INDIVIDUAL DIFFERENCES IN SPEECH PRODUCTION: WHAT IS "PHONETIC SUBSTANCE"?

Keith Johnson

University of California, Berkeley
keithjohnson@berkeley.edu

ABSTRACT

This study explored the talker as a source of phonetic variation. Two patterns emerged in a canonical correlation study of the Wisconsin X-ray Microbeam Database relating individual differences in vocal tract anatomy to differences in the phonetic realization of vowels and consonants. Talkers with a large vocal tracts tended to show large tongue vertical range of motion in vowel production, while those with smaller vocal tracts showed larger jaw vertical range of motion. Talkers who have a more deeply domed palate show more horizontal lower-lip range of motion in vowels, and tend to have more tip down posture in coronal fricatives, while talkers with less deeply domed palate showed larger range of tongue horizontal position in vowels and more tip up fricatives.

Keywords: Speech articulation, vocal tract anatomy, sound change, individual differences.

1. INTRODUCTION

Many phonological patterns emerge historically from phonetically motivated, natural sound changes which are based on a "pool of synchronic phonetic variation" [1, 2]. This study explored one source of phonetic variation, individual differences in vocal tract anatomy, with a focus on the correlated phonetic bias that may result from anatomy.

2. BACKGROUND

The relationship between individual differences in vocal tract anatomy and articulatory phonetic variation has been studied many times [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. For example, the depth of the palate vault has been related to the amount of lip/tongue trading in /u/ [10], the amount of coarticulation in front consonants [11], the articulatory variability of /s/ and /r/ [3, 4], the amount of jaw or tongue movement in a vowel height distinction [8], and the amount of articulatory difference between /s/ and /ʃ/ [13]. Similarly, the length of the oral cavity has been found to relate to the verticality of high vowel variation [15], and the verticality of front consonant coarticulation [11], while the overall length of the vocal tract has been related to speed and magnitude of articulatory movements [9, 12, 6], and to the verticality of vowel [a] to [i] movements [7].

Many of these studies rely on a small number of speakers, and most of them focus on only one or two anatomical or articulatory features. The present study contributes to this literature, and our overall understanding of the pool of phonetic variation, by looking at several dimensions of consonant and vowel articulation and several anatomical features in a mutually constraining statistical model. Using canonical correlation [16, 17], we jointly constrain the search for anatomical and articulatory patterns that are related to each other.

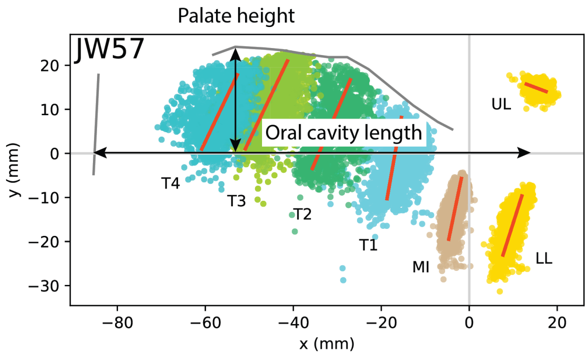

Figure 1: Example data, showing about 1500 stressed vowel midpoint pellet locations for talker JW57. The anatomical measures "palate height" adn "oral cavity length" are illustrated.

3. METHODS

3.1. Dataset

The Wisconsin X-ray Microbeam Database [18, 19] was used in this study. The database (for which textgrids are now available [19]) contains audio, articulatory point tracking data, and time aligned word and phone tags, for 48 dialectically homogeneous talkers reading word lists, sentences, and short stories. Additionally, a palate-trace was produced for each talker and the approximate location of the pharyngeal wall was determined. Talkers also were recorded in jaw-wagging and tongue protrusion trials. The articulatory data are the x and y locations of gold pellets glued onto the upper lip (UL), lower lip (LL), lower incisor (MI), and four pellets on the tongue (T1-4) in the mid-sagittal plane (see Fig. 1). These were recorded at a sampling rate of 145.64 Hz, and the data were rotated to a standard occlusal coordinate system, with (0,0) at the tip of the upper incisors.

3.2. Anatomical Measures

Table 1 lists the nine measures of vocal tract anatomy that were taken for each talker. Vocal tract length was calculated from vowel formants [20], and pharynx length was VTL minus oral tract length (Fig. 1). Maximum tongue protrusion and maximum jaw opening were measured from recordings in which talkers were asked to perform these motions. The other measurements were taken from the palate trace.

Anatomical measures
Vocal tract length
Oral tract length
Pharyngeal cavity length
Depth of the palate dome
Area under the palate
Area relative to length
Location of palate peak
Maximum tongue protrusion
Maximum jaw opening

Table 1: Anatomical measures.

3.3. Articulatory Parameters

Table 2 lists the articulatory parameters that were measured. The first three are measures of the articulatory vowel space. The input for these measures is a collection of data from the midpoints of all stressed vowels produced by a speaker (see Fig. 1). For each pellet the horizontal and vertical ranges (at the 10th and 90th percentiles) of the cloud of data points were recorded. Thus, there were seven horizontal range measures (one for each pellet) and seven vertical range measures. A measure of the verticality of the data cloud for each pellet was taken from the orientation of the first principal component of variation - the lines in each cloud in Fig. 1. The remaining measures listed in the table were taken from fricative midpoints, grouped by place of articulation into dental [θ, ð], alveolar [s,z], and post-alveolar [ʃ, ʒ , ʧ, ʤ]. Tip posture was measured as the mean difference between T1y and T2y for each place of articulation. The remaining measures of retraction, rounding, fronting etc. were taken by comparing the mean posture in dentals or post-alveolars to that in the alveolar fricatives. For example, the θ blade lowering score for a talker is the difference between the mean vertical location of the T2 pellet in their [s,z] and the mean vertical location in their [θ,ð]. There were 33 articulatory parameters in all.

Articulatory parameters
Horizontal range in vowels
Vertical range in vowels
Verticality of vowel variation
Tip posture
/ʃ/ retraction
/ʃ/ rounding
/θ/ fronting
/θ/ blade lowering
/θ/ jaw opening

Table 2: Articulatory parameters.

3.4. Statistical Analysis

Conceptually, canonical correlation analysis (CCA) finds principal components of variation in one matrix (X, which will be our anatomical measures) that are maximally correlated with principal components of another matrix (Y, which will be our articulatory parameters), and it solves for PCs in the two matrices simultaneously in order to maximize the correlations between patterns of variation in X with patterns of variation in Y.

(1)

$$\begin{pmatrix} 0 & C_{XY} \\ C_{YX} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \rho^2 \begin{pmatrix} C_{XX}+\lambda I & 0 \\ 0 & C_{YY}+\lambda I \end{pmatrix}$$

In practice (Eq. 1, [17]) this is solved as an eigenvalue problem with the covariance matrices for X and Y, C_{XY} and C_{YX}, and their autocorrelation matrices, C_{XX} and C_{YY}. The weight matrix **a** defines a pattern in X that is correlated with a pattern in Y which is defined by the weight matrix **b**. The most highly correlated pattern (i.e. the highest value of ρ) is the first canonical correlation (ρ_{CC1}) and it is achieved with canonical weights $\mathbf{a}_{CC1}$ and $\mathbf{b}_{CC1}$.

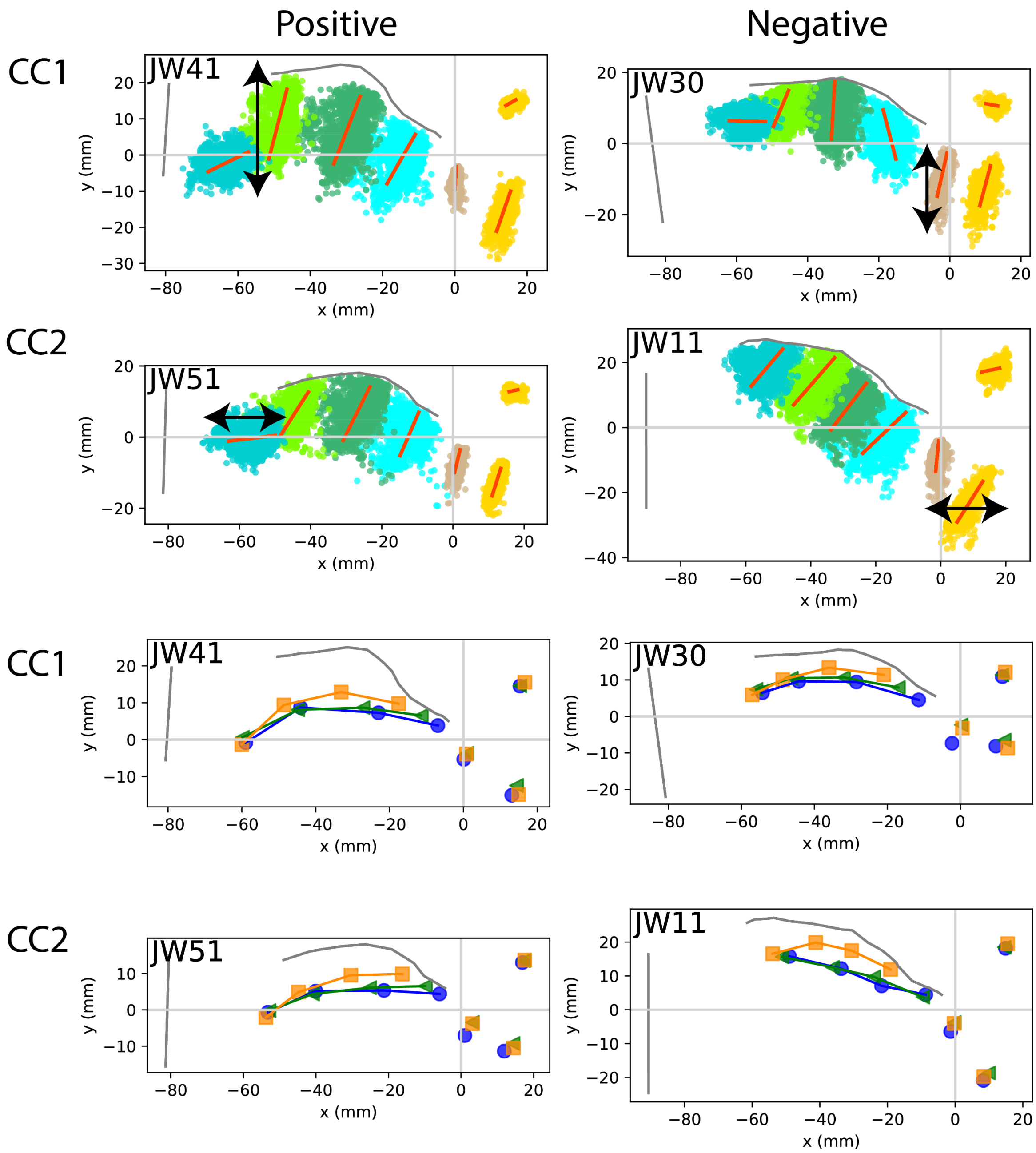

Figure 2: Talkers exemplifying the articulatory patterns discovered in the canonical correlation analysis. The **top row** shows the pattern of vowel articulation that was captured by the first canonical correlation (CC1). Talker JW41 illustrates the pattern for a talker who had a positive score on CC1 and talker JW30 illustrates the pattern for a negative score on CC1. The **second row** illustrates the patterns of vowel variation that were captured by the second canonical correlation (talker JW51 had a positive score on CC2 and talker JW11 had a negative score for CC2). The clouds of points show the midpoint pellet locations for the stressed vowels produced by the talker. Lines drawn in each cloud of points shows the orientation of the first principal component of variation of that cloud. The **bottom four panels** illustrate the patterns found for coronal fricatives for the same example talkers. The orange squares show the average pellet locations for post-alveolar fricatives, the green triangles for alveolar fricatives, and the blue circles for dental fricatives.

In this study, two canonical correlation patterns (CC1 and CC2) between speech anatomy and speech articulation were found. These are patterns in vowel production and coronal fricative production that are correlated with aspects of vocal anatomy.

Imposing L2 regularization (with λI in Eq. 1) constrains the norms of canonical weights **a** and **b** which stabilizes the solution when there are many variables in the matrices X and Y.

This study used a cross-validation procedure which was repeated 100 times and then the most frequently selected analysis parameters were used in the final analysis reported here. Based on this procedure, the regularization parameter λ was set to 0.04, and only the first 2 canonical correlations were kept. The canonical correlations between the anatomical variables and the articulatory parameters were $\rho_{CC1} = 0.81$ for the first component, and $\rho_{CC2} = 0.73$ for the second component.

4. RESULTS

Fig. 2 illustrates the main findings of this study. The first canonical correlation (CC1) was most highly weighted for vocal tract length, and pharyngeal cavity length in the anatomical measures. Among the vowel articulatory features, CC1 had positive weight for tongue body vertical range, and had negative weight for jaw x and y range. This is illustrated by the top row of Fig. 2 by talkers who had large positive or negative scores on CC1. Talker JW41 who has a long vocal tract and a large vertical vowel space for pellets T2 and T3, is compared to talker JW30 who has a shorter vocal tract and more jaw range in vowels than tongue body range. Talkers like these were called "jaw movers" and "tongue movers" in [8]. In coronal consonants CC1 was associated with TH_blade lowering.

The second canonical correlation (CC2) was related to palate doming independent of VTL. In vowels, this component linked shallow palate with larger horizontal range of tongue positions, while deeper palate was association in CC2 with larger horizontal range of lower lip positions. This is illustrated by talkers JW51 and JW11 in the second row in Fig. 2. CC2 also captured variation in coronal consonant articulation; shallow palate was associated with tip up posture, while deep palate was associated with tip down posture for all three places of articulation.

5. CONCLUSION

This analysis of the Wisconsin X-ray Microbeam Database found that the pool of phonetic variation is tied to talker anatomical differences; talkers contribute unique patterns of phonetic variation (and covariation across segments), resulting in talker specific phonetic substance. This means, for example, that patterns of coarticulation are probably talker-specific because the articulatory movements in producing vowel or consonant contrasts differ in terms of which articulator makes the larger movement. Similarly, trading relations between articulatory gestures in producing phonetic contrast appear from these data to be talker-specific. In general, because perfect articulatory imitation may not be possible, we have to conclude that phonetic knowledge is necessarily somewhat abstract.

6. REFERENCES

[1] J. Ohala, "The phonetics of sound change," in *Historical Linguistics: Problems and Perspectives*, C. Jones, Ed. Logman, 1993, pp. 237–278.

[2] A. Garrett and K. Johnson, "Phonetic bias in sound change," in *Origins of Sound Change: Approaches to Phonologization*, A. Yu, Ed. Oxford, 2012.

[3] S. Bakst and S. Lin, "An ultrasound investigation into articulatory variation in American English /r/ and /s/," in *Proc. 18th ICPhS*, Glasgow, 2003.

[4] S. Bakst, "Differences in the relationship between palate shape, articulation, and acoustics of American English /r/ and /s/," *UC Berkeley PhonLab Annual Report*, no. 12, 2016.

[5] J. Brunner, S. Fuchs, and P. Perrier, "On the relationship between palate shape and articulatory behavior," *J. Acoust. Soc. Am.*, vol. 125, pp. 3936–3949, 2009.

[6] S. Fuchs, R. Winkler, and P. Perrier, "Do speakers' vocal tract geometries shape their articulatory vowel space?" in *8th International Seminar on Speech Production, ISSP'08*, Strasbourg, France, 2008, pp. 333–336.

[7] K. Honda, S. Maeda, M. Hashi, J. Dembowski, and J. Westbury, "Human palate and related structures: Their articulatory consequences," in *Proc. ICSLP 2*, 1996, pp. 784–787.

[8] K. Johnson, P. Ladefoged, and M. Lindau, "Individual differences in vowel production," *J. Acoust. Soc. Am.*, vol. 94, pp. 701–714, 1993.

[9] D. Kuehen and K. Moll, "A cineradiographic study of VC and CV articulatory velocities," *J. Phonetics*, vol. 4, pp. 303–320, 1976.

[10] J. Perkell, M. Matthies, M. Svirsky, and M. Jordan, "Goal-based speech motor control: A theoretical framework and some preliminary data," *J. Phonetics*, vol. 23, pp. 23–35, 1995.

[11] K. Rudy and Y. Yanusova, "The effect of anatomic factors on tongue position variability during consonants," *J. Speech, Lang. Hearing Res.*, vol. 56, pp. 137–149, 2013.

[12] A. Simpson, "Dynamic consequences of difference in male and female vocal tract dimensions," *J.*

Acoust. Soc. Am., vol. 109, pp. 2153–2164, 2001.
[13] M. Weirich and S. Fuchs, “Palate morphology can influence speaker-specific realizations of phonemic contrasts,” *J. Speech, Lang. Hearing Res.*, vol. 56, pp. 894–908, 2013.
[14] R. Winkler, S. Fuchs, and P. Perrier, “The relation between differences in vocal tract geometry and articulatory control strategies in the production of French vowels: Evidence from MRI and modeling,” in *7th International Seminar on Speech Production ISSP*, Ubatuba, 2006, pp. 509–516.
[15] R. Winkler, S. Fuchs, P. Perrier, and M. Tiede, “Biomechanical tongue models: An approach to studying inter-speaker variability,” in *Proc. 12th Annual Conference of the International Speech Communication Association (Interspeech 2011)*, Florence, Italy, 2011, pp. 273–276.
[16] H. Hotelling, “Relations between two sets of variates,” *Biometrika*, vol. 28, pp. 321–377, 1936.
[17] N. Bilenko and J. Gallant, “Pyrcca: Regularized kernel canonical correlation in python and its applications to neuroimaging,” *Frontiers in Neuro-informatics*, vol. 10, p. 49, 2016.
[18] J. Westbury, “X-ray Microbeam Speech Production Database User’s Handbook, version 1.0,” unpublished manuscript, Madison, WI.
[19] Wisconsin X-ray Microbeam Database. [Online]. Available: https://github.com/rsprouse/xray_microbeam_database
[20] A. Lammert and S. Nrayanan, “On short-time estimation of vocal tract length from formant frequencies,” *PLoS ONE*, vol. 10, 2015.