Published in final edited form as:
Comput Methods Biomech Biomed Engin. 2010 ; 13(4): 493–503. doi:10.1080/10255842.2010.484809.

A preliminary application of principal components and cluster analysis to internal tongue deformation patterns

Maureen Stone[a,*], **Xiaofeng Liu**[b], **Hegang Chen**[c], and **Jerry L. Prince**[d,b]

[a] Department of Neural and Pain Sciences, Dept of Orthodontics, University of Maryland Dental School, Baltimore, MD, USA

[b] Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

[c] Department of Epidemiology and Preventive Medicine, University of Maryland Medical School, Baltimore, MD, USA

[d] Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

Abstract

Complex patterns of muscle contractions create gross tongue motion during speech. It is of scientific and medical importance to better understand speech motor strategies and variations due to language or disorders. Dense patterns of tongue motion can be imaged using tagged MRI, but characterization of motion strategies is difficult using visualization alone. This paper explores the use of principal component analysis for dimensionality reduction and cluster analysis for tongue motion categorization. Velocity fields were acquired and analyzed from midsagittal tongue slices during motion from /i/ to /u/ for 8 data sets containing multiple languages and a glossectomy patient. The analyses were carried out on tongue-only and tongue-plus-floor of mouth regions. Results showed that both analyses were sensitive to region size, and that cluster analysis was harder to interpret. Both analyses grouped the Japanese speaker with the glossectomy patient which although explicable with biologically plausible reasons, highlights the limitations of extensive data reduction.

Keywords

tongue; Principal Components Analysis; Cluster Analysis; MRI; tags

1. Introduction and Background

Speech creation and intelligibility are highly dependent on tongue deformation because change in the shape of the vocal tract is caused primarily by movement of the tongue. Tongue deformations may appear to be nearly unlimited in number when one considers the variety and quantity of speech sounds that appear across the world's languages. In addition, studies of motor equivalence and inverse models of the vocal tract indicate that many vocal tract shapes can produce similar speech spectra. The classic example of this is the English sound /r/, which uses several different tongue surface shapes, yet produces identical percepts and highly similar waveforms. Despite this variety of shapes, the majority of individual sounds within a language appear to be produced with fairly specific tongue and vocal tract shapes. The question arises, therefore, as to whether multiple speakers (or one speaker at

*Corresponding author. mstone@umaryland.edu.

different times) produce the same tongue surface shape by using essentially the same motor control strategy, or whether speakers can use entirely different, motor equivalent muscle activity patterns. The former case would allow for minor muscular variation meant to accommodate individual differences, such as oral cavity size and shape. The latter case would increase the complexity of speech motor control, but would provide more opportunities for production strategies when dealing with coarticulation, learning a new language, or compensating for changes in oral morphology due to surgical, medical, or dental procedures.

The ideal way to determine the extent of between-subject motor differences would be to directly measure all the tongue muscles while speaking using electromyography (EMG). However, the muscle fibres of the tongue are interdigitated, which makes EMG of most tongue muscles an extremely challenging, if not impossible, task. Therefore motor control strategies must be studied in a more oblique manner. One approach is to compare patterns of tissue-point motion in the internal tongue among different speakers; these patterns can be extracted from velocity fields in tagged MRI images. Tissue-point motion is the behaviour that is intermediate between muscle activity and tongue surface shape. Determining commonalities in tissue-point motion patterns among subjects is the first step toward determining common features in muscle compression patterns and, ultimately, in control strategies used to create the same speech sound.

The present study compares eight subjects saying the concatenated vowels /i/-/u/. The speakers have several native languages and one has had surgery to remove part of the tongue due to cancer (partial glossectomy). Despite these demographics, which should increase the variety of patterns seen, all the subjects produced a normal sounding /i/-/u/, including the patient.

Imaging was carried out using tagged magnetic resonance imaging and processed using harmonic phase (HARP) method (Osman and Prince, 1999). Data analysis was carried out using principal components analysis and clustering. Magnetic resonance imaging (MRI) has been used for many years to capture both the shape of the oral cavity and vocal tract (Lufkin et al. 1987, McKenna et al. 1990, Engwall, 2003) and in imaging the motion of the tongue (Masaki et al. 1999, Shadle et al. 2006, Stone et al. 2002, Narayanan et al. 1997). Detailed motion of the pattern of muscle contraction within the body of the tongue was made possible through the advent of tagged MRI (Zerhouni et al. 1988, Axel et al. 1989) and has now been used in several studies of tongue motion (Niitsu et al. 1994, Napadow et al 1999a,b, Dick et al. 2000, Parthasarathy et al, 2007). The tagged MRI method called CSPAMM (Fischer et al. 1994) and its enhancement MICSR (NessAiver and Prince, 2003) are ideally suited for the HARP method (Osman and Prince, 1999), and provide the data that are used in the statistical analyses described in this paper. The mechanics of passive deformation due to contact with the hard palate, teeth and floor of mouth are not considered in the present paper. The patterns of motion alone are studied, irrespective of active vs passive origin. Future studies, which consider both boundary and muscle contributions, will provide more complete interpretations of tongue motion.

HARP is capable of generating a wide array of motion-related quantities (Osman and Prince, 2000a), including sequences of velocity fields that provide highly detailed (pixel-by-pixel) patterns of incremental motion as the tongue moves from one time frame to the next during speech. Velocity fields were extracted from the tagged MRI data at the time-frame with the largest overall observed tissue velocities in the tongue (hereafter, the target frame). The target frame, was always the first or second frame of the transition between the two sounds. The extracted velocity fields were compared among the eight data sets using Principal

Components Analysis and Cluster Analysis to examine differences and similarities among their internal motion patterns.

Because of the large quantity of data present, even with eight subjects, methods for simplifying and grouping the data were important. Principal Components Analysis (PCA) is an excellent standard method to extract and represent patterns in high-dimensional data for which no expectations or *a priori* models are available. PCA reduces the dimensionality of a data set by determining the main orthogonal directions of data variability and is typically applied to a data set after removing the common component, that is, the mean. A set of principal components (PCs) are then produced representing the most dominant variations (from the mean) that are present within the observed data. PCA (or its close relative factor analysis) has been used to characterize speech related motion of the midsagittal tongue surface (Harshman et al, 1977, Jackson, 1988, Hoole et al, 1999, Maeda, 1990) and the coronal tongue surface (Stone, et al., 1997, Slud et al, 2002). The speech of tongue cancer patients, pre and post glossectomy surgery, has also been characterized using PCA, and the results were able to distinguish tongue surface motions resulting from different reconstruction procedures (Bressman et al, 2004Bressman et al, 2007).

One question asked by the present study is whether the deformation within the midsagittal tongue proper is sufficient to represent its key motions, or whether a tongue-plus-floor of mouth (hereafter: tongue-plus-floor) region of interest (ROI) is needed. To answer this question several PCA's were performed. PCA 1 compared an ROI that included the tongue-plus-floor muscles for the eight data sets (see Figure 1a). PCA 2 was performed on a smaller region of interest, the tongue-only (see Figure 1b). Hold-one-out analyses were also performed for each of the individual subjects to see if the patient was represented more poorly by PC's derived from a group which excluded him than were the other subjects. We must also note that the patient has left-right asymmetries in his tongue motion due to loss of tissue on one side. Midsagittal motion may be a poorer representation of his 3D tongue motion than the other subjects.

Cluster analysis is a common technique for statistical data analysis used in many fields, including image analysis and bioinformatics. Clustering is the assignment of a group of subjects into subgroups (clusters) so that subjects within a subgroup are more similar (patterns) to one another than subjects in different subgroups. Hierarchical clustering algorithms are among the best known clustering methods (Duda et al., 2001). The algorithms can be divided according to two distinct approaches: agglomerative (bottom-up, clumping) and divisive (top-down, splitting). Agglomerative algorithms begin with each subject as a separate cluster and merge them into successively larger clusters. Divisive algorithms begin with the whole group and proceed to divide it into successively smaller clusters. The clustering method requires specification of both a similarity metric and a linkage. The similarity metric is defined for pairs of subjects, with the goal to group similar subjects together. *Euclidean distance*, *Manhattan distance*, *Mahalanobis distance* and *Pearson correlation* are the bases for some of the most common similarity metrics. Although the similarity metric reflects distance between two subjects, additional specifications of distance between clusters are required to define distance between two clusters. The specification of distance between clusters is determined by the linkage method. Average, complete and single linkages are the most commonly used ones. There are many applications of hierarchical clustering, for example, Alizadeh et al. (2000) discovered new subgroups of lymphomas. Similarly, Bittner et al. (2000) found structure among otherwise morphologically indistinguishable melanoma tumors.

In this study the subjects were clustered based on co-registered velocity vectors at a single instant of time for each ROI and their component motions. The agglomerative hierarchical

clustering algorithm was used for classifying subjects. The clustering trees were generated to explore the features that could be useful for categorization. With a larger data set, such analysis may reveal different velocity field patterns among American English (AE) speakers, non-native speakers, or patients. The second question asked by this study is whether the patient will be distinguished from the normal subjects.

2. Methods

2.1. Data used in the analyses: Subjects and speech material

Eight data sets were available for this analysis, each consisting of the velocity field extracted from the target frame. Demographics for the data sets appear in Table 1. This was a non-homogeneous data set that contained: (1) three data sets spoken by the same subject (data sets 1, 2, and 3); the latter two recorded after two months, and one year, respectively; (2) three different native languages; and (3) one speaker who underwent glossectomy surgery about 1 year prior to the study (subject 8). The surgery removed 1/3 of his tongue on the left side and replaced it with a radial forearm free flap, while preserving the tongue tip. Differences among subjects in slice thickness and tag separation, which are matched within a subject to create square voxels, changed the resolution of the data, but did not noticeably affect the goals of this preliminary study. All subjects were male.

To record the data, the subjects repeated /i/-/u/ to the first two beats of a four beat metronome set at 0, 333, 800, 1400 ms in a 2 second repeat time. The last two beats were used for a controlled inhalation and exhalation. The timing was coordinated to the trigger of the MRI machine so that the first beat occurred at the onset of the MRI acquisition and tags were applied 16 ms before the beat. The triggering method is based on that of Masaki and colleagues (Masaki et al. 1999, Shimada et al., 2002). A full explanation of the recording and analysis procedures can be found in Parthasarathy et al (2007) and Stone et al. (2009).

2.2. Data collection

To acquire each tagged cine series the subject performed 3 repetitions of each speech task per slice in each of 4 aquisitions. The four acquisitions included two orthogonal, independent, tag directions, and two complementary tagging phases for each direction; these were combined to generate a single MICSR image. Each image was acquired in *k*-space with a matrix size of 64 × 22 in 3 repetitions. This relatively small matrix acquisition size is optimized to work with the HARP analysis technique, allowing us to reduce the number of repetitions of the speech task, minimizing the potential errors associated with multiple speech task repetitions.

The first of the three repetitions was a preparation cycle that is necessary for steady state imaging, and 11 *k*-space lines were acquired in each of the other two repetitions. For 7 sagittal slices, this resulted in 84 repetitions, including four pauses. The non-tagged cine-MRI images were used to register the data sets across subjects prior to the PCA and cluster analysis. These HARP and MICSR procedures are explained in detail in Osman and Prince (1999 in Osman and Prince (2000b), NessAiver and Prince (2003), and Parthasarathy, et al. (2007).

2.3. Pre-processing of data: Registration of subject data using Cine-MRI images

To spatially align the velocity fields from the eight different data sets the target frames were identified in the *cine-MRI* images and the tongue surfaces were aligned using nine landmark points, as shown in Figure 1. This was possible because the time-frames are the same in the cine and the tagged data sets of each subject, as the subject spoke to a metronome. In the present study we normalized only the target frame for each subject. The tissue points in

Figure 1b, on the right, were determined first. These points are: (1) the base of the valleculae; (2) the upper tip of the epiglottis (projected onto the tongue surface); (4) the point on the tongue surface that lies between the elbow of the velum (or the midway point of the velum if no elbow is visible) and the upper tip of the marrow (white) visible within the mandible (black); (3) the point midway between points 2 and 4; (5) the mid palate; (7) the tongue tip; (6) the point midway between 5 and 7; (8) the origin of genioglossus; and (9) the inner aspect of the mandible.. On the left, the floor muscles were included by moving the two lowest points below the soft tissue of the chin. The anterior one is positioned to include as much of the floor musculature as possible, but not the jaw bone itself. We denote the i^{th} landmark on the j^{th} subject as $\mathbf{P}_{ij}$. The tongue region of each subject is defined as the area inside the polygon formed by connecting these landmarks.

The tongue regions in all subjects are registered to data set 1using rigid transformation plus a global scaling computed from manually picked landmark points. Without loss of generality, we pick the first data set as the reference coordinate to which all the other data sets are registered. The transformation $[s_j, \mathbf{R}_j, \mathbf{t}_j]$ of the j^{th} data set is determined by minimizing

$$E_j = \sum_{i=1}^{9} \left\| \mathbf{P}_{i1} - (s_j \mathbf{R}_j \mathbf{P}_{ij} + \mathbf{t}_j) \right\|^2 \tag{1}$$

where s_j is a scalar, $\mathbf{R}_j$ is a rotation matrix, and $\mathbf{t}_j$ is a translation vector. The registered landmark points are illustrated in Figure 2, and show the variability inherent in different subjects' resting tongue and head positions. The common region of the registered tongues is then determined (the white area in Figure 2), and we denote it as C.

Next, we must transform the velocity field inside the common region of each dataset to the reference coordinates. This is accomplished in three steps. For each point (pixel) $\mathbf{p}_k \in C$ and the j^{th} subject: (1) compute its location $\mathbf{p}_{kj}$ in the j^{th} dataset by applying inverse transform, i.e., $\mathbf{p}_{kj} = s_j^{-1} \mathbf{R}_j^T (\mathbf{p}_k - \mathbf{t}_j)$; (2) compute the velocity $\mathbf{v}(\mathbf{p}_{kj}) = [u(\mathbf{p}_{kj}), v(\mathbf{p}_{kj})]^T$ at point $\mathbf{p}_{kj}$ using HARP and linear interpolation, with $u(\mathbf{p}_{kj})$ being the velocity component in the vertical (y) direction, and $v(\mathbf{p}_{kj})$ being the velocity component in the horizontal (x) direction; and (3) transform the velocity back to the reference coordinate and scale it using $\mathbf{v}_k^{(j)} = [u_k^{(j)}, v_k^{(j)}]^T = s_j \mathbf{R}_j \mathbf{v}(\mathbf{p}_{kj})$. These steps are executed for every pixel $p_k \in C$ and every dataset.

2.4. Principal component analysis

After the tongue shapes and velocity fields are aligned, we perform PCA on all the subjects and quantify the component motions of the midsagittal velocity patterns. Suppose there are N subjects, and M points in the common region. The velocity field of the j^{th} subject can be represented as a $2M \times 1$ vector. The number of points in the common region is always much larger than the number of subjects in these data sets as there are always a large number of pixels in the tongue. In this experiment, the number of subjects, N, is 8 or 7, and the numbers of points in the common region vary from 290 (tongue) to about 420 (tongue-plus-floor).

Through PCA, the data from any subject can be represented using a linear model

$$\mathbf{w}=\overline{\mathbf{w}}+\Phi\mathbf{b}, \tag{2}$$

where $\bar{\mathbf{w}}$ is the average velocity field for all the subjects

$$\overline{\mathbf{w}}=\frac{1}{N}\sum_{j=1}^{N}\mathbf{w}_j \tag{3}$$

The columns of matrix Φ represent the modes of variation of the velocity fields, and are called principal components (PCs). They are computed from the 2M × 2M covariance matrix S, given by

$$\mathbf{S}=\frac{1}{N}\sum_{j=1}^{N}(\mathbf{w}_j-\overline{\mathbf{w}})(\mathbf{w}_j-\overline{\mathbf{w}})^T \tag{4}$$

The PCs are the eigenvectors φ_i of S with corresponding eigenvalues λ_i sorted so that $\lambda_i \geq \lambda_{i+1}$. The PC corresponding to the largest eigenvalue, i.e. φ_1 represents the direction of maximum variability in the velocity fields across the subjects.

A data set $\mathbf{w}_j$ can be fitted to the PCs by finding the coefficient vector **b** that minimizes the residue

$$E=\left\|\overline{\mathbf{w}}+\Phi\mathbf{b}_j-\mathbf{w}_j\right\| \tag{5}$$

with ||•|| being the Euclidean norm. The residue represents the motion pattern of the data that cannot be represented by the subjects used in the PCA, while $\mathbf{b}_j$ represents the amount of motion patterns that are represented by the corresponding PCs.

2.5. Hold–one-out analysis

To determine how well each subject was represented, and to consider whether the method might distinguish normal from patient subjects, we performed a “hold-one-out” experiment. Eight PCA’s were performed, each using 7 different subjects. The PC’s of each analysis were then fit to the “held-out” data set to determine how well it was represented by the PC’s of the other seven data sets.

2.6. Cluster analysis

Let the velocity field of the j^{th} subject be represented as a vector $\mathbf{w}_j=[u_1^{(j)},\ldots,u_M^{(j)},v_1^{(j)},\ldots,v_M^{(j)}]^T$, and let the Pearson correlation between the j^{th} and i^{th} subject velocity vectors be denoted as ρ_{ji} = cov($\mathbf{w}_j$, $\mathbf{w}_i$)/std($\mathbf{w}_j$)std($\mathbf{w}_i$), where cov stands for covariance. The similarity metric between two subjects is defined as

$$d(\mathbf{w}_j,\mathbf{w}_i)=1-\rho_{j,i} \tag{6}$$

Consider two clusters D and D^* which contain n and n^* subjects, respectively. Then the average linkage (distance) between two clusters is measured as

$$d_{avg}(D, D^*) = \frac{1}{n \cdot n^*} \sum_{V_j \in D} \sum_{V_i \in D^*} d(\mathbf{w}_j, \mathbf{w}_i) \tag{7}$$

It is understood that the larger the calculated distance value, the greater the difference between subjects (clusters).

The agglomerative hierarchical clustering algorithm with Pearson correlation and average linkage as a distance metric was used for the analysis. The cluster analysis begins with each subject as its own cluster and at each stage chooses the “best” merge of two subjects or of two clusters of subjects if their distance is minimized until, in the end, all subjects are merged into a single cluster. The end result of hierarchical clustering is a tree structure or dendrogram (seen in Figures 5 and 6 below). At the bottom of the tree, each subject constitutes its own cluster and, at the top of the tree, all subjects have been merged into a single cluster. Merges between two subjects or between two clusters of subjects, are represented by horizontal lines connecting them in the dendrogram (Duda et al, 2001).

3. Results

3.1. Velocity fields

Figure 3 depicts the midsagittal velocity fields for each subject during the maximum /i/-to-/u/ motion. Although the directions of tissue-point motion were primarily backward and converging, there were considerable subject differences. The first three datasets, who were the same subject at different dates, showed considerable difference in deformation pattern. The patient (Subject 8) had the least tissue point convergence. His entire midsagittal tongue moved straight backward. The second row of Figure 2 adds the floor muscles and shows that the small converging motion seen in the lower tongue in the first row was enlarged in the lower region of the tongue.

3.2. PCA of tongue-plus-floor vs. tongue-only ROIs

Two principal component analyses examined the tongue-plus-floor (PCA 1) vs. the tongue-only (PCA 2) ROIs for all eight subjects, after subtracting the mean, and calculated the percent variance accounted for by the PC’s. In PCA 1, the common region computed after registration contained about 420 pixels, and in PCA 2 it contained about 290 pixels. Table 2 shows the eigenvalues and variance explained by all the PC’s in both conditions. The first four PC’s accounted for 93% and 95% of the variance, respectively. The biggest difference between the two analyses occurred in PC’s 1 and 2. Although PC1 plus PC2 had similar explanatory power for both ROI’s (72% vs. 74%), PC 2 explained more variance in the tongue-plus-floor data (24%) than the tongue only data (15%) and PC1 explained less variance (47% vs. 58%). The associated hold-one-out analyses in Table 3 showed this same relationship, despite the varied subject demographics. A similar amount of variance was explained by PC1 plus PC2 for the tongue-plus-floor and the tongue-only R0Is (1% vs 4%), PC1 explained more variance (3% vs 13%), and PC2 less variance (−1% vs −10%).

Figure 4 depicts the mean velocity for the tongue-plus-floor ROI (panel 5) and the effects of adding or subtracting PC’s 1 and 2 in the other images. The middle row shows the addition of +/− 1SD of PC1 (panels 4, 6); the middle column depicts the addition of +/− 1SD of PC2 (panels 2, 8). The mean velocity indicates that the predominant motion direction from /i/ to /

u/ was back in the tongue body, up/back in the lower tongue/floor, and down/back in the anterior tongue, both of which converge with the body. The addition of PC1, which accounted for 47% of the variance, angled the motion downward; subtraction of PC1 angled it upward. PC2, which accounted for 24% of the variance, represented the degree of anterior tongue lowering, of up/back motion of the lower tongue, and overall magnitude of the vectors.

3.3. PC Representations of Tongue Velocity Patterns

We performed PC fits by adding to the mean velocity field the PC1 and PC2 loadings for each subject (Table 4). The velocity fields of four subjects (3, 4, 5, 7) were fit very well by the mean plus PCs 1 and 2 (82% – 100%). Subjects 1–6 were represented primarily by the mean plus PC1, that is, back or down/back motion of the tongue. PC2 increased (or decreased) the convergence in the anterior tongue. Subjects 6, 7 and 8 had smaller negative loadings on PC1 than the other subjects. Subject 8 also loaded on PC3 (24%) and PC4 (11%) (not shown), which further reduced his downward motion.

3.4. Clustering

Three cluster analyses were performed on velocities in both horizontal (*x*) and vertical (*y*) directions (hereafter, *x*-*y*), horizontal (*x*) direction only, and vertical (*y*) direction only for both ROI's. Figure 5 shows results of the cluster analysis on the combined x and y velocity data for the tongue-only data. The normal-subjects analysis (left) shows that the three datasets by the same speaker (1,2,3) clustered together, as did two of the three AE speakers (4,5). The third AE speaker (6) grouped with the Japanese speaker (7). Adding the patient to the analysis (right) did not change the cluster alliances; the patient grouped with the Japanese and one AE speaker. A comparison of the *x*-*y*, *x*, and *y* motion clusters (Figure 6) indicates that the dominant movement pattern was in the *x* (horizontal) direction.

The tongue-plus-floor data did not group similarly to the tongue-only data. Instead two clusters emerged (Figure 7). The left cluster contained subjects that primarily moved straight back in the upper tongue; the right cluster contained subjects that moved obliquely down and back (see Figure 2). For this data set the *x*-*y* clusters were more similar to the *y* direction clusters (vertical).

4.0. Discussion

4.1. Tongue-plus-floor (PCA 1) vs. tongue-only (PCA 2)

The floor muscles have a dual function in speech: to elevate the tongue, and to move the jaw and hyoid bones. PCA 2 excluded the floor muscles in its ROI. Without these muscles the velocity field variability was explained fairly well with a single PC. With them the second PC had a greater role, due to the upward motion of the lower tongue. It was concluded, therefore, that in PCA of the tongue it is important to include the floor muscle region. This result was consistent with Baer et al., (1988) who showed that the floor muscle mylohyoid was active for /i/, but not for /u /

The cluster analyses also showed differences between the two ROI's. Subjects 1, 2, 3 (the same subject) clustered together in the tongue-only data, reflecting coherence in tissue point motion within this subject's tongue-body across sessions. However, the tongue-plus-floor data did not group the three sessions, indicating that differences across data sessions were more prominent in the tongue root. Data sets 1 and 2 were more clustered more tightly than 3 in the *x*-*y* and *x* data, and 2 and 3 clustered more tightly in the *y* data. In other words, the two cluster analyses appeared to have different foci. In the tongue-only data the *x*-*y* clusters were more similar to the *x* clusters, whereas in the tongue-plus-floor data the *x*-*y* clusters

were more similar to the *y* clusters. As with the PCA, the tongue-only analysis focused on the large tongue body region that moved back vs. down/back. When the floor region was added, the clusters reflected the additional upward motion in the lower tongue, thus being more consistent with the *y* clusters. Thus the cluster analysis and the PCA behaved similarly for each ROI.

4.2. Individual subjects and averaged data

The datasets used in these analyses were quite inhomogeneous; there were replicates of one individual, multiple languages, and one partial glossectomy. Because of this and the small number of subjects, the PC1 x PC2 fits varied widely across subjects. The patient was no more unusual than some of the other subjects on the first two PC's and the cluster results.

It is worth noting that the average velocity field was in itself a good representation of the motion patterns. The transition from /i/ to /u/ primarily requires backward motion of the tongue and little or no motion of the jaw, as these sounds are known to use a "high-front" and a "high back" tongue position, respectively. Nonetheless, the observed motion went beyond rigid translation. The average velocity field showed lowering of the anterior tongue and some elevation of the tongue root (Figure 3, panel 5), as did many of the individual data sets (Figure 2). These vector directions occurred because the tongue, which has no internal skeleton, is moved by activating internal muscles, which insert at all locations on the surface of the tongue. As these muscles contract, they cause local deformation which moves and also deforms the tongue. The average velocity field represented this phenomenon well.

4.3. Comparison between results of cluster analysis and PCA

Two interesting examples provide insight into what the two methods reveal. In the first example, both methods captured similarity in the motions patterns of the Japanese speaker (7) and the AE glossectomy speaker (8). All the cluster analyses showed a tight clustering between these two subjects. In addition, these subjects loaded similarly on the PC's, with a relatively large loading on PC2 and a lesser one on PC1. However, their motion similarities have entirely different underlying reasons: language vs loss of muscle tissue. Both subjects moved the tongue very little (note color map in Figure 3). The patient lost a section of mucosa and muscle in the left lateral tongue, which was replaced by a flap of skin tissue extracted from the radial forearm. This additional bulk and weight, which facilitates bolus containment and the execution of consonants, nonetheless must be moved using the remaining, reduced musculature. In addition, the sensation and motor control of the tongue tip on the resected side may be reduced due to loss of nerve fibres on the resected side; the extent of this loss is unknown in this patient. Thus his tongue motions are slower, shorter and less deformed than normal AE speakers probably due to the effects of the flap. The Japanese speaker, on the other hand, was producing a Japanese /ɑ/. This is a mid-high, unrounded vowel that is in a different category from the English /u/. His tongue position for /ɑ/ was directly posterior and fairly nearby to that for /i/ necessitating a small, nearly straight-back motion between them. The higher PC's distinguished these two subjects. Table 4 shows that PC 1+2 accounted for 100% of the variance for subject 7, but only 61% for subject 8. PC's 3 and 4 accounted for 24% and 11% respectively of subject 8's variance. Interestingly, PC3 accounted for 44% of Subject 6's velocity field as well, who was the next most similar subject.

The second example shows that both techniques captured the same variation across sessions for datasets 1, 2 and 3 (the same subject) in the tongue-plus-floor data. This was the Tamil speaker. The vowels /i/ and /u/ in Tamil are not appreciably different from English in their citation form, as was spoken here. Data set 2 loaded positively on PC1 and negatively on PC2, data set 1 loaded positively on both, and data set 3 was negative on both. The cluster

analyses for both the *y*, and *x-y* data tightly grouped data sets 1 and 2, but not data set 3 (Figure 7), consistent with their loadings on PC1. The *x* clusters tightly grouped data sets 2 and 3, consistent with the loadings on PC 2. Further studies will need to be conducted to determine how typical this intra-subject variability is when repeat datasets are collected across a time span of a year.

4.4. One motion strategy or two?

In the present study the goal of determining whether the subjects all used essentially the same speech gesture with slight individual variation, or used different gestures, could not be fully achieved due to the small size of the data set and the varied demographics of the subjects. With a more homogeneous data set, such as normal subjects who are all native speakers of the same language, fewer PC's should represent more of the variance. However, the three AE speakers (Figure 2 - #4, 5, 6) showed a fair amount of variability and suggest that the differences seen in these data may be replicated in a single-language data set.

Both the PCA and cluster analysis are good first steps in understanding a potential duality between subjects' production of these deformations. The tongue-plus-floor clusters (Figure 7) and PC1 (Figure 2) categorized two groups of speakers who moved the tongue from /i/ to /u/ using back vs. down/back tongue motion. Although inference of muscle activation patterns needs to come from the strain data, which is being analyzed separately, the present data set allows for some speculation on the use of motor control strategies. Speakers 3, 6, 7, 8 comprised one cluster and loaded negatively on PC1. These subjects may have used the styloglossus muscle primarily to pull the tongue back, since their tongue-body vectors followed the line of action for that muscle quite closely (Figure 2). For the normal subjects (3, 6, 7) the deformation included upward motion of the tongue base, which is consistent with the pull of styloglossus on the tissue or with shortening of the floor muscles. Although this is not entirely consistent with the Baer et al study (1988) showing that in AE /u/ has a lower hyoid position than /i/, it can be noted that of these three, only subject 6 is an AE speaker. The other two AE speakers have downward and backward motion of the entire tongue, and tongue root, consistent with Baer et al, 1988. The lowering seen in their tip may indicate activation of inferior longitudinalis. Additional rigidity in the patient tongue (8), due to scar tissue and the flap, would have reduced local deformation within the tongue, resulting in his rigid backward motion. Subjects 1, 2, 4, 5 comprised the second cluster and loaded positively on PC 1. They had more down/back motion, and could have used the styloglossus or the hyoglossus as the primary muscle. If they used the hyoglossus, which pulls the tongue down/back, the floor muscles would be more likely to move straight back, as occurred for subjects 4 and 5. The upward motion of the tongue base in subjects 1 and 2 argue for the activation of the styloglossus, which would be consistent with this subject's other dataset (3). It is also possible that this up/back motion results from the engagement of the floor muscles, which can elevate the tongue-body. A better determination of these contributions will be made in parallel work, which is studying the 3D muscle anatomy, principal strains, and strains in the line of action of key muscles.

4.5. Methodological choices

The first methodological choice made in this study was the application of a rigid and scalar registration method to our subject data. This choice could affect the results of our analysis. Alternatively, it would have been possible to deformably register each tongue to a target tongue. In that case, homologous points (those corresponding to the same anatomy) among subjects could be more easily achieved, and the data analysis would seem to be fundamentally more sound. For example, the lack of perfect overlap, seen in Figure 2 as the disagreement between the common region and the blue landmark regions, would vanish. However, the data we are analyzing are vector data (velocities) which must be appropriately

interpreted under deformable registration. At present, images that are rotated and scaled in order to achieve registration must have their velocity field rotated and scaled similarly. But under deformable registration, the velocity vectors should be individually and uniquely scaled and rotated in order to agree with the deformations being applied to the whole tongue. It is not immediately clear how one would carry out this process. One might, for example, directly apply the deformation gradient of the computed (deformable registration) transformation to the velocity field. However, it is equally sensible to compute a polar decomposition of the deformation gradient and apply only the rotational part of that decomposition to the velocity field. These procedures would produce different results. On the other hand, both of these approaches derive their steps from a very local picture of the implied transformation (the deformation gradient) and this might not represent the best approach given the deformations taking place at a larger scale. Tradeoffs are inevitable as alternatives are considered and clear advantages of one approach over the other may emerge in time. Although these approaches are being explored, the current approach is considered appropriate for this preliminary study.

PCA and cluster analysis quantify and simplify complex relationships among subjects. PCA looks at between-subject variability, since the mean is subtracted out, and the 2 PC model projected high dimensional data onto a low dimensional representation. The first two PC's in these data represented only 72–74% of the variance. The cluster analysis represented all the data, that is, the mean plus all the variance. The techniques are related in that if the original data don't cluster, the PC's won't reveal any relationships.

Strengths and limitations of the methods can be seen in the results found. In the comparison of subjects 7 and 8 both analyses missed subtle, but obvious differences. Subject 8 moved the entire tongue almost straight back, whereas when subject 7 moved the upper tongue backward he angled the anterior part slightly downward and the posterior part slightly upward. He also moved the lower tongue upward. Although these two subjects were more similar than the others, their differences were not captured by the clusters or by the first two PC's. The higher PC's, which revealed differences in deformation between the two, may elucidate features of patient motion when a larger data set is studied.

Cluster analysis, which incorporated all the features, reduced the data to a greater extent than PCA, which extracted component features, and came out with similar results. However, cluster analysis is a black box and does not reveal what features are clustered. In the present data set, it was possible to guess what features formed some of the clusters, especially using the PC eigenvalues and direct data observation, but other cluster bases were opaque. In a PCA the eigenvectors of each PC can be drawn to reveal roughly the dimension of variation represented by each PC, but these dimensions are not always very interpretable. In the present data set PC 1 was more easily interpretable than PC 2. In both analyses there are no *a priori* models, so the input data must still be examined thoroughly to define the final model. These results make clear that one must be careful to include all the important features of the movement and not excessively reduce dimensionality.

5.0. Conclusion

This paper examined relationships between the tongue motion deformations of 8 speakers for a single speech gesture using PCA and cluster analysis. A comparison of a tongue-only ROI with a tongue-plus-floor ROI indicated that the addition of the floor muscles allows one to observe their contribution to the deformation, and equally importantly, to interpret the indirect effects of distant muscles, such as the styloglossus or hyoglossus, on the lower tongue and floor region. The larger ROI added complexity to the deformation which will help interpret the motor control strategies used by the speakers. Both analyses found key

features in the data, and had some overlap; both missed subtleties in the motions, as might be expected. This means that when answering scientific questions, such as how many gestures are used by subjects, subtle differences may need additional exploration. Additional types of data, such as strain data, will help interpret velocity field results.

Acknowledgments

This work was funded in part by NIH grant R01-CA133015.

References

1. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403(6769):503–11. [PubMed: 10676951]

2 . Axel L, Dougherty L. Heart wall motion: improved method of spatial modulation of magnetization for MR imaging. Radiology 1989;172:349–350. [PubMed: 2748813]

3. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 2000;406(6795):536–40. [PubMed: 10952317]

4 . Bressmann T, Sader R, Whitehill TL, Samman N. Consonant intelligibility and tongue motility in patients with partial glossectomy. J Oral Maxillofac Surg 2004;62:298–303. [PubMed: 15015161]

5 . Bressmann T, Ackloo E, Heng C, Irish JC. Quantitative three-dimensional ultrasound imaging of partially resected tongues. Otolaryngology - Head and Neck Surgery 2007;136 (5):799–805. [PubMed: 17478219]

6 . Dick, D.; Ozturk, C.; Douglas, A.; McVeigh, E.; Stone, M. Three-dimensional tracking of tongue motion using tagged-MRI. International Society for Magnetic Resonance in Medicine; 8th Scientific Meeting and Exhibition; Denver. 2000.

7 . Duda, RO.; Hart, PE.; Stork, DG. Pattern Classification. 2. John Wiley & Sons; New York: 2001.

8 . Engwall, O. A revisit to the application of MRI to the analysis of speech production - testing our assump-tions. Paper presented at 6th Int. Sem. Spee. Pro. (ISPS); Sydney, Australia. 2003.

9 . Fischer SE, et al. True myocardial motion tracking. Mag Res Med 1994;31:401–413.

10 . Harshman R, Ladefoged P, Goldstein L. Factor analysis of tongue shapes. J Acoust Soc Am 1977;62:693–713. [PubMed: 903511]

11 . Hoole P. On the lingual organization of the German vowel system. J Acoust Soc Am 1999;106(2): 1020–1032. [PubMed: 10462807]

12 . Jackson M. Phonetic theory and cross-linguistic variation in vowel articulation. UCLA Working Papers in Phonetics. 1988;(71)

13 . Lufkin R, Christianson R, Hanafee W. Normal magnetic resonance imaging anatomy of the tongue, oropharynx, hypopharnyx and larynx. Dysphagia 1987;1:119–127.

14 . Maeda, S. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: Hardcastle, WJ.; Marchal, A., editors. Speech Production and Speech Modelling. The Netherlands: Kluwer; 1990. p. 131-149.

15 . Masaki S, Tiede M, Honda K, Shimada Y, Fujimoto I, Nakamura Y, Ninomiya N. MRI-based speech production study using a synchronized sampling method. J Acoust Soc Jpn 1999;20:375–379.

16 . McKenna KM, Jabour BA, Luftkin R, Hanafee W. Magnetic resonance imaging of the tongue and oropharynx. Top Magn Reson Imaging 1990;2:49–59. [PubMed: 2223110]

17 . Napadow VJ, Chen Q, Wedeen VJ, Gilbert RJ. Biomechanical basis for lingual muscular deformation during swallowing. American Journal of Physiology 1999a;277:G695–G701. [PubMed: 10484396]

18 . Napadow VJ, Chen Q, Wedeen VJ, Gilbert RJ. Intramural mechanics of the human tongue in association with physiological deformations. Journal of Biomechanics 1999b;322:1–12.

19 . Narayanan SS, Alwan AA, Haker K. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals 1997;101(2):1064–1077.

20 . NessAiver M, Prince JL. Magnitude Image CSPAMM Reconstruction (MICSR). Magnetic Resonance in Medicine 2003;50(2):331–342. [PubMed: 12876710]

21 . Niitsu M, Kumada M, Campeau G, Niimi S, Riederer SJ, Itai Y. Tongue displacement: visualization with rapid tagged magnetization-prepared MR imaging. Radiology 1994;191:578–580. [PubMed: 8153346]

22 . Osman NF, Kerwin WS, McVeigh ER, Prince JL. Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. Magn Res Med 1999;42:1048–1060.

23 . Osman NF, Prince JL. Visualizing myocardial function using HARP MRI. Phys Med Biol 2000a; 45:1665–1682. [PubMed: 10870717]

24 . Osman NF, McVeigh ER, Prince JL. Imaging heart motion using harmonic phase MRI. IEEE Trans on Medical Imaging 2000b;9(3):186–202.

25 . Parthasarathy V, Prince JL, Stone M, Murano E, NessAiver M. Measuring tongue motion from tagged Cine-MRI using harmonic phase (HARP) processing. Journal of Acoustic Society of America 2007;121(1):491–504.

26. Shadle, CH. Encyclopedia of Language and Linguistics. 2. Vol. 9. Keith Brown; 2006. Acoustic Phonetics; p. 442-460.

27 . Shimada Y, Fujimoto I, Takemoto H, Takano S, Masaki S, Honda K, Takeo K. [Nippon Hoshasen Gijutsu Gakkai Zasshi] 4D-MRI using the synchronized sampling method (SSM). Japanese 2002;58(12):1592–1598.

28 . Slud E, Smith P, Stone M, Goldstein M. Principal Components Representation of the Two-Dimensional Coronal Tongue Surface. Phonetica 2002;59(2–3):108–133. [PubMed: 12232463]

29 . Stone M, Goldstein M, Zhang Y. Principal component analysis of cross-sectional tongue shapes in vowels. Speech Communication 1997;22:173–184.

30 . Stone M, Davis EP, Douglas AS, NessAiver M, Gullapalli R, Levine WS, Lundberg A. Modeling the motion of the internal tongue from tagged cine-MR images. J Acoust Soc Am 2001;109(6): 2974–2982. [PubMed: 11425139]

31 . Stone, M.; Liu, X.; Zhuo, J.; Gullapalli, R.; Salama, A.; Prince, JL. Principal Component Analysis of Internal Tongue Motion in Normal and Glossectomy patients with primary Closure and Free Flap. Proceedings of the Fifth B-J-K International Symposium on Biomechanics Healthcare and Information Science; Feb. 20–22, 2009; Kanazawa, Japan. 2009.

32 . Zerhouni EA, Parish DM, Rogers WJ, Yang A, Shapiro EP. Human heart: Tagging with MR imaging—a method for noninvasive assessment of myocardial motion. Radiology 1988;169(1): 59–63. [PubMed: 3420283]

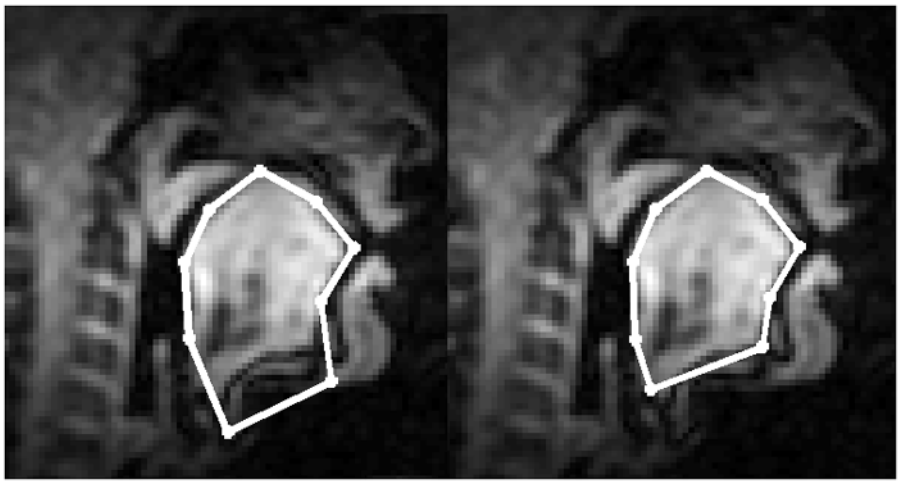

Figure 1.
Nine landmark points were used to align the ROIs for all subjects, jaw muscles were included (left) or omitted (right). The image is subject 5 at the onset of the /i/-/u/ transition.

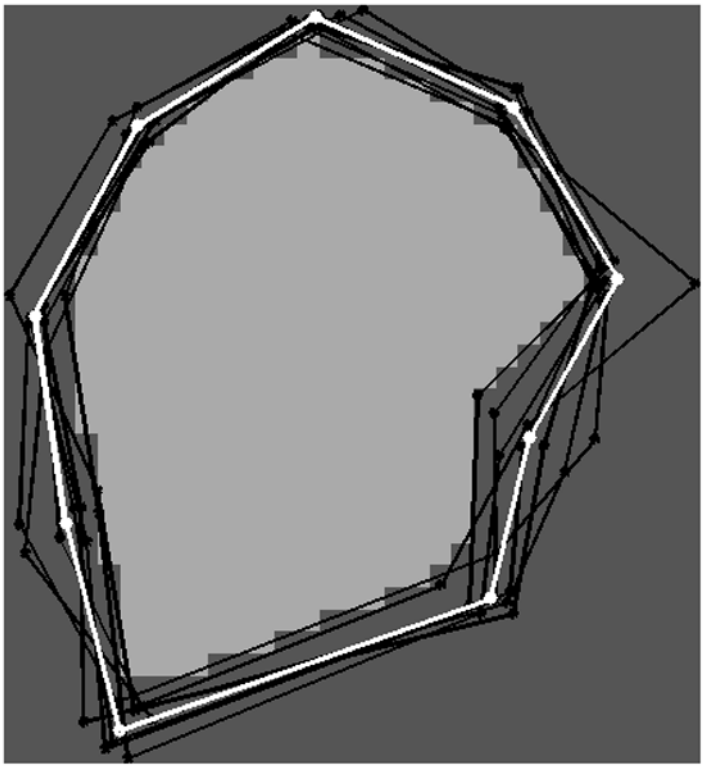

Figure 2.
The landmarks and the common region (white) for the eight tongues after alignment. The landmarks of the reference tongue are shown in white, and the other tongues are shown in black.

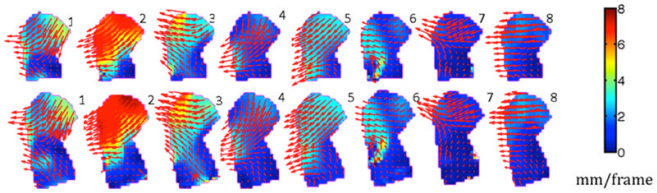

Figure 3.
Velocity fields of all subjects' regions of interest for tongue-only (top) and tongue-plus-floor (bottom)regions of interest. The velocity vectors are displayed with red arrows. The internal tongue colors reflect the magnitude of the local velocities; the colormap is on the right.

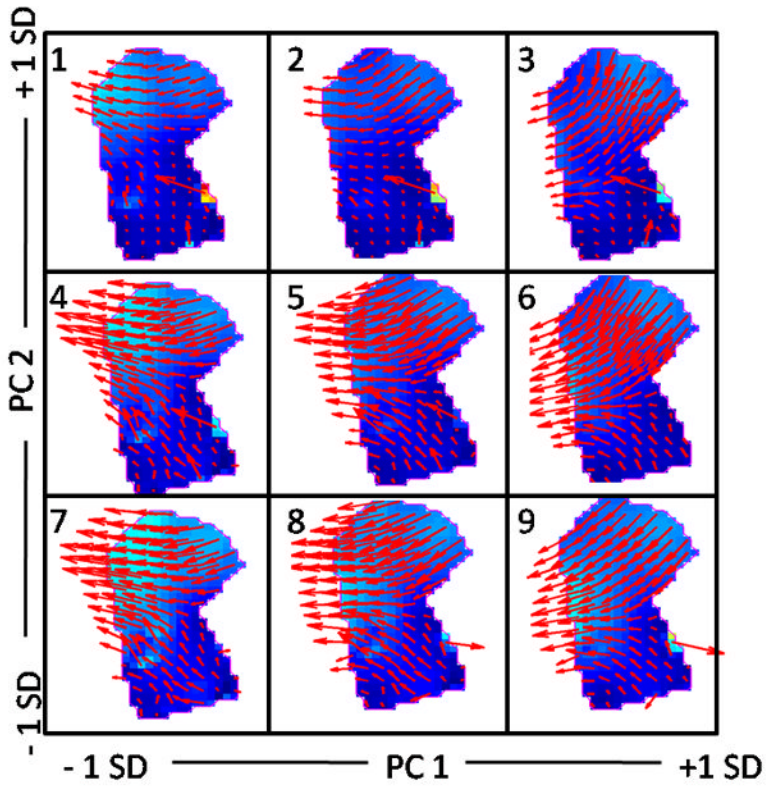

Figure 4.

Synthetic reconstructions of the effects of PC1 and PC2 added to the mean velocity field of the 8 subjects. Images consist of: the mean velocity field (panel 5), models composed by adding +/− 1SD of PC 1 (panels 4 and 6) or PC2 (panels 2 and 8), and all combinations (four corner panels) for the tongue-plus-floor data. The internal tongue colors reflect the same colormap as in Figure 3. Errors can be seen near the mental sympysis in the form of arrows of exceptional length (top row) or odd direction (bottom row). Jaw motion is an inherent part of these tongue motions.

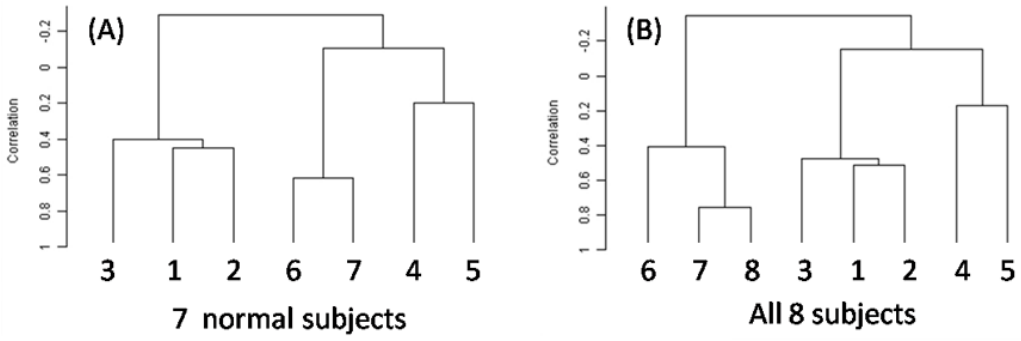

Figure 5.
Dendrograms of clusters for tongue-only data for the normal (left) and all (right) speakers.

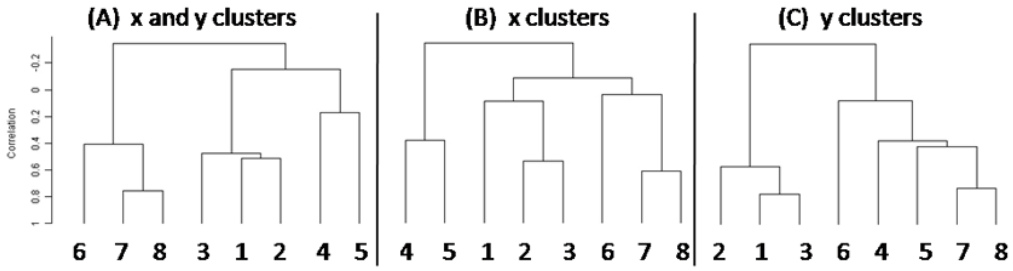

Figure 6.
Dendrograms of the tongue-only clusters for the (A) x and y, (B) x, and (C) y directions.

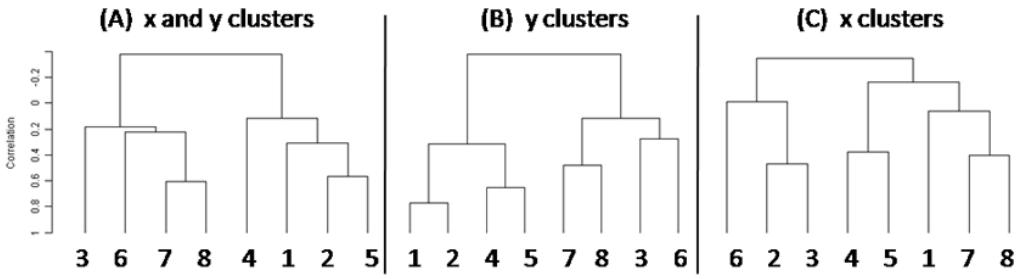

Figure 7.
Dendrograms of tongue-plus-jaw clusters for the (A) x and y, (B) y, and (C) x directions.

Table 1

Subject Demographics

Subj	Language	health	Tesla	ST / tag sep
1	Tamil	normal	1.5	7mm
2	Tamil	normal	1.5	7mm
3	Tamil	normal	1.5	7mm
4	English	normal	1.5	7mm
5	English	normal	1.5	7mm
6	English	normal	1.5	7mm
7	Japanese	normal	3.0	5mm
8	English	patient	3.0	6mm

Table 2

PCA 1 and 2. Tongue-plus-floor (T+F) vs. Tongue-only (T) data for 8 subjects.

	Eigenvalues		Percent Explained		Cumulative percent	
	T+F	T	T+F	T	T+F	T
PC 1	195	145	47%	58%	47%	58%
PC 2	99	38	24%	15%	72%	74%
PC 3	59	36	14%	15%	86%	88%
PC 4	30	16	7%	6%	93%	95%
PC 5	12	5	3%	2%	96%	97%
PC 6	10	4	2%	2%	99%	99%
PC 7	6	3	1%	1%	100%	100%

Table 3

Hold one out analyses for PCA 1 and 2. Percent variance explained by the first two PC's for the tongue-plus-floor (T+F) and the tongue (T) data.

		PC1	PC2	PC1+2
no S2	T+F	48%	25%	73%
	T	58%	16%	74%
no S3	T+F	46%	26%	71%
	T	55%	20%	75%
no S4	T+F	40%	29%	69%
	T	52%	19%	71%
no S5	T+F	47%	26%	72%
	T	58%	16%	74%
no S6	T+F	54%	28%	82%
	T	67%	18%	85%
no S7	T+F	60%	20%	80%
	T	63%	19%	83%
no S8	T+F	50%	26%	76%
	T	63%	16%	79%
min	diff	3%	−1%	1%
max	diff	13%	−10%	4%

Table 4

Percent variance explained by the first two PC's in the tongue-plus-floor data

PCs	S1	S2	S3	S4	S5	S6	S7	S8
1	63%	76%	85%	95%	79%	47%	48%	40%
2	1%	0%	8%	0%	3%	8%	52%	21%
1+2	64%	76%	93%	95%	82%	55%	100%	61%