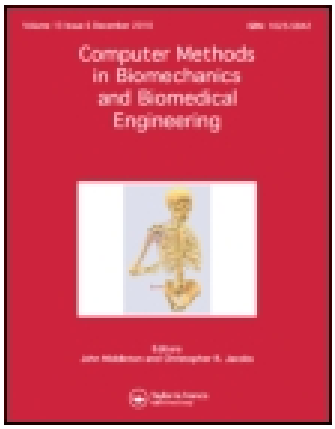

Computer Methods in Biomechanics and Biomedical Engineering
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gcmb20

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

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

[a] Department of Neural and Pain Sciences, Department of Orthodontics , University of Maryland Dental School , Baltimore, MD, 21201, USA

[b] Department of Computer Science , Johns Hopkins University , Baltimore, MD, 21201, USA

[c] Department of Epidemiology and Preventive Medicine , University of Maryland Medical School , Baltimore, MD, 21201, USA

[d] Department of Electrical and Computer Engineering , Johns Hopkins University , Baltimore, MD, 21201, USA

Published online: 15 Jul 2010.

To cite this article: Maureen Stone , Xiaofeng Liu , Hegang Chen & Jerry L. Prince (2010) A preliminary application of principal components and cluster analysis to internal tongue deformation patterns, Computer Methods in Biomechanics and Biomedical Engineering, 13:4, 493-503, DOI: 10.1080/10255842.2010.484809

To link to this article: http://dx.doi.org/10.1080/10255842.2010.484809

Computer Methods in Biomechanics and Biomedical Engineering
Vol. 13, No. 4, August 2010, 493–503

Taylor & Francis
Taylor & Francis Group

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[b,d]

[a]*Department of Neural and Pain Sciences, Department of Orthodontics, University of Maryland Dental School, Baltimore, MD 21201, USA;* [b]*Department of Computer Science, Johns Hopkins University, Baltimore, MD 21201, USA;* [c]*Department of Epidemiology and Preventive Medicine, University of Maryland Medical School, Baltimore, MD 21201, USA;* [d]*Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21201, USA*

(*Received 29 July 2009; final version received 31 March 2010*)

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 magnetic resonance imaging, but characterisation of motion strategies is difficult using visualisation alone. This paper explores the use of principal component analysis for dimensionality reduction and cluster analysis for tongue motion categorisation. Velocity fields were acquired and analysed from midsagittal tongue slices during motion from /i/ to /u/ for eight datasets containing multiple languages and a glossectomy patient. The analyses were carried out on the tongue-only and tongue-plus-floor of the mouth regions. The results showed that both the analyses were sensitive to region size and that cluster analysis was harder to interpret. Both the 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; magnetic resonance imaging; tags

1. Introduction and background

Speech creation and intelligibility are highly dependent on tongue deformation because the change in the shape of the vocal tract is primarily caused by the 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 on 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. Thus, the question arises as to whether multiple speakers (or one speaker at 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 the patterns of tissue-point motion in the internal tongue among different speakers; these patterns can be extracted from velocity fields in tagged magnetic resonance imaging (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 towards determining the 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 had several different native languages and one had undergone surgery to remove a part of the tongue due to cancer (partial glossectomy). Despite these demographics, which should

*Corresponding author. Email: mstone@umaryland.edu

ISSN 1025-5842 print/ISSN 1476-8259 online

DOI: 10.1080/10255842.2010.484809
http://www.informaworld.com

increase the variety of patterns seen, all the subjects produced a normal–sounding /i/–/u/, including the patient.

Imaging was carried out using tagged MRI and processed using the harmonic phase (HARP) method (Osman et al. 1999). Data analysis was carried out using principal components analysis (PCA) and clustering. MRI has been used for many years to capture the shape of both the oral cavity and vocal tract (Lufkin et al. 1987; McKenna et al. 1990; Engwall 2003) and in imaging the motion of the tongue (Narayanan et al. 1997; Masaki et al. 1999; Stone et al. 2002; Shadle 2006). 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 and Dougherty 1989), which is now being used in several studies of tongue motion (Niitsu et al. 1994; Napadow et al. 1999a, 1999b; 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 et al. 1999) and provide the data that have been 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 have not been considered in the present paper, and the patterns of motion alone have been 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 2000), 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. The 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 datasets using PCA and cluster analysis to examine the differences and similarities among their internal motion patterns.

Because of the large quantity of data, even with eight subjects, methods for simplifying and grouping the data were important. PCA is an excellent standard method for extracting and representing patterns in high-dimensional data for which no expectations or *a priori* models are available. PCA reduces the dimensionality of a dataset by determining the main orthogonal directions of data variability and is typically applied to a dataset 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 characterise speech-related motion of the midsagittal tongue surface (Harshman et al. 1977; Jackson 1988; Maeda 1990; Hoole 1999) and the coronal tongue surface (Stone et al. 1997; Slud et al. 2002). The speech of tongue-cancer patients, pre- and post-glossectomy surgery, was also characterised using PCA, and the results were able to distinguish tongue surface motions resulting from different reconstruction procedures (Bressmann et al. 2004, 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 PCAs were performed. PCA1 compared an ROI that included the tongue-plus-floor muscles for the eight datasets (see Figure 1(a)). PCA2 was performed on a smaller ROI, the tongue-only (see Figure 1(b)). Hold-one-out analyses were also performed for each of the individual subjects to determine if the patient was represented more poorly by PCs derived from a group that excluded him than 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 those 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 him/her into successively larger clusters. Divisive algorithms begin with the whole

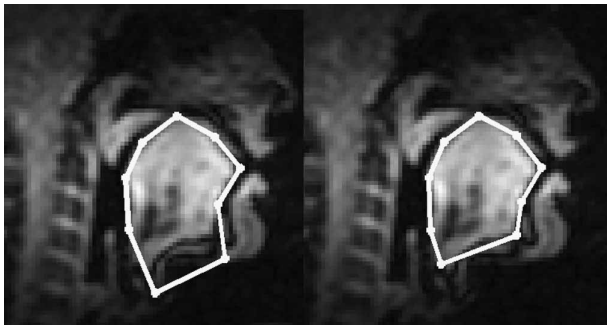

Figure 1. Nine landmark points were used to align the ROIs for all subjects; jaw muscles were included (left) or omitted (right). The image depicts Subject 5 at the onset of the /i/–/u/ transition.

group and proceed to divide it into successively smaller clusters. The clustering method requires specification of both a similarity metric and 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 the distance between two subjects, additional specifications of the distance between clusters are required to define the distance between two clusters. The specification of the 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 tumours.

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 the subjects. The clustering trees were generated to explore the features that could be useful for categorisation. With a larger dataset, 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 datasets were available for this analysis, each consisting of the velocity field extracted from the target frame. The demographics for the datasets are presented in Table 1. This was a non-homogeneous dataset that contained: (1) three datasets spoken by the same subject (datasets 1–3); the latter two were recorded after 2 months and 1 year, respectively; (2) three different native languages and (3) one speaker who underwent glossectomy surgery about 1 year prior to the study (Subject 8).

Table 1. Subject demographics.

Subject	Language	Health	Tesla	ST/tag sep (mm)
1	Tamil	Normal	1.5	7
2	Tamil	Normal	1.5	7
3	Tamil	Normal	1.5	7
4	English	Normal	1.5	7
5	English	Normal	1.5	7
6	English	Normal	1.5	7
7	Japanese	Normal	3.0	5
8	English	Patient	3.0	6

The surgery removed one-third of his tongue on the left side and replaced it with a radial forearm free flap, while preserving the tongue tip. The differences among the subjects in slice thickness and tag separation, 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 the 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 and 1400 ms in a 2 s 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 et al. (1999) and Shimada et al. (2002). A full explanation of the recording and analysis procedures can be found in the work carried out by Parthasarathy et al. (2007) and Stone et al. (2009).

2.2 Data collection

To acquire each tagged cine series, the subjects performed three repetitions of each speech task per slice in each of the four acquisitions. 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 a *k*-space with a matrix size of 64×22 in three repetitions. This relatively small matrix acquisition size was optimised to work with the HARP analysis technique, allowing us to reduce the number of repetitions of the speech task, minimising the potential errors associated with multiple speech task repetitions.

The first of the three repetitions was a preparation cycle necessary for steady-state imaging, and 11 *k*-space lines were acquired in each of the other two repetitions. For seven sagittal slices, this resulted in 84 repetitions, including four pauses. The non-tagged cine-MRI images were used to register the datasets across the subjects prior to the PCA and cluster analysis. These HARP and MICSR procedures are explained in detail in the work carried out by Osman et al. (1999, 2000), 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 datasets, 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 datasets of each subject, as the subject spoke to a metronome. In the present study, we normalised

only the target frame for each subject. The tissue points in Figure 1, 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); (3) the point midway between points (2) and (4); (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); (5) the mid palate; (6) the point midway between (5) and (7); (7) the tongue tip; (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 the subjects were registered to dataset 1 using rigid transformation plus a global scaling computed from manually picked landmark points. Without loss of generality, we picked the first dataset as the reference coordinate to which all the other datasets were registered. The transformation $[s_j, \mathbf{R}_j, \mathbf{t}_j]$ of the *j*th dataset was determined by minimising

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

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

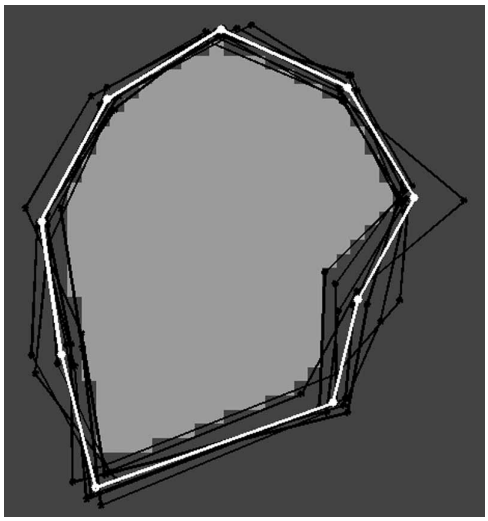

Figure 2. The landmarks and the common region (white) for the eight tongues after alignment. The landmarks of the reference tongue are shown in white, and the other tongues are shown in black.

Next, we transformed the velocity field inside the common region of each dataset to the reference coordinates. This was accomplished in three steps. For each point (pixel) $\mathbf{p}_k \in C$ and the *j*th subject, we: (1) computed its location $\mathbf{p}_{kj}$ in the *j*th dataset by applying inverse transform, i.e. $\mathbf{p}_{kj} = s_j^{-1}\mathbf{R}_j^{\mathrm{T}}(\mathbf{p}_k - \mathbf{t}_j)$; (2) computed the velocity $\mathbf{v}(\mathbf{p}_{kj}) = [u(\mathbf{p}_{kj}), v(\mathbf{p}_{kj})]^{\mathrm{T}}$ 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) transformed the velocity back to the reference coordinate and scaled it using $\mathbf{v}_k^{(j)} = [u_k^{(j)}, v_k^{(j)}]^{\mathrm{T}} = s_j \mathbf{R}_j \mathbf{v}(\mathbf{p}_{kj})$. These steps were executed for every pixel $\mathbf{p}_k \in C$ and every dataset.

2.4 *Principal component analysis*

After the tongue shapes and velocity fields were aligned, we performed PCA on all the subjects and quantified the component motions of the midsagittal velocity patterns. Let us consider that 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 datasets, because there are always a large number of pixels in the tongue. In this experiment, the number of subjects, *N*, was 8 or 7, and the numbers of points in the common region varied 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

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

The columns of matrix Φ represent the modes of variation of the velocity fields and are called PCs. They are computed from the $2M \times 2M$ covariance matrix $\mathbf{S}$, given by

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

The PCs are the eigenvectors ϕ_i of S with corresponding eigenvalues λ_i sorted, so that $\lambda_i \geqq \lambda_{i+1}$. 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 dataset $\mathbf{w}_j$ can be fitted to the PCs by finding the coefficient vector $\mathbf{b}$ that minimises the residue

$$E = \|\bar{\mathbf{w}} + \Phi\mathbf{b}_j - \mathbf{w}_j\|, \tag{5}$$

with $\|\cdot\|$ 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 PCAs were performed, each using seven different subjects. The PCs of each analysis were then fit to the 'held-out' dataset to determine how well it was represented by the PCs of the other seven datasets.

2.6 Cluster analysis

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

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

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

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

It is understood that the larger the calculated distance value, the greater is the difference between the 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 the two subjects or two clusters of subjects if their distance is minimised until, in the end, all the 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). 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/–/u/ motion. Although the directions of the tissue-point motion were primarily backward and converging, there were considerable subject differences. The first three datasets, represented by the same subject at different dates, showed considerable differences 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 is enlarged in the lower region of the tongue.

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

Two PCAs examined the tongue-plus-floor (PCA1) vs. the tongue-only (PCA2) ROIs for all the eight subjects after subtracting the mean, and calculated the per cent variance accounted for by the PCs. In PCA1, the common region computed after registration contained about 420 pixels, and in PCA2, it contained about 290 pixels. Table 2 shows the eigenvalues and variance explained by all the PCs in

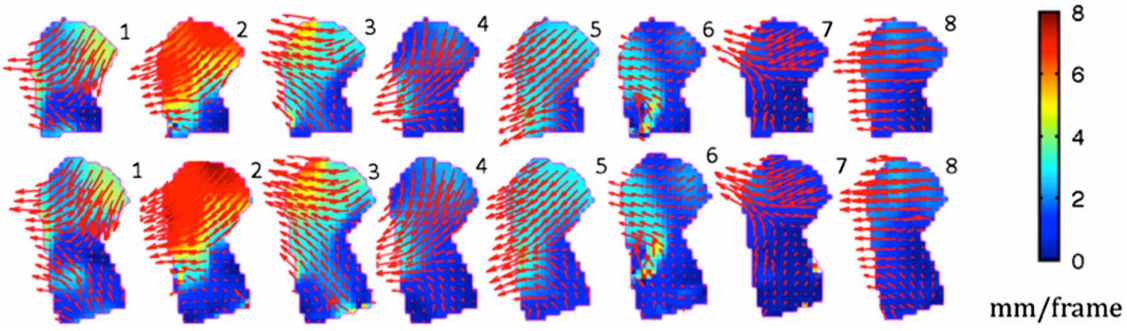

Figure 3. Velocity fields of all the subjects' regions of interest for tongue-only (top) and tongue-plus-floor (bottom) ROI. The velocity vectors are displayed with red arrows. The internal tongue colours reflect the magnitude of the local velocities; the colour map is on the right.

Table 2. PCA1 and 2. Tongue-plus-floor (T + F) vs. tongue only (T) data for eight subjects.

	Eigenvalues		Explained (%)		Cumulative (%)	
	T + F	T	T + F	T	T + F	T
PC1	195	145	47	58	47	58
PC2	99	38	24	15	72	74
PC3	59	36	14	15	86	88
PC4	30	16	7	6	93	95
PC5	12	5	3	2	96	97
PC6	10	4	2	2	99	99
PC7	6	3	1	1	100	100

both the conditions. The first four PCs accounted for 93 and 95% of the variance, respectively. The biggest difference between the two analyses occurred in PCs 1 and 2. Although PC1 plus PC2 had similar explanatory power for both the ROIs (72 vs. 74%), PC2 demonstrated more variance in the tongue-plus-floor data (24%) than the tongue-only data (15%) and PC1 showed less variance (47 vs. 58%). The associated hold-one-out analyses presented in Table 3 show similar relationship, despite varied subject demographics. A similar amount of variance was explained by PC1 and PC2 for the tongue-plus-floor and the tongue-only ROIs (1 vs. 4%); PC1 explained more variance (3 vs. 13%) and PC2 less variance (−1 vs. −10%).

Table 3. Hold-one-out analyses for PCA1 and 2. Per cent variance explained by the first two PCs 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

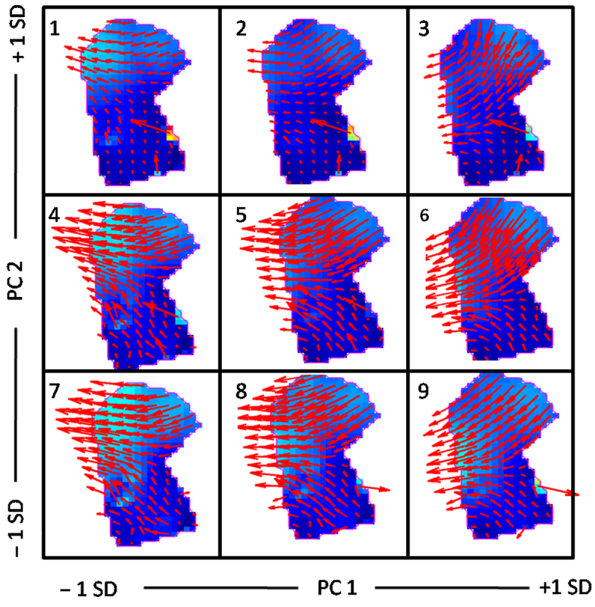

Figure 4. Synthetic reconstructions of the effects of PC1 and PC2 added to the mean velocity field of the eight subjects. Images consist of the mean velocity field (Panel 5), models composed by adding ±1SD of PC1 (Panels 4 and 6) or PC2 (Panels 2 and 8) and all combinations (four corner panels) for the tongue-plus-floor data. The internal tongue colours reflect the same colour map as in Figure 3. Errors can be seen near the mental symphysis 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 4 depicts the mean velocity for the tongue-plus-floor ROI (Panel 5) and the effects of adding or subtracting PCs 1 and 2 in the other images. The middle row shows the addition of ±1SD of PC1 (Panels 4 and 6) and the middle column depicts the addition of ±1SD of PC2 (Panels 2 and 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, while subtraction of PC1 angled it upward. PC2, which accounted for 24% of the variance, represented the degree of anterior-tongue lowering, 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 PC1 and PC2 loadings to the mean velocity field for each subject (Table 4). The velocity fields of four subjects (3–5 and 7) were fitted 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.

Table 4. Per cent variance explained by the first two PCs in the tongue-plus-floor (T + F) 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

Subjects 6–8 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 the ROIs. Figure 5 shows the results of the cluster analysis on the x-y velocity data for the tongue-only ROI. The normal-subjects analysis (left) showed that the three datasets by the same speaker (1)–(3) clustered together, as did two of the three AE speakers (4) and (5). The third AE speaker (6) was grouped with the Japanese speaker (7). Addition of the patient to the analysis (right) did not change the cluster alliances; the patient was grouped with the Japanese and the one AE speaker. A comparison of the x–y, x and y motion clusters (Figure 6) indicated that the dominant movement pattern was in the x (horizontal) direction.

The tongue-plus-floor data did not group in a similar way 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 dataset, the x–y clusters were more similar to the y-direction clusters (vertical).

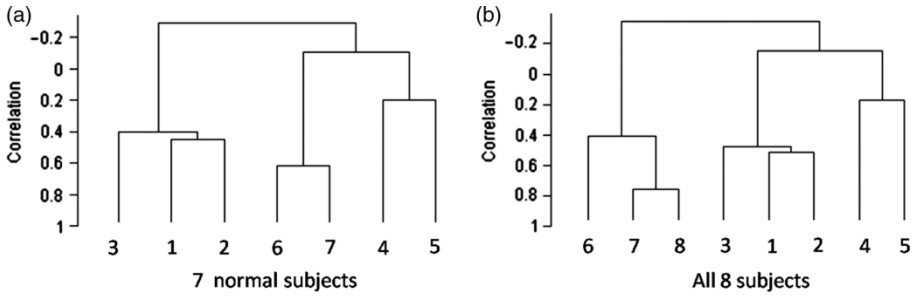

Figure 5. Dendrograms of x-y clusters for tongue-only data for the normal (left) and all (right) speakers.

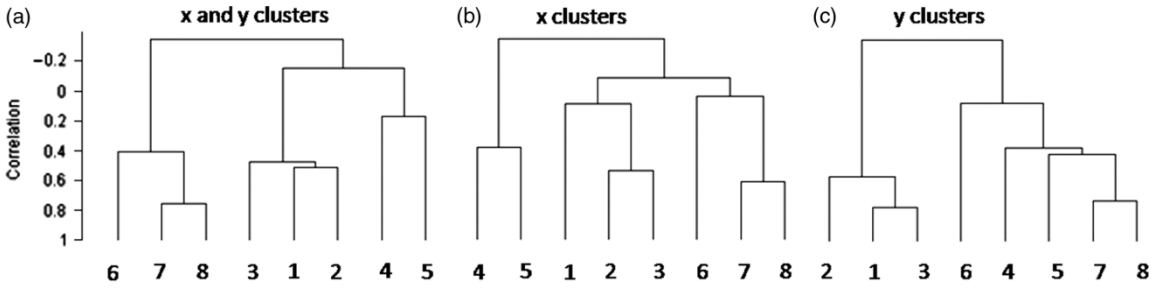

Figure 6. Dendrograms of the tongue-only clusters for the (A) x-y, (B) x and (C) y directions.

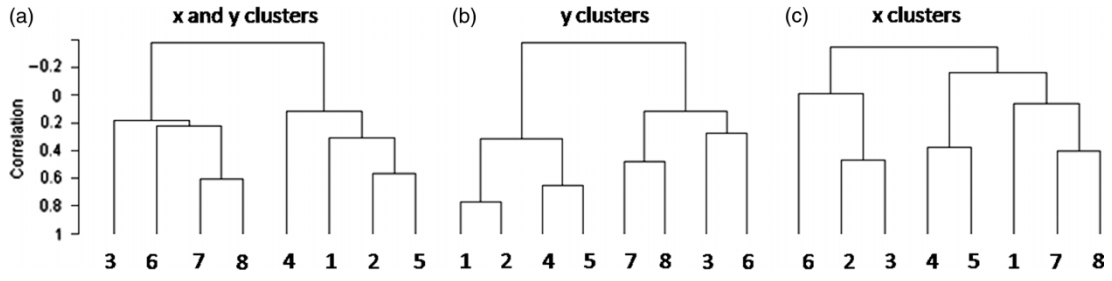

Figure 7. Dendrograms of tongue-plus-jaw clusters for the (A) x and y, (B) y and (C) x directions.

4. Discussion

4.1 Tongue-plus-floor (PCA1) vs. tongue-only (PCA2)

The floor muscles have a dual function in speech: to elevate the tongue as well as to move the jaw and hyoid bones. PCA2 excluded the floor muscles in its ROI. Without these muscles the velocity field variability was explained fairly well with a single PC. However, with them, the second PC had a greater role, due to the upward motion of the lower tongue. Therefore, it was concluded that in PCA of the tongue it is important to include the floor muscle region. This result is consistent with that observed by Baer et al. (1988), who showed that the floor muscle mylohyoid was active for /i/, but not for /u/.

The cluster analyses also showed the differences between the two ROIs. Subjects 1–3 (the same subject) clustered together in the tongue-only data, reflecting coherence in the tissue-point motion within this subject's tongue-body across the sessions. However, the tongue-plus-floor data did not group the three sessions, indicating that the differences across the data sessions were more prominent in the tongue root. Datasets 1 and 2 were 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 × PC2 fits varied widely across the subjects. The patient was no more unusual than some of the other subjects on the first two PCs 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 datasets (Figure 2). These vector directions occurred because the tongue, which has no internal skeleton, is moved by activating the internal muscles inserted at all the locations on the surface of the tongue. As these muscles contract, they cause local deformation that moves and also deforms the tongue. The average velocity field represented this phenomenon well.

4.3 Comparison between the results of cluster analysis and PCA

Two interesting examples provide insight into what the two methods reveal. In the first example, both the methods captured similarity in the motion 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 were loaded similarly on the PCs, 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 the subjects moved the tongue very little (note the colour 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 the skin tissue extracted from the radial forearm. This additional bulk and weight, which facilitates bolus containment and execution of consonants, 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 /ɯ/. This is a mid-high, unrounded vowel that is in a different category from the English /u/. His tongue position for /u/ was directly posterior and fairly nearby to that for /i/ necessitating a small, nearly straight-back motion between them. The higher PCs distinguished these two subjects. Table 4 shows that PC1 + PC2 accounted for 100% of the variance for Subject 7, but only 61% for Subject 8. PCs 3 and 4 accounted for 24 and 11% of Subject 8's variance, respectively. 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 the techniques captured the same variation across the sessions for datasets 1–3 (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. Dataset 2 loaded positively on PC1 and negatively on PC2, dataset 1 loaded positively on both and dataset 3 was negative on both. The cluster analyses for both the y and x–y data tightly grouped datasets 1 and 2, but not dataset 3 (Figure 7), consistent with their loadings on PC1. The x clusters tightly grouped

datasets 2 and 3, consistent with the loadings on PC2. Further studies must 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 all the subjects 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 dataset and the varied demographics of the subjects. With a more homogeneous dataset, such as normal subjects who are all native speakers of the same language, fewer PCs should represent more variance. However, the three AE speakers (Figure 2 and Nos 4–6) showed a fair amount of variability, suggesting that the differences seen in these data may be replicated in a single-language dataset.

Both PCA and cluster analysis are good first steps in understanding a potential duality between the subjects' production of these deformations. The tongue-plus-floor clusters (Figure 7) and PC1 (Figure 2) categorised two groups of speakers who moved the tongue from /i/ to /u/ using back vs. down/back tongue motion. Although the inference of muscle activation patterns needs to come from the strain data, which is being analysed separately, the present dataset allows for some speculation in the use of motor control strategies. Speakers 3, 6, 7 and 8 comprised one cluster and loaded negatively on PC1. These subjects may have used the styloglossus muscle primarily to pull the tongue back, because their tongue-body vectors followed the line of action for that muscle quite closely (Figure 2). For the normal subjects (3, 6 and 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 study by Baer et al. (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 flap, would have reduced local deformation within the tongue, resulting in his rigid backward motion. Subjects 1, 2, 4 and 5 comprised the second cluster and loaded positively on PC1. They had down/back motion and could have used the styloglossus or hyoglossus as the primary muscle. If they used the hyoglossus, which pulls the tongue down/back, the floor would be more likely to move straight back, as observed in Subjects 4 and 5. The upward motion of the tongue base in Subjects 1 and 2 argues for the activation of the styloglossus, which is 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 a parallel work 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 the 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 that we analysed are vector data (velocities) that 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. However, 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 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 the subjects. PCA looks at between-subject variability, because the mean is subtracted out and the two PC models project high-dimensional data onto a low-dimensional representation. The first two PCs in these data represent only 72–74% of the variance. The cluster analysis represents all the data, that is, the mean plus all the variance. The techniques are related in that if the original data do not cluster, the PCs would not reveal any relationships.

Strengths and limitations of the methods can be seen in the results obtained. In the comparison of Subjects 7 and 8, both the analyses missed the 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 PCs. The higher PCs, which revealed differences in the deformation between the two, may elucidate features of patient motion when a larger dataset 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 dataset, 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 roughly reveal the dimension of variation represented by each PC, but these dimensions are not always very interpretable. In the present dataset, PC1 was more easily interpretable than PC2. In both the analyses, there are no *a priori* models; hence, 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 the dimensionality.

5. Conclusion

This paper examined the relationships between the tongue motion deformations of eight 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 helped to interpret the motor control strategies used by the speakers. Both the analyses found key features in the data and had some overlap; both missed the subtleties in the motions, as might be expected. This means that when answering scientific questions, such as how many gestures are used by the subjects, subtle differences may need additional exploration. Additional types of data, such as strain data, will help to interpret velocity field results.

Acknowledgement

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

References

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X et al., 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 403(6769):503–511.

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

Baer T, Alfonso PJ, Honda K. 1988. Electromyography of the Tongue Muscles During Vowels in /apvp/ Environment. Research Bulletin of the Institute of Logopedics and Phoniatrics. 22:7–19.

Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A et al., 2000. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature. 406(6795):536–540.

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

Bressmann T, Ackloo E, Heng C, Irish JC. 2007. Quantitative three-dimensional ultrasound imaging of partially resected tongues. Otolaryngol – Head Neck Surg. 136(5):799–805.

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

Duda RO, Hart PE, Stork DG. 2001. Pattern classification. 2nd ed. New York: Wiley.

Engwall O. 2003. A revisit to the application of MRI to the analysis of speech production – testing our assumptions. Paper presented at: Proceedings of the Sixth International Seminar on Speech Production (ISPS); Sydney, Australia.

Fischer SE, McKinnon GC, Scheidegger MB, Prins W, Meier D, Boesiger P. 1994. True myocardial motion tracking. Magn. Reson. Med. 31:401–413.

Harshman R, Ladefoged P, Goldstein L. 1977. Factor analysis of tongue shapes. J Acoust Soc Am. 62:693–713.

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

Jackson M. 1988. Phonetic theory and cross-linguistic variation in vowel articulation., UCLA Working Papers in Phonetics. No. 71.

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

Maeda S. 1990. 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. p. 131–149.

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

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

Napadow VJ, Chen Q, Wedeen VJ, Gilbert RJ. 1999a. Biomechanical basis for lingual muscular deformation during swallowing. Am J Physiol. 277:G695–G701.

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

Narayanan SS, Alwan AA, Haker K. 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and

EPG data. Part I. The laterals. J Acoust Soc Am. 101(2):1064–1077.

NessAiver M, Prince JL. 2003. Magnitude image CSPAMM reconstruction (MICSR). Magn Reson Med. 50(2):331–342.

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

Osman NF, Prince JL. 2000. Visualizing myocardial function using HARP MRI. Phys. Med. Biol. 45:1665–1682.

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

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

Parthasarathy V, Prince JL, Stone M, Murano E, NessAiver M. 2007. Measuring tongue motion from tagged Cine-MRI using harmonic phase (HARP) processing. J Acoust Soc Am. 121(1):491–504.

Shadle CH. 2006. Acoustic phonetics. In: Brown K, editor. Encyclopedia of language and linguistics. Vol. 9, 2nd ed. p. 442–460.

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

Slud E, Smith P, Stone M, Goldstein M. 2002. Principal components representation of the two-dimensional coronal tongue surface. Phonetica. 59(2–3):108–133.

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

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

Stone M, Liu X, Zhuo J, Gullapalli R, Salama A, Prince JL. 2009. 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; 20–22 February, Kanazawa, Japan.

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