s/ and /r/
	(b) between-category	Johnson, et al., 1993	flat palate -> more jaw movement in distinguishing /i/ and /ɛ/
		Perkell, 1997	flat palate -> smaller tongue position difference between /i/ and /ɛ/
		Weirich & Fuchs, 2013	flat palate -> bigger articulatory difference between /s/ and /ʃ/
(2) Vocal tract length	(a) within-category	Winkler et al., 2011	longer % oral -> /u/ variation is more horizontal; /i/ variation is more vertical
		Rudy & Yunusova, 2012	longer palate length -> more vertical coarticulation in front C’s
	(b) between-category	Kuehn & Moll, 1976	longer VT -> higher velocity in CV sequence
		Honda et al. 1996	longer VT -> more horizontal (45˚) in /a/ to /i/ movement. Short vocal tract is more vertical
		Simpson, 2001; 2002	male -> larger and faster tongue movement in /a/, /i/ sequence; and in /e/, /a/
		Fuchs et al., 2008	longer VT -> low to high V has larger vertical distance.

Palate shape is also implicated in how front segment contrasts are produced, section (1)(b) in Table 1. For example, Johnson et al. (1993) suggested that talkers with domed palates had a greater degree of independent tongue movement in making front vowel height distinctions than was found for talkers with flatter palates, who tended to change the height of the tongue, primarily by moving the jaw (jaw movers versus tongue movers). Similarly, Perkell (1997) reported that the difference in height between /i/ and /ɛ/ was larger for talkers who have a domed palate. On the other hand, with a different sort of phonetic contrast, Weirich & Fuchs (2013) found that the articulatory difference between /s/ and /ʃ/ was greater for talkers with flat palates.

Within-category variation has also been tied to the length of vocal tract (or the relative size of the oral cavity) (Table 1, (2)(a)). The pattern of variation that has been studied is the orientation of within-category variation, rather than the amount of variability (as was discussed above for palate doming). The main findings in the literature are that the vowel /i/ (Winkler et al., 2011) and the coronal consonants /t/, /d/, etc. (Rudy & Yunusova, 2012) tend to show a more vertical pattern of within-category variation when the oral cavity is long relative to the pharynx. Winkler et al. (2012) also found that within-vowel variation of /u/ was more horizontal for talkers who have a greater oral/pharyngeal ratio.

Several studies have also considered whether the size of the vocal tract may have an effect on individual differences in between-category phonetic realization (Table 1, (2)(b)). Kuehn & Moll (1976) found that tongue velocity in CV syllables was greater for talkers with larger jaw and tongue volume dimensions. They argued that talkers with larger vocal tracts must move the articulators further and faster to utilize the same relative articulatory/acoustic space used by talkers with smaller vocal tract dimensions.

Most other studies dealing with vocal tract length differences and between-category variation have looked at the pattern of motion involved in moving from /a/ to /i/ (Honda et al., 1996; Simpson, 2001; 2002; Fuchs et al., 2008). The general finding is that talkers with longer vocal tracts (who differ primarily in pharynx length) tend to have larger differences between /a/ and /i/ tongue positions (Simpson, 2001; 2002; Fuchs et al., 2008), and to have greater horizontal displacement between /a/ and /i/ (Honda et al., 1996).

One weakness of prior research is that the effects of vocal tract anatomy on speech production have been carried out in a somewhat piecemeal way. Studies that look at effects of vocal tract length, don't simultaneously consider possible confounding or contrasting effects of palate shape or size; and similarly, studies with a focus on the palate don't consider vocal tract length. The piecemeal approach also is seen in our focus on particular segments, not considering how general trends in articulation might be found by looking at a wider range of phonetic material than just one or two phones.

The present study was more comprehensive than prior research. Several measures of vocal tract anatomy, dealing with palate size and shape, as well as with vocal tract length and the relative size of the oral and pharyngeal portions of the vocal tract, were used to predict individual articulatory variability in vowel and consonant contrasts, with a single group of talkers. One focus of the current study is the between-category shape of the articulatory vowel space, the range of articulatory postures taken when different vowel contrasts are made. The other main focus of the paper is also a study of phonetic contrast between categories, focussing on coronal fricative production. Within-category variability will be the focus of a future paper.

II. Methods

A. The Database

i. Talkers. This study makes use of the Wisconsin x-ray microbeam speech production database (Westbury, 1994), which consists of audio and articulatory recordings from 48 talkers (22 male, 26 female). Average age at the time of recording was 22.3 years. The majority of these talkers were from the state of Wisconsin in the United States and speak in a variety of English that was typical for that area.

ii. Articulatory data. The articulatory data are the locations of small gold pellets in the sagittal plane (see Figure 1a). The pellets were glued onto key articulators in the mid-sagittal plane - the upper lip (UL), lower lip (LL), lower incisor (MI), and four pellets on the tongue (T1, T2, T3, and T4) at approximately 15 mm spacing between the tongue pellets. The locations of these pellets were tracked by an x-ray control and detection system that focussed a small beam of x-ray in a sagittal section patch predicted from the last sample, and then determined the precise location of the pellet within the small patch of x-ray. Each pellet was sampled 145.64 times per second, producing smooth trajectories of articulator movement.

iii. Coordinate system. A bite plate was used to register the occlusal plane (Westbury, 1994) and the data was rotated so that the occlusal plane was horizontal, and the data were translated so that the (0,0) point in the coordinate system was at the tip of the upper incisor. Thus, in this study ‘horizontal’ means parallel to the occlusal plane in the sagittal plane, and ‘vertical’ means perpendicular to the occlusal plane in the mid-sagittal plane. These are illustrated by the measures in Figure 1a of “Oral cavity length” which is a purely horizontal measure, and “Palate height” which is a purely vertical measure.

iv. Palate trace. A palate cast was made for each participant (Westbury, 1994) and a trace of the sagittal midline of the hard palate was taken from the cast using the x-ray microbeam apparatus to register the locations of pellets in a chain draped into the inverted cast. The same customized calibration bite plate that was used to register the occlusal plane for the talker’s data was also used to calibrate their palate trace.

v. Pharynx Trace. From a sagittal x-ray of the vocal tract, two points along the posterior pharyngeal wall (one relatively superior and one relatively inferior) were located. The line connecting these points “provides a useful, albeit coarse, approximation of the position and orientation of the dorsal-most surface bounding the pharynx” (Westbury, 1994, p. 51).

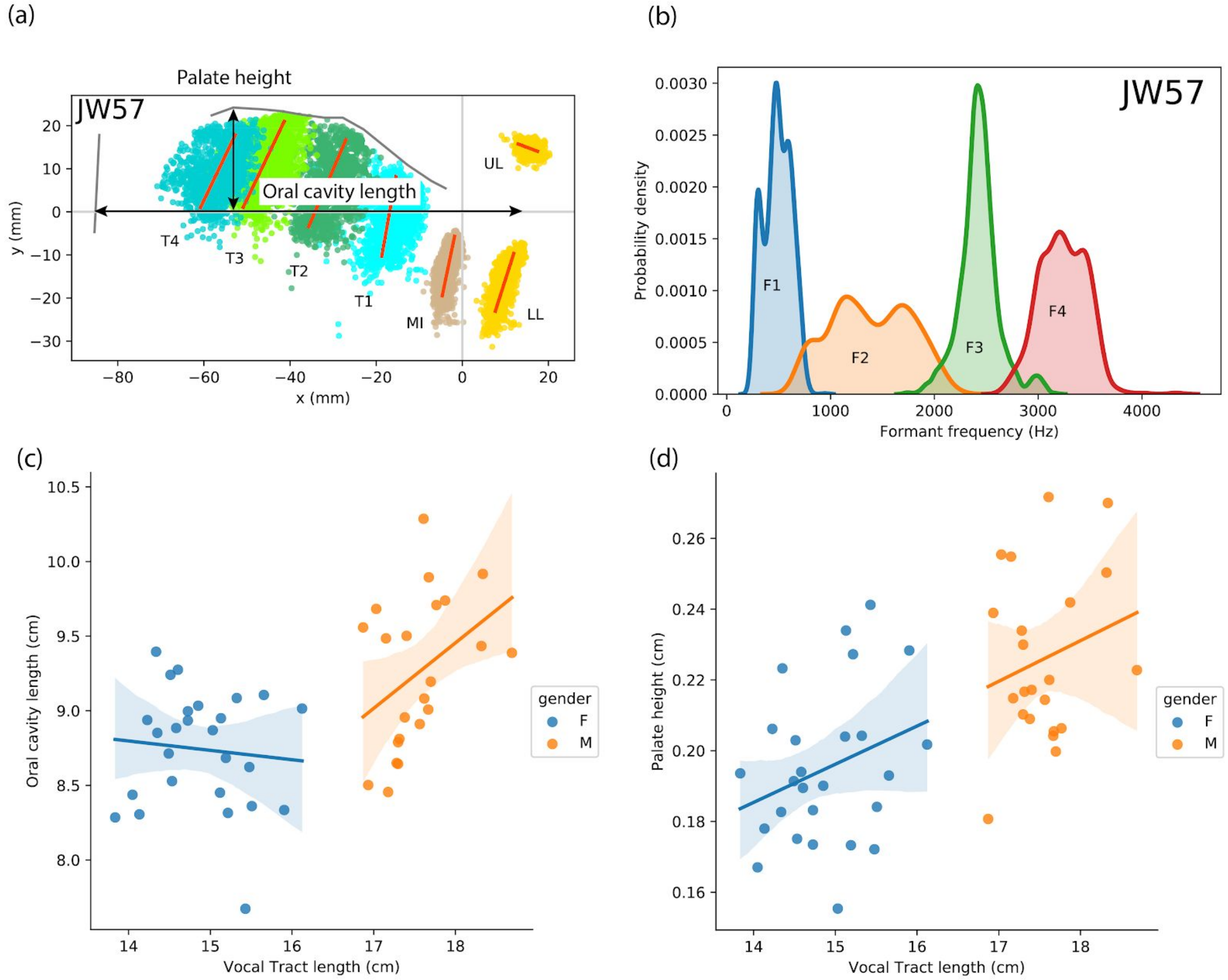

FIG 1. Illustration of the data and some measures of vocal tract anatomy. (a) A view of the sagittal section of the vocal tract showing locations sampled for each of the pellets attached to the speech articulators. Pellets UL and LL were attached to the upper and lower lips. MI was attached between the lower incisors, and pellets T1-T4 were attached to the tongue. The data points are the vowel midpoint pellet locations for talker JW57 for all of his vowel tokens. The first principal component of variation for each pellet is shown as a red line, and the "oral cavity length", and "palate height" measures are illustrated with the black double-headed arrows. (b) Example of vowel formant distributions for talker JW57. (c) Vocal tract length (horizontal axis) and Oral cavity length (vertical axis). Each dot is a talker (n=48) and separate linear fits are given for the men and the women. (d) Vocal tract length (horizontal axis) and palate height (vertical axis).

vi. Speech tasks. In a full recording session, for each talker, there were 118 separate recordings ranging in length from 3 to 25 seconds. Fifty of these were recordings of short lists of isolated words or word-like sequences (V, VCV, sVd). Forty-five of the recordings were of lists of disconnected sentences (e.g. “Across the street stands a country school.”). Six of the recordings were of connected speech in reading meaningful paragraphs. Seventeen of the recordings can be classified as nonspeech or ‘non-language’ records of various types including swallowing, tongue protrusion, and jaw-wagging, but also rapid repetition of speech gestures like [tətətətətə…]. Non-speech recordings with no transcribable acoustic speech signal are not included in the data here, but I did use tongue protrusion and jaw wagging as anatomical features (see section B, below).

vii. Phonetic tagging. Audio was simultaneously recorded with the articulatory data (Westbury, 1994). Transcripts of the words and phones were time aligned with the audio recordings so we can identify acoustic onsets and offsets of segments for the analyses reported here. The initial transcripts were produced automatically using forced alignment (Yuan & Liberman, 2008), and then hand corrected in Praat (Boersma & Weenik, 2018) by a team of phonetically trained undergraduate Linguistics students at UC Berkeley. Hand correction was done by two workers, in sequence, reviewing and correcting each other’s work. These transcripts are available at https://github.com/rsprouse/xray_microbeam_database.

B. Measurement of anatomical features

The **length of each talker’s vocal tract** was estimated from their vowel acoustics. The acoustic theory of speech production (Fant, 1960) suggests that vocal tract length can be accurately estimated from the average spacing of the formants (the ΔF). ΔF is the slope of the line fitting formant frequencies to formant number, and can be calculated as in (1). Vocal tract length is a function of ΔF (2). The line-fitting calculation of ΔF correlates (r = 0.99) with Lammert & Narayanan’s (2015) linear regression formula to estimate the length of the vocal tract from short stretches of speech. Using vocal tract length measurements from MRI scans, they estimated that the error of this approach using short stretches of speech was under 1cm.

1) $\Delta F = 1/mn \sum_{j}^{m} \sum_{i}^{n} [F_{ij}/(i-0.5)]$, where *i* = formant number (1...4), and *j* is token number.

2) $VTL = c/2\Delta F$, *c* = 34000 cm/s, the speed of sound, warm moist air

The formant frequencies for this study were measured at the midpoints of stressed instances of the vowels [u], [ʊ], [oʊ], [ɑ], [ɔ], [ʌ], [æ], [ɛ], [eɪ], [ɪ] and [i]. Three different algorithms were used to estimate the formants: autocorrelation LPC with formant trajectory smoothing (Markel & Gray, 1976; Talkin, 1987), inverse filter control (Watanabe, 2001), and robust LPC (Lee, 1988). At intervals of 5ms, the median of these three measures was recorded as the estimated formant value for F1-F4 at that time point. Figure 1b shows an example for one talker of the distributions of the vowel midpoint formant measurements that went into the estimation of vocal tract length in this study.

The **oral cavity length** was taken as the horizontal distance in centimeters between the average horizontal location of the upper and lower lips and the average horizontal location of the pharyngeal trace. The **pharyngeal cavity length** was then taken as the difference between the overall length of the vocal tract and the length of the oral cavity. Figure 1c shows vocal tract and oral cavity length estimates for the men and women in this study. The men generally had longer vocal tracts than the women, while there was substantial overlap between the two groups in oral cavity length. There was a tendency for men with longer vocal tracts to have longer oral cavities as well.

The **doming of the palate** was measured in three ways. These are highly correlated with each other ($r > 0.8$ for each comparison). The first and simplest measure, illustrated in Figure 1a, is to record the maximum vertical value on the sagittal palate trace **-** the **palate height**. A second measure of palate doming takes the product of palate height and the absolute horizontal maximum of the palate trace, which corresponds roughly to the location of the posterior nasal spine (PNS) at the boundary between the hard and soft portions of the palate. This is a rough measure of the **total area** under the palate. The **mean area** under the palate was computed by taking the area under the palate trace curve and then normalizing it by the horizontal extent of the palate trace. For the fricative analysis, I also measured the horizontal **location of the maximum** vertical value on the palate trace. This gives a measure of whether the peak occurred close to the alveolar ridge or further back (see Lammert et al., 2011). As shown in Figure 1d, there was a positive relationship between vocal tract length and palate height ($r = 0.618$) and this relationship holds within men and women.

Westbury (1994) describes two nonspeech tasks that subjects were asked to complete which may be related to articulatory range of motion. In one, the talker was asked to stick his/her tongue out of the mouth as far as possible. The 95% quantile of horizontal location of pellet T1 during this recorded task is a measure of **maximum tongue protrusion**. There was a rough correlation ($r = 0.397$) between oral cavity length and maximum tongue protrusion. In a second range of motion task, the subject was asked to open and close the jaw repeatedly, opening to the maximum extent possible. The vertical location of pellet MI at the 95% quantile during jaw wagging is a measure of **maximum jaw opening**.

Finally, Westbury (1994) presented a table showing a **bite classification** for each of the subjects in the database. Simplifying his nuanced discussion for the sake of grouping subjects, 38 of the 48 subjects were classified as having a normal bite. Of the remainder, eight (4 men) were classified as having a class II “overbite”, and one man was classified as having a class III “underbite”, and one man was characterized as having a bilateral posterior “crossbite”. Given the emphasis put on jaw angle by Honda et al. (1996), I briefly explored the possible articulatory correlates of bite classification in these data despite the imbalance in number of subjects in each group.

C. Measuring the Articulatory Vowel Space

Pellet locations in the sagittal plane were taken at the midpoints of stressed instances of the vowels [u], [ʊ], [oʊ], [ɑ], [ɔ], [ʌ], [æ], [ɛ], [eɪ], [ɪ] and [i]. For each talker there were about 1500 tokens. The lowest number of tokens for a talker was 842 (talker JW46), followed by 1183 (talker JW19) and 1313 (talker JW32). The lowest number of tokens for any vowel per talker was 11 tokens of [ʊ] for talker JW46, and the greatest number of tokens was 217 instances of [ɪ] by talker JW12.

The number of words representing any one vowel varied. For example, for talker JW11 the 141 tokens of [u] came from 32 lexical items. Most of these were said only once, but there were 38 instances of 'two', 20 instances of 'too', 15 instances of 'school', 11 instances of 'you', 10 instances of 'shoot', and 8 instances each of 'do' and 'use'. Each vowel is represented by a somewhat unbalanced set of coarticulatory and prosodic contexts. For this study of the articulatory vowel space, I decided to take all of the tokens of these vowels because I was interested in a comprehensive sample of each talker's articulatory range for vowel production. Note that reduced productions were transcribed as such. For example, an instance of 'to' where the vowel is reduced to schwa is transcribed as [tə], and so doesn't contribute to the variation found in [u].

For each x-ray microbeam pellet we have a cloud of data points in the (x,y) mid-sagittal plane. These clouds were illustrated above in Figure 1a for talker JW57. In this figure there are 1449 data points in each pellet's cloud of observations. The articulatory space was then characterized by three measurements for each pellet: (1) the horizontal and (2) vertical range of pellet positions, and (3) the orientation of the cloud. The **horizontal range** was calculated by finding values on the x-axis that encompassed 80% of the x-axis variation (the 10% and 90% x-axis quantiles) and expressing the magnitude or range of the x variation as the absolute value of the difference between the quantiles. This captures in millimeters the range of positions taken by the pellet on the x-axis of the mid-sagittal plane. The **vertical range** of variation was calculated in the same way for the range of positions taken on the y-axis. The **orientation** of the cloud of variation for each pellet was found by calculating the first principal component of variation (this is illustrated by the red lines in Figure 1a) and then calculating the angle of the principal component relative to the occlusal plane. An angle of 0 (or 180) means that the first principal component of variation is horizontal (parallel to the occlusal plane), while an angle of 90 means that PC1 is vertical. The angle of the first principal component was then transformed into a **verticality measure** (formula 3), which ranges from 0 (horizontal) to 1 (vertical). Pellets for jaw (MI) and lower lip (LL) tended to have more vertical orientation, while the orientation of tongue pellet distributions varied from talker to talker.

3) $\text{verticality} = (90 - |A-90|) / 90$, where A is an angle from 0 to 180 degrees

D. Measuring Coronal Fricative Production

In a procedure that was similar to that for the vowels, the vertical and horizontal pellet locations in the sagittal plane were taken at the segment midpoint for all instances of [s, z, ʒ, ʃ, ʧ, ʤ, ð, θ] that were longer than 50 ms. This resulted in 38,192 tokens, with over 800 from the typical talker, and at least 415 for the talker with the smallest amount of data. The segments were grouped into three classes, and data from these classes were pooled. [s] and [z] tokens were pooled as alveolar fricatives, the voiceless and voiced dental fricatives [ð] and [θ] were pooled,

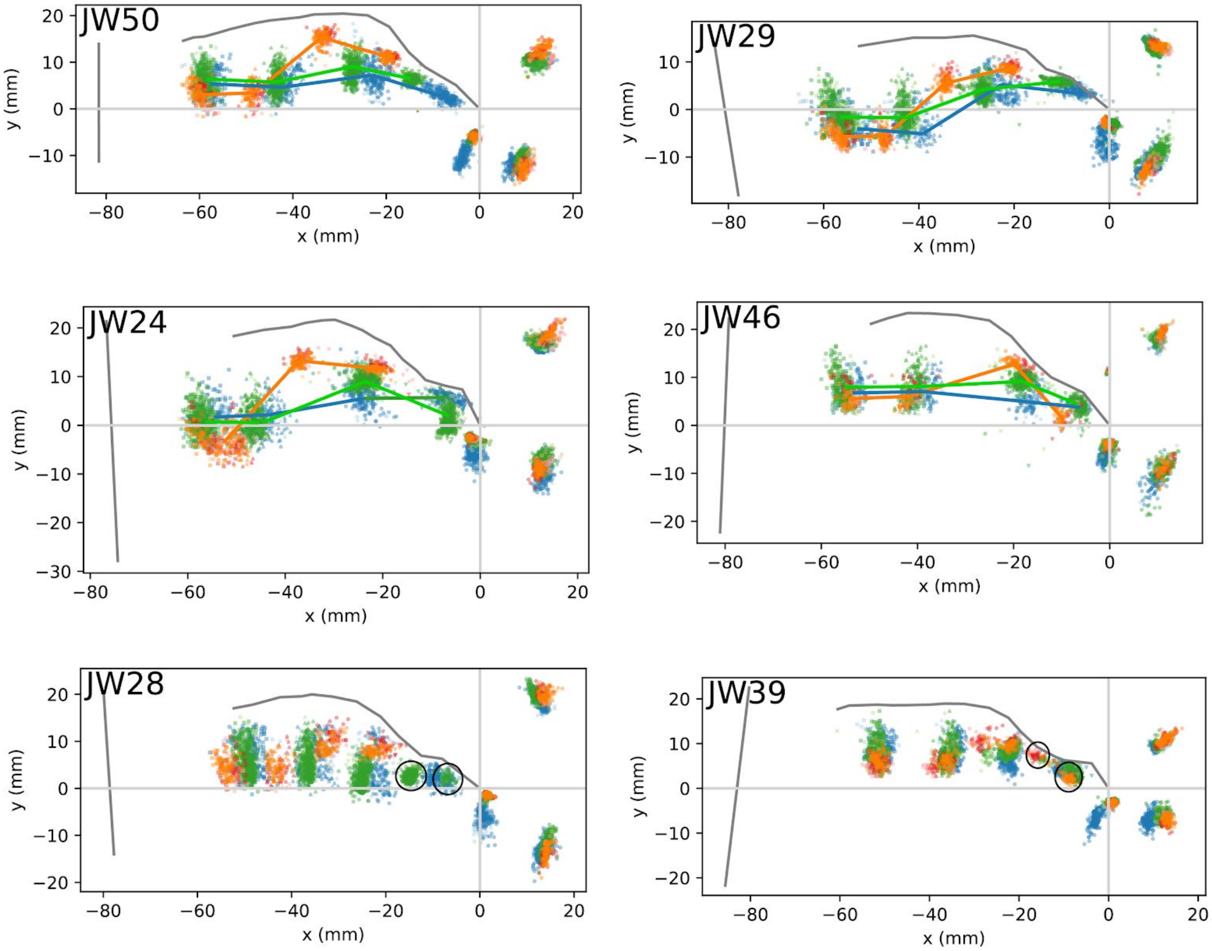

FIG 2. Fricative data from six talkers illustrating some individual differences. The graphs on the left are male talkers and those on the right are female talkers. Data for the segments are plotted in different colors: dental segments in shades of blue, alveolar segments in shades of green, and post-alveolar segments in shades of orange or tan. In the top four graphs, lines connect the tongue pellet clouds for each of the places of articulation. In the bottom two graphs, circles indicate distinct bimodal clusters of T1 location for alveolars in JW28 and for post-alveolars in JW39.

and the post-alveolar fricatives [ʃ] and [ʒ], together with the post-alveolar affricates [ʧ] and [ʤ] were pooled to form a third class. I decided to take instances of both post-alveolar fricatives and affricates in the third group because (1) this class of segments was less frequent in the corpus, and (2) upon inspection of the data, it appeared that talkers produced them without much difference. This can be seen in Figure 2 in the tight clustering of orange (fricative) and tan (affricate) data points.

Visual inspection of the data (Figure 2) suggested that a few parameters could be used to probe for individual differences and whether they are related for vocal tract anatomy. These measures start by taking the mean values of the pellet sagittal x and y locations separately for the three places of articulation - dental, alveolar, and post-alveolar. Measures of within-class **posture**, or between-class **difference** were then calculated for each talker.

First, three measures of **tongue posture** during a segment were taken. These had to do with tongue tip raising. To measure this, I took the vertical location of tongue pellet T1 relative to the vertical location of pellet T2 (the difference in value of T1y and T2y) for the dentals, alveolars and post-alveolars. Notice for example in Figure 2 that talker JW46 has a **tip-down** posture for both the alveolar and the post-alveolar segments while talker JW24 has a tip-down posture for the alveolars, but not for the post-alveolars.

Second, based on visual inspection of the data, I took six measures of **differences** between segments. **Post-alveolar tongue retraction** was taken by subtracting the mean T1x value for post-alveolars from the mean T1x for alveolars. **Dental fronting** was measured by subtracting T1x in alveolars from T1x in dentals. **Dental blade lowering** was measured by subtracting T2y in alveolars from T2y in dentals. **Post-alveolar lip rounding** was calculated separately for the upper and lower lip subtracting the LLx or ULx of post-alveolars from that for alveolars. Finally, **dental jaw lowering** was measured as the difference between the mean MIy value in dentals and alveolars.

The previous literature on individual differences in vowel production lead to the expectation that overall vocal tract length, and perhaps palate doming will be tied to the vertical range of the tongue body and horizontal range of the tongue blade during vowels. Additionally, based on prior research, we would also expect that vocal tract length may not have much impact on fricative production, but palatal doming may.

III. Results

A. The Articulatory Vowel Space

Correlations among vowel measurements (articulatory range, and verticality) and the anatomical features are shown in Figure 3. The top portion of the correlation matrix shows how the anatomical features are correlated with each other. For example, the three measures of palate doming - palate height, mean area, and total area under the palate are positively correlated. We also see that overall vocal tract length is correlated with several of the other anatomical features including pharynx length, oral cavity length, and the measures of palate doming. A couple of these correlations were shown earlier in Figures 1c and 1d.

Figure 3 also reveals several interesting relationships among the measures of articulatory vowel space. For example, logically enough, there is generally a negative correlation between verticality and the x range of a pellet and a positive correlation between verticality and the y range. The between-articulator correlations also probably reflect anatomical links between vocal tract structures.

Turning now to the relationship between vocal tract anatomy and individual differences in speech production, it is interesting to note that the anatomical features were not highly correlated (|r| < 0.4) with any of the lip or jaw range measures in vowel production. Instead, measures of vocal tract size and shape were more strongly correlated with T1/T2 horizontal range and T3/T4 vertical range. Thus, also, the verticality of the T4 cloud of data was also related to vocal tract length, length of the pharynx cavity, and to palate height.

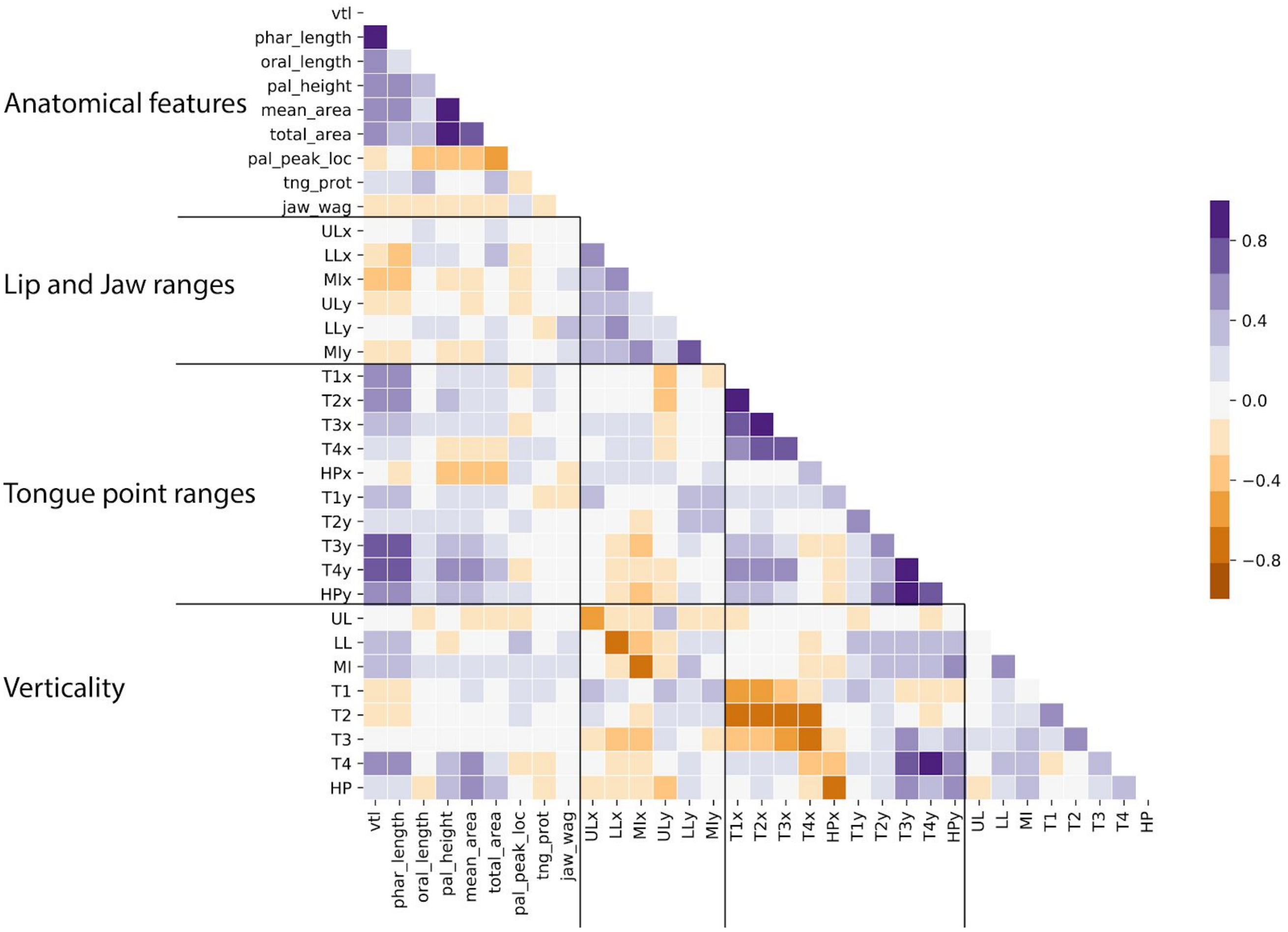

FIG 3. Correlations between quantitative measures of vocal tract anatomy and articulatory data from vowels. 'Vtl' is vocal tract length, 'tng_prot' is maximum protrusion of the tongue, 'jaw_wag' is maximum jaw opening.

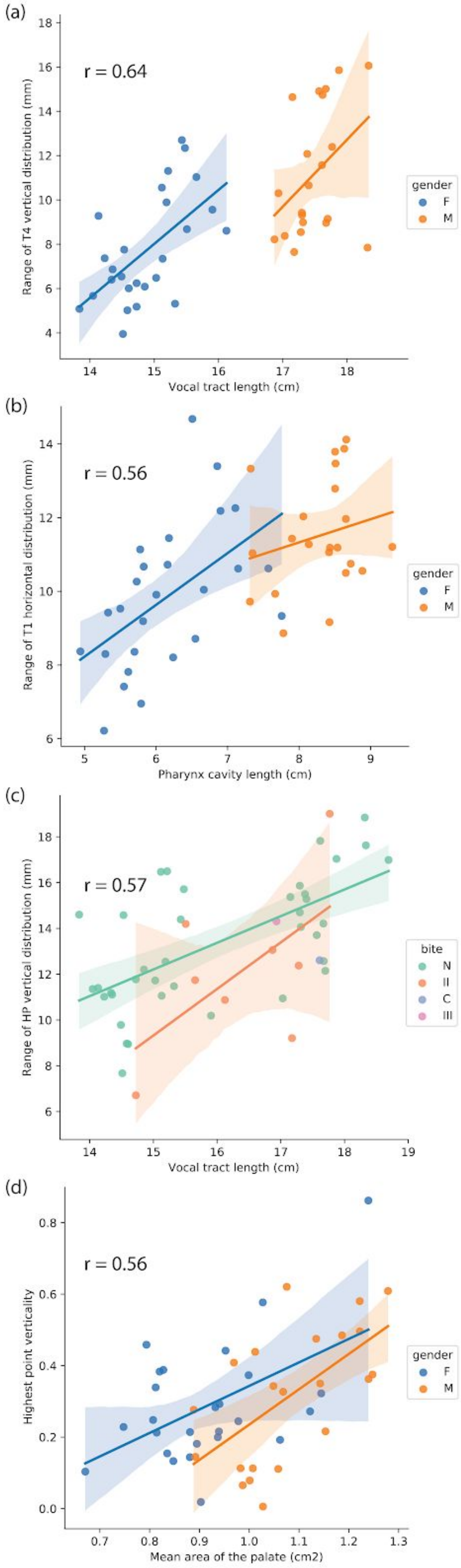

FIG 4. Visualizing select correlations from Figure 3. In these graphs each dot is a talker. (a) Relationship between vocal tract length and the vertical range of pellet T4. (b) Relationship between pharynx length and the horizontal range of the tongue tip pellet, T1. (c) Relationship between vocal tract length and the vertical range of the highest point of the tongue. (d) Relationship between mean area under the palate and the verticality of HP.

Figure 4 shows the strongest correlations in more detail. These graphs show separate linear fits (with bootstrapped confidence intervals) for men and women (4a, b, d) or in one case for normal versus type II overbite (4c). Three of the main correlations between anatomy and vowel production relate the vocal tract length or pharynx length to an articulatory phonetic feature of the vowel space, and one (4d) relates a feature to the palate shape.

The correlation between vocal tract length and the vertical range of the tongue body (T4y) (r=0.64) is shown in Figure 4a. Both men and women showed a positive correlation between vocal tract length and the vertical range of the tongue body position in English vowels. The longer the vocal tract the greater the tongue body vertical range.

There was also a correlation between pharynx length and the horizontal range of tongue tip (T1x) (r=0.56; Figure 4b). With shorter vocal tract size we are more likely to see less tongue horizontal range in vowel production.

Figure 4c shows the relationship between the vertical range of the cloud of “highest point on the tongue” data points (HPy) and the length of the talker’s vocal tract. Talkers with a long vocal tract tended to have more vertical orientation of the highest point comparing vowels. Interestingly, this relationship seemed to be modulated by the talker's bite. Talkers with a type II bite (an overbite) tended to have bit more horizontal orientation, than other talkers with about the same vocal tract length. The sample of talkers with an overbite is quite small; only 8 talkers all together. Still, there is some possibility that overbite (essentially a retraction of the mandible) may counter the tendency for long vocal tract to have more vertical orientation of tongue body.

Finally, Figure 4d shows that the relationship between palate doming and highest point vertical range (HPy) is positive. The deeper the palate the larger the T4y range. Both men and women showed this positive relationship.

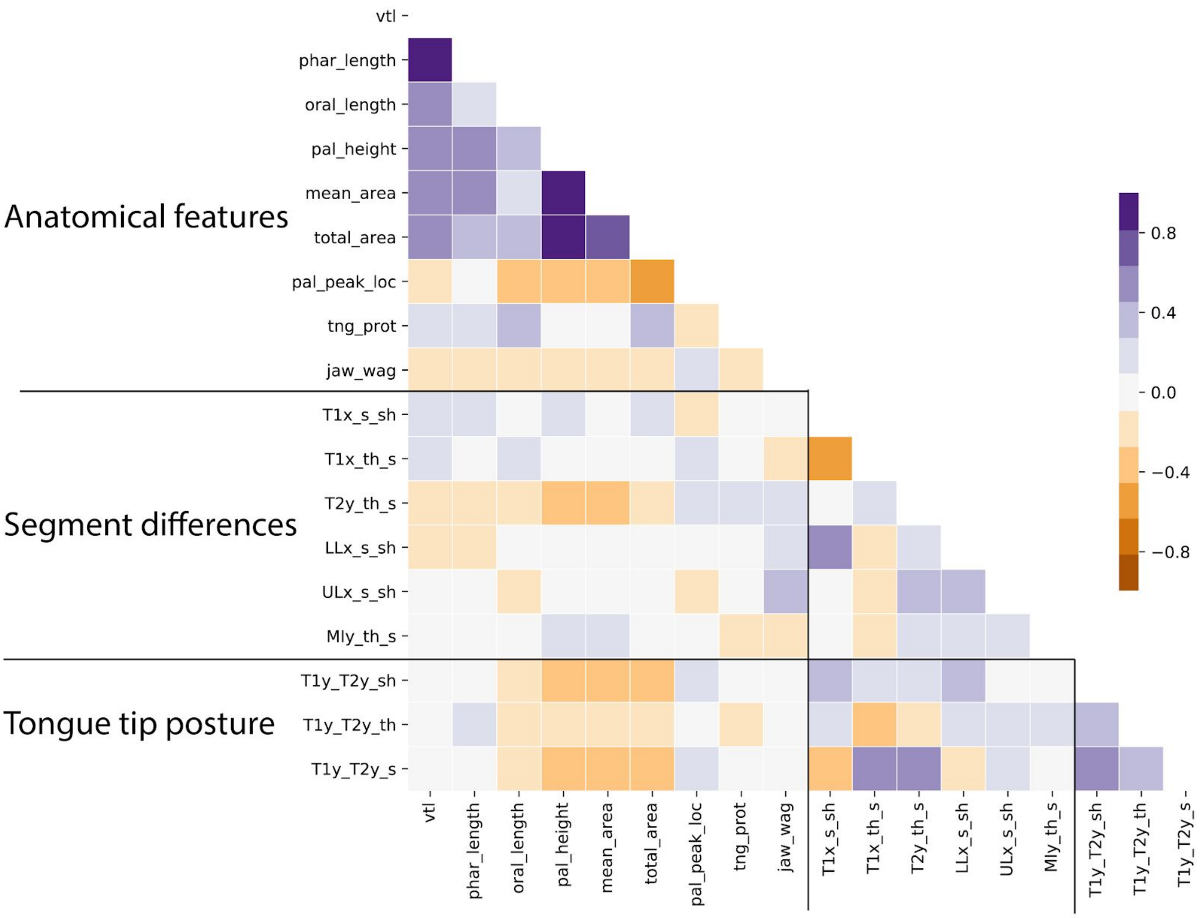

FIG 5. Correlations between quantitative measures of vocal tract anatomy and articulatory data from fricatives.

B. Coronal Fricative Production

Turning now to fricative production, the correlation matrix relating anatomical features to measures of fricative articulation (Figure 5), presents the same anatomical feature correlations that were plotted in figure 3, but in this figure we relate them to measures of coronal fricative posture or contrast.

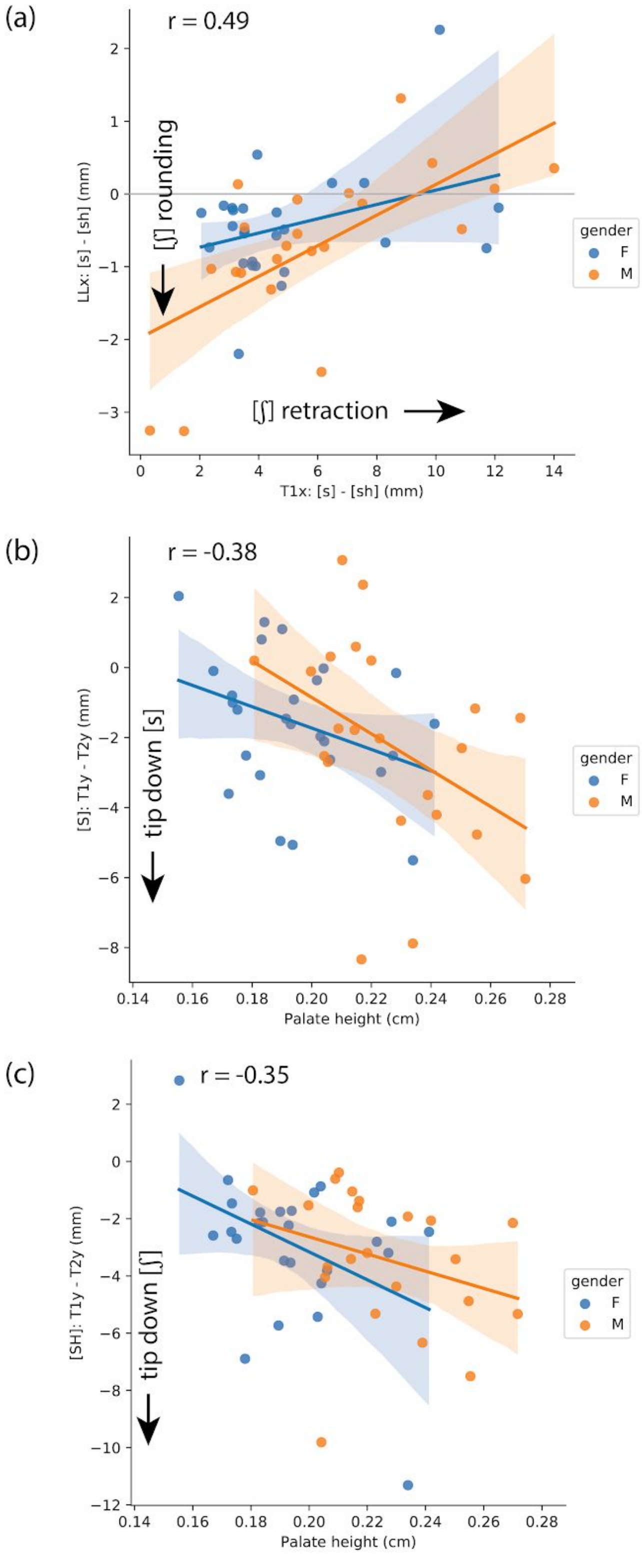

FIG 6. Selected correlations in coronal fricative production. (a) a trading relation between lip rounding and tongue tip retraction in the production of [ʃ]. The variables are the difference in LLx position between [s] and [ʃ] versus the difference in T1x position, (b) the relationship of palate height with tip posture (T1y - T2y) in [s], and (c) the relationship of palate height with tip posture (T1y - T2y) in [ʃ].

Figure 6 shows three of the strongest correlations from Figure 5. The first correlation (fig 6a) is an interesting trading relation in the production of [ʃ]. The figure shows that there is a tendency (r=0.49) to have a greater amount of [ʃ] lip rounding, compared to the posture of the lips in [s], if the tip of the tongue in [ʃ] is not retracted. The greater the [ʃ] tip retraction the less the lip rounding.

Figure 6b shows the negative correlation of palate height with tongue tip posture in [s] (r = -0.38). There is a trend for talkers who have a higher/deeper palate vault to have a more tongue tip down posture in [s]. Fig 6c shows that this relationship between deep palate and tip down posture also holds for [ʃ] (r = -0.35). One interesting aspect of these observations is that there is no hint of a correlation between overall vocal tract length (a feature that was so dominantly important for the vowel articulatory space) and tip-down posture (all |r| < 0.1).

However, pharynx length was moderately correlated with tongue tip retraction in [ʃ] (T1x(s) - T1x(ʃ)). This positive correlation (r=0.28) suggests that talkers with longer pharynx tended to have a more retracted [ʃ].

We will detour briefly in section C to discuss coronal fricative production in three specific individuals because of their unique patterns of production, and then in section D we will return to the correlation matrices and use canonical correlation to discover multivariate patterns of relating vocal tract anatomy to the combined vowel and fricative production data.

C. Three Unique Coronal Fricative Patterns

Figure 2 showed two different clusters for the tongue tip (T1) pellet in the realization of the post-alveolars in the speech of **talker JW39**. This talker, unlike the others, had a difference between the post-alveolar fricatives [ʃ] and [ʒ] and the post-alveolar affricates [ʧ] and [ʤ]. The tongue tip was more anterior in [ʃ] and [ʒ] than in [ʧ] and [ʤ]. Out of the 79 instances of the fricatives, 69 of them (87%) were produced with a fronted tongue tip (T1x > -12.5), while only 25% (18/73) of the instances of the affricates were. In a sense, this isn't surprising because we routinely see that the alveolar fricatives [s] and [z] differ from [t] and [d] in their particular place of articulation. Interestingly, most talkers in the database produce the post-alveolar affricates and fricatives with more nearly the same tongue posture, making talker JW39's pattern the exception rather than the rule.

We also saw in Figure 2 that **talker JW28** produced the alveolar fricatives [s,z] with two distinct tongue tip positions, one more fronted than the other. It seemed reasonable to expect that there might be some segmental contexts where this talker produced a fronted [s,z] (for example if the next phone was dental [θ] or [ð]). To explore this possibility, I classified the [s] tokens produced by talker JW28 as "fronted" if the T1 pellet in [s,z] was located in front of x= -10 (T1x > -10). This threshold is based on visual inspection of the two clusters in figure 2. The other alveolar tokens were classified as "retracted". This classification was then cross tabulated with the previous and following phones. The usual pronunciation of alveolar fricatives for this subject was in the 'retracted' category, which was actually a posture that was still closer to the dental fricatives than to the post-alveolar fricatives. So the unusual, perhaps contextually conditioned, posture for this talker seems to be the fronted position of the tongue tip.

There were 59 'fronted' tokens of [s,z] for this talker and they were distributed over a range of segmental contexts. There was a tendency for alveolar fricatives to be fronted before [ɛ], but of the 60 instances before [ɛ] only twelve of them were fronted. There was a stronger tendency for [s,z] to be fronted before [f], but even here only about half of the [sf] or [zf] sequences (8/17) showed the fronted variant. Preceding context also had an inconsistent effect on [s,z] realization. The most frequent occurence of the fronted variant was after an [l] where the probability of fronting was 9/15, and after [u] the probability was 5/12. The overall finding here is that the two variants of the alveolar fricatives [s] and [z] in the speech of talker JW28 do not seem to be conditioned in any simple way by the neighboring phonetic context.

Finally, **talker JW58** produced the post-alveolar fricatives and affricates in a unique way. What we see for the post-alveolar segments in Figure 7 (the orange, red, and tan plot symbols) is a velarized post-alveolar. A sound like this - simultaneous [ʃ] and [x] - has been described in some dialects of Swedish and has the IPA symbol [ɧ]. It isn't clear how this talker got this variant of the post-alveolar place of articulation, though he is listed as having studied Hebrew, a language that has the velar fricative [x], from the age of 6, so perhaps there could be some influence of an L2 gesture on his L1 production. Also, he is one of the few talkers in the database who identified his dialect 'home' as New Jersey. In listening to this talker I wouldn't have guessed that he had this pronunciation of [ʃ], and it is striking that he is very consistent in using this articulatory posture for all of the post-alveolar tokens - i.e. there are not separate clouds of points for his post-alveolar affricates and fricatives, or for any other subsets of segments or contexts.

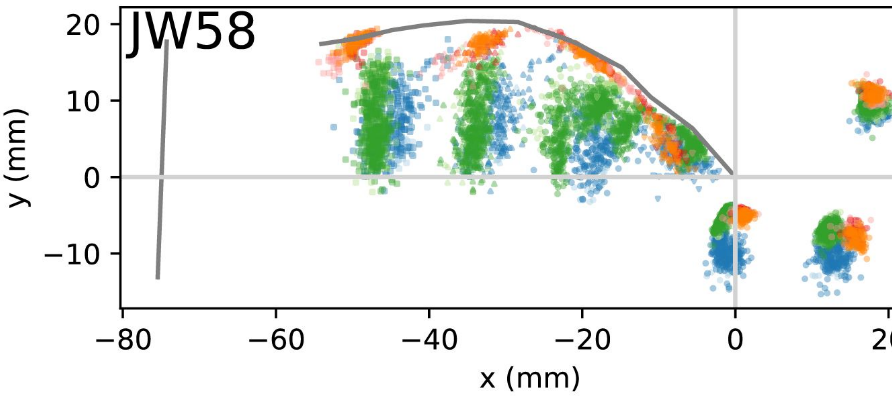

FIG 7. Mid-sagittal view of pellet locations during talker JW58's coronal fricatives and affricates at the segment temporal midpoints.

There is also some indication in Figure 7 that this talker has both retracted and fronted variants of [s,z]. All of the instances of [s] or [z] that were preceded by the rhotic vowel [ɚ] were retracted. This is a clear case of a 'categorical' coarticulatory effect (Nolan, Holst & Kühnert, 1996) where context dictates a change in tongue posture for a segment. However, in carry-over or anticipatory coarticulation at a distance this retracted/rhoticized variant of [s,z] occurs only stochastically. For instance, two of the five instances of "dark suit" have retracted [s]; retraction occurs in 2 of 10 instances of 'street', and 1 of 5 instances of 'stream'. The retracted variant also sometimes appears without the 'conditioning' environment. For example, one of the eight repetitions of the nonword [sa] had the retracted [s]. It is perhaps important to note that this retracted/rhoticized [s] is not the same as this talker's [ʃ], which, as noted above, is very unique.

D. Multivariate analysis

This section presents a multivariate analysis that jointly models covariation among the various anatomical measurements, together with covariation among the combined articulatory measures from the articulatory vowel space and the coronal fricative parameters. Using the anatomical and articulatory variables introduced in section II, there are 9 measures of vocal tract anatomy for each of the 48 talkers (we will call this the **X** dataset) and 33 articulatory phonetic variables for each of the 48 talkers (we will call this the **Y** dataset). The correlation matrices in figures 3 and 5, above, indicate that there are several strong covariances among these measures. The analysis in this section looks at multivariate projections of the datasets to find correlations between latent variables in **X** and **Y**, i.e. the strongest correlations between patterns in the VT anatomy variables and patterns in the articulatory phonetics variables.

1. Canonical correlation

The canonical correlation analysis (CCA) implementation that was used in this study is from the pyrcca python library described by Bilenko & Gallant (2016). They give the following description of CCA:

> Given two zero-mean datasets **X** and **Y**, $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_n}) \in \mathbb{R}^{d\times n}$ and $\mathbf{Y} = (\mathbf{y_1}, \mathbf{y_2}, \ldots \mathbf{y_m}) \in \mathbb{R}^{d\times m}$ (where $\mathbf{x_i}$, $\mathbf{y_i}$ are d-dimensional vectors), CCA finds a canonical coordinate space that maximizes correlations between the projections of the datasets on to that space. For each dimension of this coordinate space, there is a pair of projection weight vectors, $\mathbf{a_j} = (a_{1j}, a_{2j}, \ldots a_{nj})$ and $\mathbf{b_j} = (b_{1j}, b_{2j}, \ldots b_{mj})$ called *canonical weights*. The resulting projections of datasets **X** and **Y** onto each dimension of the canonical space are a pair of d-dimensional vectors, $\mathbf{u_j} = \langle \mathbf{a_j}, \mathbf{X_i} \rangle$ and $\mathbf{v_j} = \langle \mathbf{b_j}, \mathbf{Y_i} \rangle$, that are called *canonical components* or *canonical variates*. (Bilenko & Gallant, 2016, p. 2)

Conceptually what CCA does is correlate a principal components of variation in X with principal components of Y, and it solves for PCs in the two data sets simultaneously in order to maximize the correlations between the “projections” of the two datasets. This is called the canonical coordinate space.

Bilenko & Gallant (2016) commented on the tendency of CCA to “overfit to spurious noise correlations between datasets,” and point out that “overfitting can be avoided by curbing the size of the canonical coordinate space, by regularization, or both” (p. 2). In fitting CCA to the z-scored data in this study, I used cross-validation to select the number of canonical correlations to report, thus curbing the size of the canonical coordinate space, and to select a regularization constant (which constrains the weight vectors **a** and **b**). Given the relatively small size of the datasets (only d=48 observations in each), the cross-validation procedure was repeated 100 times and the most frequently selected analysis parameters were used in the analysis here. Thus, the regularization parameter was set to 0.04, and only the first 2 canonical correlations were kept. The canonical correlations between the anatomical canonical component scores <**u**>, and the phonetic canonical component scores <**v**> were 0.81 for the first dimension and 0.73 for the second dimension.

The weight vectors **a** and **b** for the canonical coordinate space with 2 dimensions (CC1 and CC2) are illustrated in figure 8. The first canonical correlation (CC1) has positive weights for vocal tract length (vtl) and length of the pharynx (phar_length), and to a lesser extent also positive weights on palate height and mean area under the palate. CC2 also has positive weights for vocal tract length and pharynx length, but has large negative weights for palate height, mean palate area, and total palate area. So, a talker with a large positive CC1 component <**u**> and a CC2 component close to zero (illustrated by talker JW41 in figures 9 & 10) will have a long vocal tract and a deep palate, while a talker with a large positive CC2 component and a CC1 component close to zero (illustrated by talker JW51) will have a long vocal tract but a relatively shallow palate.

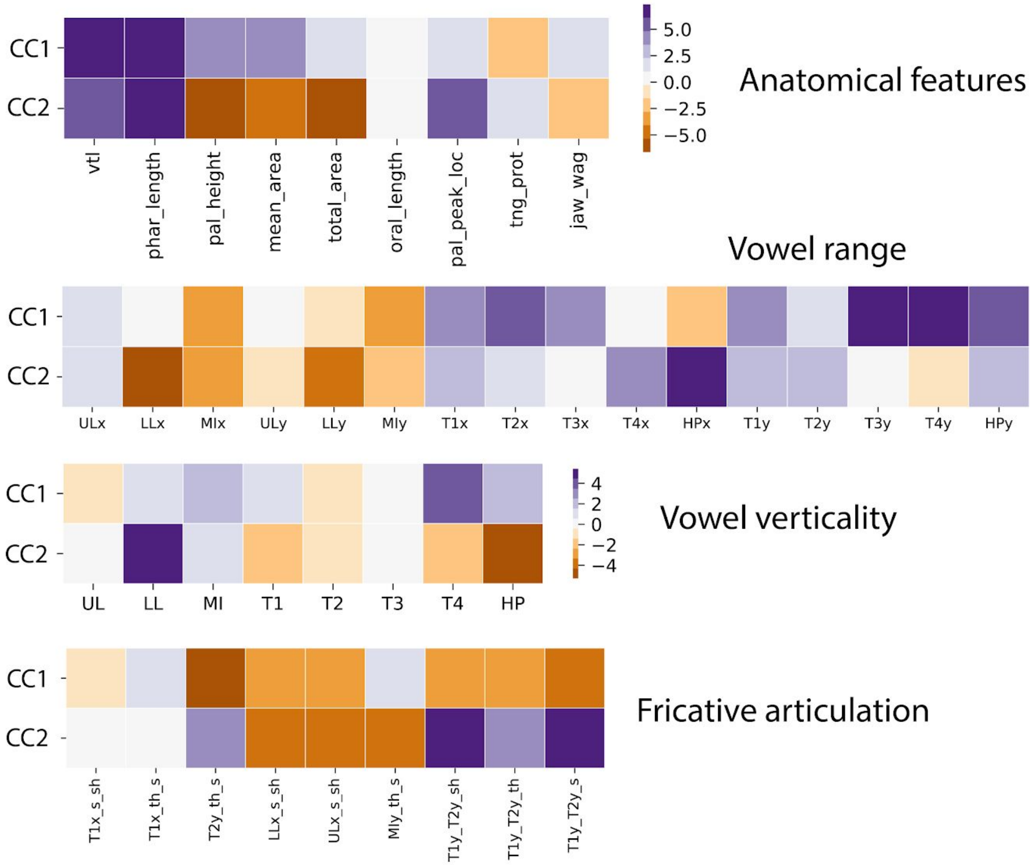

*FIG 8. The canonical weights for the two canonical correlations (CC1 and CC2) are illustrated by a diverging color map from red to blue. The nine <**a**> weights are in the first panel (Anatomical features) and the thirty-three <**b**> weights are shown in the lower three panels.*

Note that for figures 9 & 10, I selected four talkers based on their scores in the anatomical canonical components in the <**u**> matrix - to have the largest magnitude positive or negative scores on one canonical dimension and the closest to zero on the other. Which is to say that the talkers were selected for their vocal tract geometry not for their articulatory patterns.

In vowel production, CC1 has positive weights (figure 8) for tongue body vertical range (T4y, HPy), and for tongue tip horizontal range (T1x, T2x). There is also a negative weight for jaw motion (MIx, MIy). CC1 also has a positive weight for pellet T4 verticality. So talkers (like JW41) who score positively on CC1 will have a more vertically oriented T4 distribution, while talkers (like JW30 in figures 9 & 10) who have a negative score on CC1 will have a more horizontal orientation of the T4 cloud of observations, and a greater range of jaw positions during vowels. These patterns for vowel production are illustrated in the top row of figure 9.

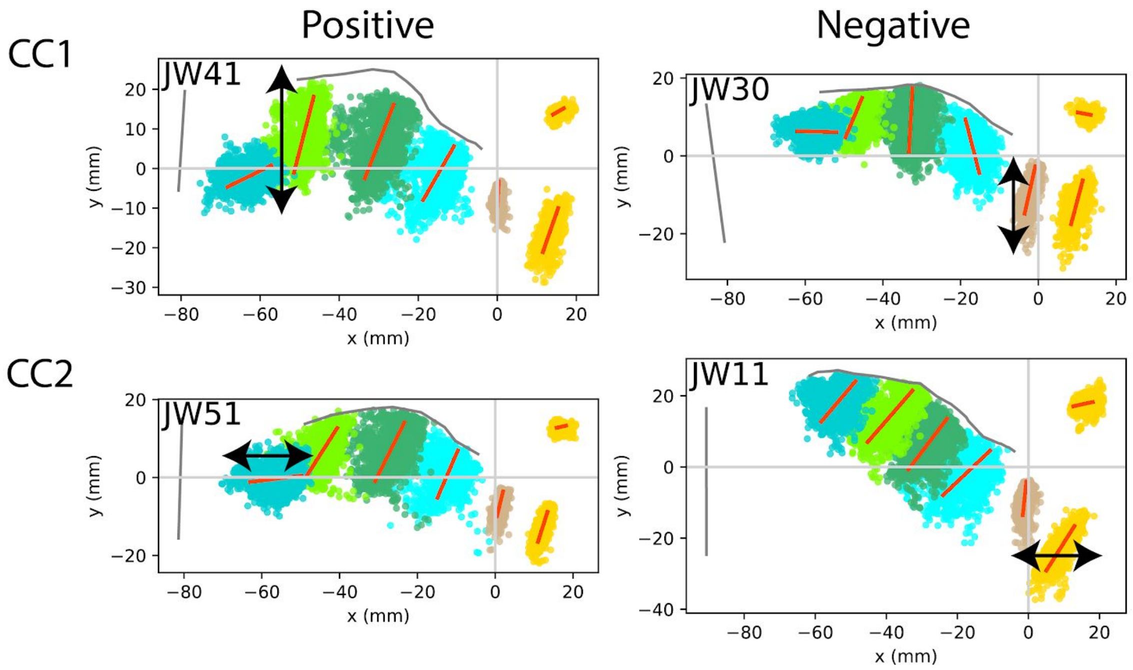

FIG 9. Articulatory vowel spaces for talkers who score highly (positive or negative) on one of the canonical correlations, while scoring close to zero on the other. The arrows illustrate the main vowel articulatory patterns associated with the CCs.

In vowels, CC2 has positive weight for HPx - the range of horizontal location of the highest point of the tongue. Talkers (like JW51) who score positively on CC2 showed a larger horizontal range of HP variation in vowels than talkers who have a negative score on CC2 (like JW11). CC2 had negative weight for lower lip horizontal range (LLx) in vowels (and thus positively weighted for LL verticality). Talkers who scored negatively on CC2 (illustrated by talker JW11) tended to have a wider range of LLx values in their vowels.

When it comes to fricatives, CC1 has positive weight for dental blade lowering in [θ] (relative to the posture for [s]) and for a tendency for a tip-up [s]. These patterns are not well illustrated well in the talkers selected for figure 10, and indicate the general lower weight magnitude for fricative dimensions for CC1 compared with CC2. The CC2 pattern in fricatives is related to tongue tip posture - a positive score on CC2 (JW51) was associated with a more raised tip posture for each of the coronal fricatives. The weights are negative because the feature is calculated as T1y-T2y.

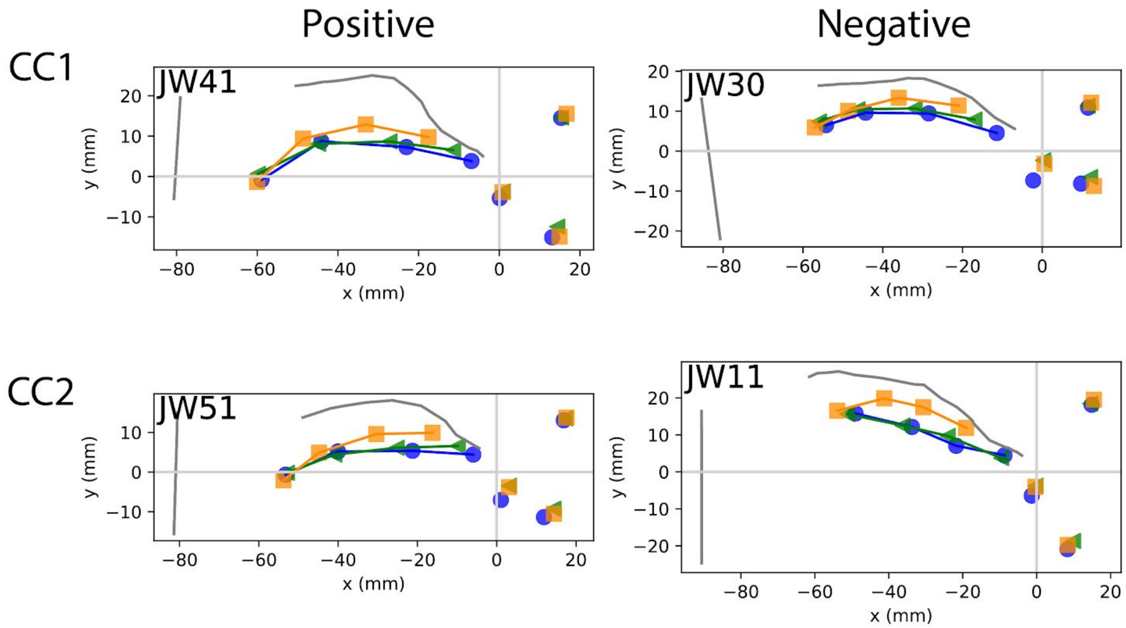

FIG 10. Average pellet locations at fricative midpoint [green = ʃ, yellow = s, blue = θ] for talkers who score highly (positive or negative) on one of Canonical correlations, while scoring close to zero on the other CC. The arrows illustrate the main vowel articulatory patterns associated with the CCs.

In summary, canonical correlation analysis found two reliable factors relating vocal tract anatomy with articulatory patterns in vowel and coronal fricative contrasts. The first dimension of the canonical correlation space contrasted big vocal tracts (long vocal tracts with deep palates) with small, and was primarily associated with tongue vertical range in vowels. This dimension is reminiscent of the parameter identified by Johnson, Ladefoged, & Lindau (1993) distinguishing ‘tongue movers’ from ‘jaw movers’. The second dimension of the CCA highlighted palate doming that is anti-correlated with vocal tract length. In vowels, this dimension primarily distinguished talkers who used tongue body horizontal movement versus talkers who used lip rounding, while in fricatives it was associated with tongue tip posture. Talkers with shallow palates had more tongue horizontal range (and less lip rounding) contrasting vowels, and a more tongue-tip up posture in coronal fricatives.

IV Conclusion

This is the first study in the literature to combine several measures of vocal tract anatomy with several measures of speech articulation. The study has focussed on between-category variation, looking at how segmental contrasts are produced. Future research can use the same data set to examine within-category variation, as has been done in several previous studies.

This study confirms several findings from previous research. (1) Men and women have different length vocal tracts. This difference is more pronounced in the length of the pharynx than in the length of the oral cavity. (2) Vocal tract length is associated with the range and orientation of tongue positions in making vowel contrasts, the longer the vocal tract the larger the vertical range of tongue body motion, and the larger the tongue blade horizontal range of motion. Talkers who have a longer pharynx/vocal tract tend to have a more vertical orientation of tongue body motion. (3) Palate height is also positively correlated with vocal tract length and with greater vertical range of tongue body motion, and vertical orientation of tongue body. (4) Individual differences in fricative production are tied to palate shape more than to overall vocal tract length.

What kind of theory is suggested by these findings? We will close with three observations.

First, in this study, the effect of anatomical constraints was not deterministic. Though we found correlations between anatomical characteristics and speech production patterns, the relationship is partial. The highest canonical correlation seen here (r=0.81, for CC1) means that, for this dimension of the canonical correlation, only 66% of the phonetic variance can be attributed to anatomical variation. One caveat to this conclusion is that it may be possible to find stronger relationships if we have better anatomical descriptions. Perhaps with more detail of the 3D vocal tract, rather than only sagittal dimensions, or perhaps with more information about the lower vocal tract - dimensions of the pharynx, epilarynx, or piriform sinus (Godoy, et al., 2015) -- better phonetic predictions might be possible.

The way that a gestural approach to speech communication must deal with between-talker variation is that we need to be able to say that all talkers are basically doing the same thing when they make the phonetic contrasts of their language, and that a part of perception is to normalize the differences in terms of vocal tract actions. But, as it stands now, it would be difficult to devise an articulatory normalization algorithm that would remove between-talker articulatory variability based on some inferences about vocal tract geometry because much of the between-talker variability is not predictable from knowledge about vocal tract geometry.

Second, some of the phonetic differences found here are not necessarily problematic for an articulatory normalization theory. For example, the first canonical correlation -- that found a basic difference between ‘jaw movers’ and ‘tongue movers’ -- is not a huge complication for a theory based on vocal tract constriction actions (Goldstein & Fowler, 2003). Talkers may have different patterns of coordination among “model articulators” (Saltzman & Munhall, 1989) while still performing the same constriction actions. So, whether the jaw is more or less involved in moving the tongue around in the mouth is not a between-talker articulatory difference that necessarily matters much for a theory based on constriction actions. The fact that the inter-articulator coordination of tongue and jaw is partly attributable to vocal tract geometry is interesting, and may be relevant for considerations of the visual phonetic information available to listeners/viewers, or perhaps (more speculatively) for the possible patterns of phonetic coarticulation, but this is not a theoretical problem for a gesture-based model because the same constriction can be accomplished with different mixtures of jaw and tongue involvement in the constriction forming action.

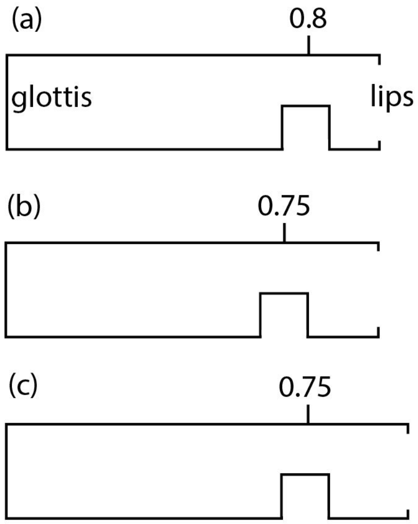

Fig. 11. Two articulatory moves starting from the configuration in (a) with a constriction at 0.8 (glottis = 0, and lips = 1) that can result in a tongue constriction at 0.75. One, panel (b), is to move the tongue constriction back (see JW51 in Figure 9). The other, panel (c), is to lengthen the vocal tract with lip protrusion (see JW11 in Figure 9).

Third, the second canonical dimension found in this study trades lip horizontal movement for tongue body horizontal movement, as illustrated in figure 11. This trading relationship is of a different type than jaw movers versus tongue movers. If we wanted to, we could describe the similarity between the configurations in (b) and (c) in articulatory terms - the constriction is located at 75% of the length of the tube. However, the constriction action that starts in Fig 11(a) and moves to (b) involves a different 'tract variable' (Saltzman & Munhall, 1989) than the action that gives (c). What the different constriction actions have in common is that they produce a constriction at about the same location in the standing waves of the vocal tract resonances - an acoustic description of what the moves have in common.

One can imagine a similar type of 'outcome' equivalence in how a person may press a doorbell. They could press with any finger of their right hand, or left hand; or if carrying a package, press with an elbow or knee. The equivalence of these gestures is in the outcome that they produce. Trading relations such as the one in figure 11, and in the data in figure 9, suggests that one of the main ways that talkers differ in how they produce vowels is that they deploy different patterns or combinations of constriction actions to produce comparable acoustic outcomes. This supports the auditory parity hypothesis.

Acknowledgements. Many thanks to the phonetic tagging team: Sofea Dil, Eric Chen, and Steven Ho. Thanks also to Matt Faytak for comments and suggestions on an earlier version of this paper.

References

Alfwaress, F.; Al Maaitah, E.; Al-Khateeb, S.; & Zama, Z.U. (2015). "The relationship of vocal tract dimensions and substitution of the palatal approximant /j/ for the alveolar trill /r/," *Int. J. of Sp.-Lang. Pathology.* **17**(5), 518-526.

Bakst, S. (2016). Differences in the relationship between palate shape, articulation, and acoustics of American English /r/ and /s/. *UC Berkeley PhonLab Annual Report*, 12. Retrieved from https://escholarship.org/uc/item/0sj6b0zp

Bakst, Sarah; Johnson, Keith. (2018) Modeling the effect of palate shape on the articulatory-acoustics mapping, *JASA Express Letters* **EL71**

Bakst, S. & Lin, S. (2015) An ultrasound investigation into articulatory variation in American /r/ and /s/. *Proceedings of the 18th International Congress of Phonetic Sciences*. Paper number 0696.

Bilenko, N.Y. & Gallant, J.L. (2016) Pyrcca: Regularized Kernel Canonical Correlation in Python and its applications to neuroimaging. *Frontiers in Neuroinformatics*. **10**, 49, doi: 10.3389/fninf.2016.00049

Boersma, P. & Weenink, D. (2018). "Praat: doing phonetics by computer [Computer program]," Version 6.0.40, retrieved 11 May 2018 from http://www.praat.org/

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

Edwards, J. & Harris, K.S. (1990) "Rotation and translation of the jaw during speech," *J. Speech and Hearing Res.* **33**(3), 550-562.

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

Susanne Fuchs, Ralf Winkler, Pascal Perrier. (2008) Do Speakers' Vocal Tract Geometries Shape their Articulatory Vowel Space?. *8th International Seminar on Speech Production*, ISSP'08, Dec 2008, Strasbourg, France. pp.333-336.

Garrett, Andrew & Keith Johnson (2012). Phonetic bias in sound change. In Alan Yu (ed.), *Origins of sound change: Approaches to phonologization*. Oxford University Press. pp. 51-97

Godoy, E.; Malyska, N; & Quatieri, T.F. (2015) Estimating lower vocal tract features with closed-open phase spectral analyses. *INTERSPEECH-2015*, 771-775.

Goldstein, L. & Fowler, C.A. (2003) Articulatory Phonology: A phonology for public language use. In Schiller, N.O. & Meyer, A.S. (eds.) *Phonetics and Phonology in Language Comprehension and Production*, pp. 159-207. Mouton de Gruyter.

Goldstein, U. (1980). "An articulatory model for the vocal tracts of growing children," Ph.D. Thesis, MIT, Cambridge, MA.

Honda, K.; Maeda, S.; Hashi, M.; Dembowski, J.S. & Westbury, J.R. (1996) "Human palate and related structures: Their articulatory consequences," *Proc. ICSLP* **2**, 784-787.

Houde, J.F. & Nagarajan, S.S. (2011) Speech production as state feedback control. *Frontiers in Human Neuroscience* **5**. doi: 10.3389/fnhum.2011.00082

Johnson, K., Ladefoged, P. & Lindau, M. (1993). “Individual differences in vowel production,” *J. Acoust. Soc. of Am.* **94**, 701-714.

Kuehn, D. P., & Moll, K. L. (1976). “A cineradiographic study of VC and CV articulatory velocities,” *J. Phonetics, **4***(4), 303-320.

Laine, T. (1986). “Articulatory disorders in speech as related to the size of alveolar arches,” *European Journal of Orthodontics*, **8**, 192-197.

Lammert, A.C. & Narayanan, S.S. (2015) On short-time estimation of vocal tract length from formant frequencies. *PLoS ONE* **10**(7), e0132193. https://doi.org/10.1371/journal.pone.0132193

Lammert, A.C.; Proctor, M.; Katsamanis, A. & Narayanan, S. (2011) Morphological variation in the adult vocal tract: A modelling study of its potential acoustic impact. *Interspeech 2011*, 2813-2816.

Lee, Chin-Hui. (1988) “On robust linear prediction of speech,” *IEEE Transactions on Acoustics, Speech and Signal Processing* **36**, 642-650.

Markel, J.D. & Gray, Jr., A.H. (1976). *Linear Prediction of Speech.* Springer Verlag, Berlin.

Nolan, F.; Holst, T.; & Kühnert, B. (1996). “Modelling [s] to [ʃ] accommodation in English,” *J. Phonetics*, **24**, 113-137.

Ohala, J.J. (1993) The phonetics of sound change. In Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives*. London: Longman. pp. 237-278

Perkell, J.S. (1997) Articulatory processes. In Hardcastle, W.J. & Laver, J. (eds) *Handbook of Phonetic Sciences*, Oxford: Blackwell Publishers, pp. 333–370.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. & Zandipour, M. (2000) A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *J. Phonetics*, **28**, 233-272.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1995). “Goal-based speech motor control: A theoretical framework and some preliminary data,” *J. Phonetics*, **23**, 23–35.

Pisoni, D.B. & Sawusch, J.R. (1975) Some stages of processing in speech perception. In Cohen, A. & Nooteboom, S.G. (eds.) *Structure and Process in Speech Perception*, pp. 16-34, NY: Springer-Verlag.

Rudy, K. & Yanusova, Y. (2013) The effect of anatomic factors on tongue position variability during consonants. *J. Speech, Lang. Hearing Res*. **56**, 137-149.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333-382. http://dx.doi.org/10.1207/s15326969eco0104_2

Simpson, A.P. (2001) “Dynamic consequences of difference in male and female vocal tract dimensions,” *J. Acoust. Soc. of Am.* 109(5), 2153-2164.

Simpson, A.P. (2002) “Gender-specific articulatory-acoustic relations in vowel sequences,” *J. Phonetics, **30***, 417-435.

Talkin, D. (1987) “Speech formant trajectory estimation using dynamic programming with modulated transition costs,” *J. Acoust. Soc. of Am.* 82(S1): S55.

Watanabe, A. (2001). Formant estimation method using inverse-filter control. *IEEE Transactions on Speech and Audio Processing, 9,* 317-326.

Weirich, M. & Fuchs, S. (2013) Palate morphology can influence speaker-specific realizations of phonemic contrasts. J. Speech Lang Hear Res. 56(6), S1894-908. doi: 10.1044/1092-4388(2013/12-0217).

Westbury, J.R. (1994) *X-ray Microbeam Speech Production Database User's Handbook, Version 1.0*. Madison, WI.

Winkler, R.; Fuchs, S. & Perrier, P. (2006) The relation between differences in vocal tract geometry and articulatory control strategies in the production of French vowels: Evidence from MRI and modeling. *7th International Seminar on Speech Production ISSP* Ubatuba, 509-516

Winkler, R.; Fuchs, S.; Perrier, P.; Mark Tiede, M. (2011). "Biomechanical Tongue Models: An Approach to Studying Inter-speaker Variability," *12th Annual Conference of the International Speech Communication Association (Interspeech 2011), Aug 2011, Florence, Italy. Proceedings of Interspeech*, pp.273-276, 2011. <hal-00619238>

Yuan, J. & Liberman, M. (2008). Speaker identification in the SCOTUS corpus. *Proceedings of Acoustics 2008,* 5687-5690.