

Author Manuscript

Comput Methods Biomech Biomed Eng Imaging Vis. Author manuscript.

Patterns of variance in /s/ during normal and glossectomy speech

Maureen Stone,

Department of Neural and Pain Science, Department of Orthodontics, University of Maryland School of Dentistry, Baltimore, MD, USA. telephone: 410-706-1269, fax: 410-706-0865, mstone@umaryland.edu

Jonghye Woo,

Neural and Pain Science, University of Maryland School of Dentistry, Baltimore, MD; telephone: 410-706-1270, Fax: 410-706-0865, jwoo@umaryland.edu

Jiachen Zhuo,

Diagnostic Radiology, University of Maryland School of Medicine, Baltimore, MD; 410-328-5974, Fax: 410-328-5937, jzhuo@umm.edu

Hegang Chen, and

Epidemiology, University of Maryland School of Medicine, Baltimore, MD; 410-706-4067, Fax: 410-706-4031, hchen@epi.umaryland.edu

Jerry L. Prince

Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD; telephone: 410-516-5192, jprince@jhu.edu

Abstract

The production of speech includes considerable variability in speech gestures despite our perception of very repeatable sounds. Variability is seen in vocal tract shapes and tongue contours when different speakers produce the same sound. This study asks whether internal tongue motion patterns for a specific sound are similar across subjects, or whether they indicate multiple gestures. There are two variants of the sound /s/, which may produce two gestures, or may represent a multitude of gestures. The first goal of this paper is to quantify internal tongue differences between these allophones in normal speakers. The second goal is to test how these differences are affected by subjects expected to have different speech gestures: normal controls and subjects who have had tongue cancer surgery. The study uses tagged MRI to capture midsagittal tongue motion patterns and Principal Components Analyses to identify patterns of variability that define subject groups and /s/-types. Results showed no motion differences between apical and laminal controls in either the tongue tip or whole-tongue. These results did not support unique tongue behaviours for apical and laminal /s/. The apical patients, however, differed from all other speakers and were quite uniform as a group. They had no elevation and considerable downward/backward motion of the tongue tip. This was consistent with difficulty maintaining the tip-blade region at the proper distance from the palate.

Keywords

tongue motion; MRI; speech; principal components

Corresponding Author: Maureen Stone, mstone@umaryland.edu.

1. Introduction

The motor control of the tongue is quite complicated due to the tongue's location, sandwiched between the palate and the jaw. This location limits the tongue's range of motion and produces boundary constraints that require the tongue to use complex deformations to create the myriad gestures associated with speech production. The tongue is well equipped to produce complex gestures, because it is a volume preserving, soft tissue structure, with a complex architecture consisting of 3D interdigitated, orthogonal muscles capable of both global and local stretch. The complex muscle architecture is highly innervated, with at least 8,000 motor units (cf, Wozniak and Young, 1969; Atsumi and Miyatake, 1987) allowing the possibility of very local control. The volume preservation means that compression in one region results in expansion elsewhere (Kier and Smith, 1985) allowing deformation during motion. In fact, "volume shifting" is the primary mechanism by which the tongue moves. The material presented in this paper is about speech, but the principals and the tools can be applied to broader applications, such as swallowing and respiration.

Tagged-MRI was originally developed to create temporary magnetic patterns in heart tissue that could be used to measure motion (Axel & Dougherty, 1989). The same principal can be used to create temporary patterns in tongue muscle in order to measure motion. Cine-MRI provides an MRI sequence with crisp structural boundaries of the tongue and vocal tract suitable for modelling the tongue and airway surface (cf. Ventura, Freitas and Tavares, 2010). Tagged-MRI, on the other hand, magnetically marks, or tags, all planes of tissue within the soft tissue structures of the vocal tract, prior to collecting the MR image sequence. When the MRI sequence is collected, the motions of the tagged planes are visible in each time-frame. Usually motion of the tagged planes is tracked and motion of the points between the tags is interpolated. The present study uses harmonic phase (HARP) imaging and image processing methods, which were developed to track every tissue point in the tongue independently with no interpolation, resulting in more reliable tracking of tissue motion (Osman, Kerwin, McVeigh & Prince, 1999).

Tagged MRI is not the most direct way to study motor control; the direct method would be to measure muscle activity using electromyography (EMG). However, the muscle fibres of the tongue are interdigitated and multidirectional, which makes it extremely challenging to collect and interpret EMG of most tongue muscles. As an alternative method, tagged-MRI indirectly studies motor control strategies by capturing tissue point motion, which is the behaviour intermediate between muscle activity and tongue surface shape. Tissue-point motion patterns reveal commonalities and divergences among subjects. Strains are closely related to muscle activity since compression results from muscle contraction, as well as passive compression, while muscle expansion is passive in these speech tasks. Examples of strain revealing interaction between different tongue regions and the tongue and jaw can be seen in Unay, Ozturk and Stone (2012). In that study, strain fields extracted from tagged-MR images of /Ca/ syllables showed the effect of coarticulation on the deformation of the tongue into /a/. The present study examines velocity fields because they indicate magnitude and direction of motion, not compression, which should be closely linked with the surface tongue displacements observed in cine-MR images. This is the first step towards categorization of groups and determining the number and types of control strategies used to create the speech gesture.

Two variables were controlled in this study to help ascertain whether motion pattern differences among subjects were due to individual fine-tuning of a single gesture or truly different motor-equivalent gestures. The first variable was /s/-type. There are two variants of /s/ in English: apical and laminal (cf. Dart 1991, 1998, Narayanan, et al., 1995, Yunusova,

Stone et al.

et al., 2012). The apical /s/ elevates the tongue tip to contact the palate, create a narrow, grooved constriction, and focus the jet stream of air onto the incisors. The laminal /s/ uses the tongue blade, just behind the tip, to create the grooved constriction, and the tip is kept lower in the mouth. The effect of this natural variation is to provide two groups of /s/ producers whose variability should be smaller within than across /s/-types. Previous work from our lab used midsagittal cine-MRI, not tags, to classify apical and laminal /s/ tongue surface shapes as apical and laminal based on Dart's (1991) categorization scheme (Stone, Rizk, Woo, Murano, Chen, Prince, 2013). That work found a correlation between palate height and /s/ type for controls; low palate speakers were more likely to produce an apical /s/, and high palate speakers were more likely to produce a laminal one.

The second variable was subject group. Stone et al., (2013) found that glossectomy patients, unlike controls, were more likely to produce laminal /s/ irrespective of palate height. That result was part of the motivation to define tag motion patterns for patients in the two /s/ types. The two subject groups were normal controls and glossectomy patients. Glossectomy surgery is performed on people with tongue cancer to remove the malignant tissue and a margin of healthy tissue. The patients in the present study lost tissue in the lateral portion of their tongue, behind the tip, on one side. The effect of this loss is to reduce control of the tongue tip due to severance of the nerve supply and of muscle fibres running from back to front of the tongue. Reduced control of the tongue tip and tissue loss make the tongue groove needed for /s/ more challenging, and glossectomy patients often exhibit difficulty with voiceless fricatives, such as /s/ (Nicolletti, et al., 2004).

The patients used in the present study have small to moderate tumour sizes: up to 4 cm long before surgery. For smaller excisions speech quality has been shown to be consistently better than larger excisions, containing mild distortions at worst (Heller, Levy and Sciubba, 1991, Nicolletti, et al, 2004). This suggests only a mild reduction in motor control and the expectation that patients who cannot continue to produce a tip-up apical /s/ will be able switch to the tip-down laminal /s/, as both are commonly used and not likely to contain unusual motor requirements.

Mathematical quantification of the tongue's complex internal motion patterns is challenging because the tongue deforms in three-dimensions, and the deformations are local, large and rapid. The present paper simplifies the challenge by quantifying two-dimensional motion patterns in the midsagittal plane, and also by using Principal Components Analyses (PCA) to reduce the dimensionality of the motions. Previous work from our lab used PCA to examine tag motion in the midsagittal tongue. First, a study examined the simple backward motion of the tongue from /i/ to /u (Stone et al, 2010). Seven controls speaking three native languages, and one glossectomy patient were studied. The motion was well represented by PC's 1 and 2, which accounted for 74% of the data. Subsequent hold-one-out analyses were used to assess the contribution of the glossectomy patient and the languages to the PC features. It was determined that the different native languages produced as great an effect on the PC's as the glossectomy surgery, Therefore, our subsequent studies have used only native American English speakers from the local Maryland, Pennsylvania region of the US.

A second paper looked at the motion patterns of the midline, a left and a right sagittal slice from 10 controls and 3 glossectomy patients (Stone, Langguth, Woo, Chen, Prince, 2013). That study found that motions of the tumor/small motion side were significantly different between patients and controls, while the non-tumor/large motion side of the tongues were not and patients did not show increased or adaptive motion in the preserved side.

Two later PCA studies described motion of the whole midsagittal tongue, defined as in the present study. The first examined the 5 apical and 5 laminal controls used here. That study

used visual inspection to assess differences due to /s/-type, based on PC1+PC2 models of velocity fields. No categorically related differences were observed between /s/-types (Ding, et al, 2012). The second compared 5 controls and 5 patients, all apical /s/ producers. Once again, PC1+PC2 models of the midsagittal tongue were visually examined and did not show clear distinctions between subject groups. Focusing just on the central tongue region, near the location of the scar, greater uniformity of motion was observed visually in PC models for patients, presumably due to scar rigidity (Gallagher, et al, 2012). Neither of these studies examined the tongue tip specifically, however, where notable differences might occur due to /s/-type and subject groups. All the above studies, motivated the present study which uses 19 subjects: 5 apical controls, 5 laminal controls, 4 apical patients, 5 laminal patients and focuses on tag patterns in the tongue tip.

In the present study PCA is first used to quantify motion patterns of the whole tongue during /s/ in the control subjects. The goal is to determine whether known differences in the tongue tip between /s/-types would be propagated into the whole tongue due to the tongue's muscular architecture and volume preservation. Thereafter, the tongue tip-blade region is the main focus of this study for several reasons. First, the tip-blade of the tongue is the end-effector that produces the /s/ sound (Goldstein, Byrd and Saltzman, 2006). Second, the /s/ sound is challenging for glossectomy patients after surgery (Nicolletti, et al, 2004). Third, the tongue tip and the tongue body are somewhat independent in their control, so it may be that /s/-type differences are very local (Elfring, Boliek, Seikaly, Harris, Rieger, 2012, Kuruvilla, Green, Yunusova, Hanford, 2012).

Three hypotheses have been formulated for these data. The first proposes that for the apical /s/, the whole-tongue will lower or remain steady to facilitate the elevation of the tip. In contrast, the whole-tongue will move upward during laminal /s/ as will the tongue blade and tip. This second pattern is consistent with a simpler motor program for laminal /s/, which is better managed by the patients. The second hypothesis is that in the tip-blade region, the apical /s/ will exhibit an internal tongue motion pattern that shows separate motion for the tip and the blade; the tip will elevate and the blade will lower. This pattern is consistent with activation of the anterior genioglossus (GGA). To facilitate bend in the tongue surface as required by a muscular hydrostat a lengthwise and a crosswise muscle must activate; the lengthwise muscle alone will only shorten the structure (Kier and Smith, 1985). For the laminal /s/gesture, the tip and blade should show an elevation pattern that reflects motion of the whole-tongue, or less elevation for the tip than the blade since it is farther from the body. The first and/or second PC, which account for the most variance, are expected to distinguish motion pattern variability between /s/ types. Such a result would support two different gestures for /s/ in the tongue tip-blade region. The third hypothesis proposes that patients and controls will use different tongue tip-blade gestures to produce the /s/. Apical patients are expected not to produce tongue tip elevation independent of the tongue blade. Laminal patients are expected to have tip-blade motion patterns similar to the controls.

2. Methods

2.1. Subjects

From an MRI database of control participants recruited from the Baltimore area by advertisement, five apical /s/ and five laminal /s/ producers were chosen. Similarly, five post-glossectomy participants with apical /s/ and five with laminal /s/ were chosen from an MRI database recruited from the Department of Oral and Maxillofacial Surgery at the University of Maryland and the Department of Otolaryngology-Head and Neck Surgery at Johns Hopkins Hospital by their surgeons. All participants were native American-English speakers with normal hearing.

The ten patients received surgical excision of lateral lingual tumours that were closed primarily. The surgical procedure requires removal of 1–1.5 cm of healthy tissue on all sides of the tumour. Thus every resection is 2–3 cm longer than the original tumour. Subjects 11,12,13,15,16, 19 had small size tumours, less than 2cm in the largest direction prior to surgery. Subjects 14, 17,18 had moderate size tumours, between 2–4 cm in the largest direction prior to surgery. Informed consent was obtained for all participants. All subjects were screened for normal hearing, and the controls had normal oral motor examinations. One patient (small size tumour) was later excluded because her data were poor in quality suggesting erratic motions while in the MRI machine. Thus nineteen subjects were used in the study, 5 apical controls, 5 laminal controls, 4 apical patients, 5 laminal patients. Of the four apical patients, two had moderate size tumours.

2.2. Speech Material

The speech task "a geese" was chosen for this study because it meets several criteria. First, it starts with a neutral tongue position "uh," so that the original tongue position is minimally inflected. Second, the tongue motion within the word "geese" is unidirectional (back to front), which reduces complexity of motion. Third, it contains the high vowel /i/; this decreases jaw motion, which increases tongue deformation. Finally, it takes less than 1 second to repeat, which allows the tongue motion to be collected before the tags fade. The data analyzed in the study began with the release of the /g/ and ended 2 frames after the first tongue-palate contact for /s/. This eliminated between-subject differences in the length of the "uh."

2.3. Tagged MRI Methodology

A tag-trigger algorithm was implemented on a Siemens 3T Trim Trio MRI system using a 12-channel head coil and 4-channel neck coil. The imaging parameters were 6 mm slice thickness, 6 mm tag spacing, 1.875 mm by 1.875 mm in-plane spatial resolution, 1 second recording time, 26 time-frames per second. Tagged MRI data were collected using Complementary Spatial Modulation of Magnetization (CSPAMM) tagging (Fischer, McKinnon, Maier, and Boesiger, 1993) and reconstructed using Magnitude Image CSPAMM Reconstruction (MICSR) (NessAiver & Prince, 2003). CSPAMM acquisition first acquires a cosine tag pattern and then a minus cosine tag pattern. Standard CSPAMM reconstruction subtracts these two images yielding a perfect cosine tag pattern (distorted by any motion that may occur, of course) regardless of how much the tags fade (Fischer et al., 1993). MICSR uses the same two acquisitions but combines them using only the magnitude data (without requiring their phase). In addition to avoiding the need for collection of phase data, MICSR has improved contrast-to-noise ratios over CSPAMM combination, especially at later times in the image sequences (NessAiver & Prince, 2003).

A MICSR dataset is composed of four data acquisitions. Two of them contain horizontal tags and two contain vertical tags; each tag direction is acquired twice, one with a cosine tag pattern and one with a minus cosine tag pattern. Each of these four acquisitions requires three repetitions per slice, in order to acquire adequate Fourier data for analysis. Thus for 7 sagittal slices there are 4 separate acquisitions each containing 21 repetitions of the task, with 3 intervening pauses (Parthasarathy, Prince, Stone, Murano and NessAiver 2007). The midsagittal plane of the tongue was chosen for these 2D measurements as it is considered the best single representative of the total motion. This is because for almost all lingual sounds the midsagittal slice represents the vocal tract airway, which resonates the speech sounds, whereas the sides of the tongue contact the palate and reflect its shape (Stone & Lundberg, 1996).

Speaker precision was optimized by training the subject to speak to the same metronome beat that is used in the scanner. Since each tagged MR image sequence is a combination of multiple repetitions, subject variability across repetitions will cause blurring when the images are combined. To maximize speaker precision, subjects were trained prior to the MRI scan to precisely repeat the speech tasks. Subjects were also trained to inhale and exhale at fixed places within each cycle to further improve task repetition. The training used a metronome with a 4 beat sequence (for 2 syllables, inhalation, exhalation) based on the work of Masaki, et al. (1999).

Speech recordings were made in the MRI scanner with a noise reduction fibre optic microphone (Optoacoustics, Ltd., Israel) with no metallic components. These data were used only to corroborate accuracy of phoneme segment breaks (Boersma & Weenink, 2013) and are not discussed further.

2.4. Principal Components Analysis (PCA)

PCA is a dimension reduction method that represents high-dimensional data as a set of new orthogonal variables called principal components (PCs). The central idea of PCA is to retain the variation present in a data set while reducing its dimensionality, which consists of a large number of interrelated variables. In this paper subjects are represented by models composed of the average velocity field plus the first two modes of variation of the velocity fields, that is the first two PCs. A detailed description of the PCA computation used here is found in Stone, et al. (2010).

Two PCA's were performed. The first PCA used data from the 10 controls (apical vs. laminal), and quantified the variance in velocity for the whole-tongue (midsagittal). The goal was to determine whether normal controls who produced apical vs. laminal /s/ utilized different global tongue motion patterns. The second PCA used data from all 19 subjects, and quantified the variance in the tongue tip. The goal was to determine whether groups (subject vs. patient) used different tip-blade behaviours and whether /s/ type interacted with group. The 19 subjects each used slightly different speech rates and the word length for "geese" was, therefore, between10 and 15 time-frames. In total there were 135 velocity fields used in the 10 subject tip-blade PCA.

2.5. Preprocessing of Data for PCA

2.5.1. Registration of multiple subject data using landmark points—PCA

requires corresponding tissue points in space and time to be used across subjects. That is, the pixels (tissue-points) to be compared must correspond across subjects in order to directly compare their motion, therefore, the subjects were registered to each other. The method used was a weighted landmark-based rigid registration that accounts for 2 translations and 2 rotations. More specifically, let $p^{R,k}$ be landmark points for the reference R, and $p^{T,k}$ be those for the template T. We wish to align $p^{R,k}$ to $p^{T,k}$ by a rigid transformation T such that $T(p^{R,k}) \sim p^{T,k}$ by minimizing the following landmark distance function:

$$D(u) = \sum_{k=1}^{9} w_k \| p^{T,k} - T(p^{R,k}) \|^2,$$

where $T(p^{R,k}) = p^{T,k} + u(p^{R,k})$ and w_k denotes the a weight applied to the *k*th point.

Nine landmark-based points were chosen on the surface of the tongue for each subject in time-frame 1 of the tagged MRI dataset to enclose the region of the tongue to be compared across speakers. Some of the landmarks were actual tissue points, such as the tongue tip, the

high point of the tongue, etc. Others were midway between two identified points, such as the point midway between the high point and the tongue tip (see Figure 1a). Labelling was done manually by the first author and two students under her direction using strict criteria defined previously in (Stone, et al., accepted, JSLHR). The nine points were tracked using HARP through the following 25 time-frames to determine their location at every moment in time and rule out mistracking. If mistracking occurred, adjacent points, usually deeper in the tongue, were selected until no mistracking occurred. For the whole-tongue analysis the landmark points were given equal weight and the resulting common region was centred within the tongue (Figure 1b, white region). However, this registration method contained little of the tongue tip for most subjects. Since the tip and blade are crucial for /s/ production, a second registration was performed to better align the region containing the tipplus-blade, hereafter 'tip-blade' using a weighted landmark-based registration. For the tipblade registration, the three anterior points were weighted ten times more heavily than the others and the resulting common region (white) was more anterior (see Figure 1c, white region). The tissue points used in the tip-blade analysis were those in the common region that were also within the blue contour (Figure 1d). Once the common tissue points were identified in the whole-tongue and the tip-blade registration, each subject's dataset was transformed back into its original orientation so that the data were no longer transformed, and the tissue points were corresponding across subjects.

2.5.2. Choice of Velocity Field—PCA also requires the tissue points to have an identical relationship in time across subjects. It would have been optimal to use all the time-frames between /g/ and /s/, and to include time in the analysis. However, the subjects spoke at different rates of speed and thus had different numbers of frames between the /g/ and /s/, thus their rates would affect any overall patterns seen. Therefore, a single time-frame was decided upon for the analysis. However, the choice of frame required thought. Slower speakers might use a longer, slower period of maximal velocity than faster ones. This would affect maximum velocity during the pre-/s/ motion, as rate of speech would affect maximum velocity across speakers independently of velocity pattern. Thus, the decision was made to use the final time-frame in the motion toward /s/, that is, the velocity field that ended in contact between the tongue and palate for the /s/-constriction. The /s/ contact frame was determined using visual inspection and group discussion between several of the co-authors to be sure of consensus. This velocity field was considered to be the least influenced by rate of speech.

2.6. Statistical Comparison of PC Loadings

Although this data set is small and thus not ideal for parametric statistics. These analyses are meant to shed light on the relationships between variables, and to be used to indicate the direction for future studies. Two types of analysis were performed on the PC results. Analyses of Variance (ANOVA) were used to determine whether any of the individual PC's distinguished between subject group and /s/-type. Two-way ANOVAs were performed on the first 17 PC's to determine whether /s/-type or subject group had significantly different loadings on each PC. Then one-way ANOVA's were used as post-hoc tests to determine the significance of subject group or /s/-type on the various PC's. Linear Discriminant Analysis (LDA) was performed on the first 17 PC's to determine how many were needed to perfectly categorize subjects into group and /s/ type.

3.0. Results

All the velocity fields comprising the word "geese" were input to the PCA's performed on the whole-tongue (10 controls, 125 fields) and tip-blade data (10 controls, 9 patients, 251 fields). In the whole-tongue analysis the first two PC's explained 67.8% of the variance, and

the first 15 PC's explained 90.2%. In the tip-blade analysis the first two PC's explained 79.3% of the variance, and only the first 5 PC's were needed to explain 90.6%. The smaller number of PC's indicated that the motion patterns in the tip-blade were more uniform than in the tongue body even with the addition of patient data. The vector fields of the first two PC's, which represent the largest patterns of variance, are depicted in Figure 2 for the whole-tongue, and in Figure 3 for the tip-blade. In both figures the middle image depicts the average velocity field; the apex of the tongue is on the left. The average velocity field need not have a meaningful shape as it sums motion in many different directions. The principal direction of variance in tongue motion appears in the models to the right and life of the average, on the centre row, which are made by adding and subtracting 1 standard deviation of PC 1. PC 1 explained a third of the variance in the whole-tongue (36.1%), and over half the variance in the tip-blade region (58.2%). The second principal direction of variance, modelled by adding and subtracting 1 standard deviation of PC 2 to the average, is shown in the middle column, PC 2 explained 31.7% in the whole-tongue and 21.2% in the tip-blade. The PC models also need not have a meaningful shape as they are statistical constructs of the directions of variance, not dependent on physical motions. However, sometimes they do represent reasonable motion pattern components as are seen in these data.

3.1. Representation of motion by PCs 1 and 2

Figure 2 shows that in the whole-tongue data, PC's 1 and 2 described variance in the horizontal and vertical directions, respectively. Oblique motion in the upper tongue was also included: up/down for PC 1 (row 2) and back/front for PC 2 (column 2). When PC's 1 and 2 were added, they described variance as mostly upward or downward in direction (upper right, lower left corner), and convergence or divergence between the upper and lower tongue (upper left, lower right corner).

Figure 3 depicts the average and PC models for the tip-blade region only. PC 1 indicated that the primary direction of variance was forward/upward vs. downward/backward (see Row 2). PC 2 showed the secondary direction of variance to be upward/backward vs. downward/forward (see Col. 2). Combinations of the PC 1×2 models indicated that the primary directions of motion variability in the tip-blade covered the four major directions, up, down, back, front, with little local deformation (see Figure 3, the four corners).

In sum, tip-blade motion tended to be more unidirectional than motion of the whole-tongue and its motion variance was better represented by PC 1and 2 (79.4% vs. 67.8% of the variance) than the whole-tongue model. The PC models are based on the entire word 'geese', however. To focus the analyses on the /s/ motion specifically, ANOVA's were performed on the chosen velocity field for each subject, that is, the final one leading into the palatal constriction for /s/.

3.2. Statistical Analyses of /s/ velocity fields

3.2.1. Whole-Tongue Analysis: ANOVA—In the whole-tongue data, one-way ANOVA's were performed on the first 10 PC's, which together explained 87% of the variance, to see if any distinguished between the apical and laminal speakers. None were significant, and no further analyses were performed. Figure 4 shows the loadings of the apical (square) and laminal (circle) speakers on PC 1 and 2. /s/-types were not categorized by the loadings on either PC. Additionally, the subjects all loaded positively on PC 2 and negatively on PC 1 indicating backward and upward motion of the whole-tongue for all subjects into the /s/. Further examination was made of the velocity fields prior to the key velocity field. This examination showed that for at least 3 velocity fields preceding the one used in the ANOVA, 8 of the 10 controls had negative loadings on PC 1 and positive loadings on PC 2. Only one subject changed loading on PC1 at the target velocity field.

Stone et al.

These loadings indicate that after the rapid forward motion of the midsagittal tongue for the /g/ and /i/, the tongue tended to move backward and upward into the /s/ for both /s/-types.

3.2.2. Tip-blade Analysis: ANOVA—Backward and upward motion into the /s/ was confirmed in the tip-blade data. Two-way ANOVA's were performed on the first 17 PC's with subsequent one-way ANOVA's used post hoc to distinguish whether subject group, /s/-type, or both had significantly different motion patterns (see Table 1). PC's 1 and 2 showed a significant interaction between subject group and /s/ type. The one-way ANOVA's showed that both PC's distinguished between subject group in the apical /s/ data. PC 1 also distinguished between /s/ type in the patients, and PC2 was almost significant between /s/ type in the controls. These effects occurred primarily because of the very tight clustering of the patients with apical /s/. PCs 4 and 5 showed a significant difference for groups irrespective of /s/ type. PC 4, 7 and 10 distinguished between /s/ types and groups. They also accounted for most of the variance.

Figure 5 shows that fifteen of the 19 subjects loaded negatively on PC 1, indicating down/ back motion of the tip-blade. Nine of 10 controls (black) and 2 patients (gray) also loaded positively on PC 2, indicating up/back motion. The combination of +PC2 and -PC1 produces primarily backward motion with a slight downward angle. The four apical patients (gray squares) formed a tight cluster with the highest loadings on -PC1 and virtually no PC2; their tip-blade moved uniformly down/back. Further examination of PC2 shows differences between the controls and the laminal patients. The 5 laminal patients (gray circles) were distributed on PC 2 with positive or negative loadings. All the controls but one loaded positively on PC2 (black tokens). The apical controls (black squares) tended to have slightly higher loadings on +PC2 than the laminal controls (black circles).

3.2.3. Subject Categorization: Linear Discriminant Analysis—The LDA was used to determine how many PC's were needed to correctly separate the subject groups and /s/-types in the tip-blade data set. Results showed that when apical and laminal speakers were combined, it took 14 PC's to correctly categorize patients and controls into separate groups. Similarly, when patients and controls were combined, even 17 PC's still misclassified 2 subjects' /s/-type. When the data were separated by /s/-type, however, categorization into patient vs. control groups was achieved with just PC 1 for the apical subjects and the first 6 PC's for the laminal ones. Similarly, when the data were separated by subject group, categorization of subjects into apical and laminal /s/-type was achieved by the first 3 PCs for patients and the first 5 PC's for controls.

4.0. Discussion

The typical description of /s/ production in the midsagittal plane, indicates that the tongue body develops a groove at midline along its entire length, and the tongue tip or tongue blade elevates to the alveolar ridge to form a circular constriction. This shape funnels a narrow airstream onto the upper or lower incisors causing the sibilant noise characteristic of /s/ (cf. MacKay, 1991). The tongue behaviours in the present study, however, were heavily influenced by the other sounds in the word "geese." The tongue began the word by moving forward from /g/-to-/i/ very rapidly and sometimes downward. At the release of /g/ (timeframe 1–2) the loadings on PC 1 were typically very high positive values (whole-tongue loading (mean) = 18.93, range=4.4 to 49.3; tip-blade (mean) = 11.10, range=2.2 to 26.8) (these data are not depicted). These numbers contrast starkly with the loadings of the /s/ onset time-frame on PC 1 (whole-tongue (mean)= -3.99, range=-9.5 to 0; tip-blade (mean)=

-2.45, range=-5.4 to 1.8). Thus, the primary motion from /i/-to-/s/ was clearly backward for both the tip-blade and whole-tongue, and not reflective of the canonical form.

4.1. Whole tongue vs. tongue tip: Control Subjects

The first hypothesis predicted that the whole-tongue, in order to support the different tipblade manoeuvres, would move downward or remain immobile for apical /s/ and upward for laminal /s/. Results showed that all subjects moved the whole-tongue upwards (+ PC2) and backwards into the /s/ (– PC1). Thus, the whole-tongue patterns were consistent with the same motion pattern for both /s/ types.

Another component in the first hypothesis was that the tip-blade motion pattern would be similar to the whole-tongue for the laminal, but not the apical /s/. While not exactly a part-whole comparison, a relationship was postulated between the tip-blade and whole-tongue because tongue motion is based on principals of volume preservation. Speech gestures are expected to involve more than just the end-effector region, and muscles in the tongue body may be needed to correctly position the tip. Moreover, some of the anterior tongue was included in the whole-tongue data (Figure 1). However, in the whole-tongue analysis, none of the 10 PC's were significantly different between the apical and laminal controls. Similarly, in the tip-blade analysis there were no PC's that had a significant main effect for /s/ type, though it is possible that PC 2 (p=0.059) would distinguish /s/-type if more subjects were included.

4.2. One gesture vs. many for /s/: Control subjects

One assumption of this study was that the two /s/ gestures, apical and laminal, would reflect two different tongue gestures. The second hypothesis specified that in controls, the apical /s/ would have an upward direction of motion for the tip and a downward one for the blade, whereas the laminal /s/ would elevate both the tip and blade. The results did show that the apical /s/ controls were more likely to have two directions of motion (higher positive loading on PC 2) than the laminal ones (Figure 5), although the directions were not those predicted. Moreover, Figure 5 shows that 9 of the 10 controls loaded positively on PC2 (upward / backward motion), indicating that the difference between apical and laminal /s/ was more likely one of degree, than an entirely different gesture. Thus, neither the whole-tongue nor the tip-blade data distinguished between /s/-type in the control subjects based on the PC1+2 models or the ANOVA's.

4.3. Tongue tip differences: Patients vs. controls

The third hypothesis was that apical patients would not show an independent tongue tip and blade motion, but apical controls would. The major difference was found to be direction of motion, not division into a tip vs. blade region. The apical patients had zero or negative loadings on PC 2. In addition, there were clear differences between apical patients and almost all other subjects. The ANOVA's in Table 1 show that for the patients, PC1 alone was sufficient to distinguish apical from laminal /s/ producers. Similarly, when looking only at apical speakers, PC 1 and 2 both distinguished between patients and controls. These results occurred because the apical patients were highly uniform in their motion. This uniformity is quite visible in the PC1 × PC 2 loadings (Figure 5), which show that for apical patients (gray squares) the tip-blade had absolutely no elevation, and considerable back/ down motion. Only one control showed this pattern. Downward motion of the tip-blade unit also was found in 2 of the laminal patients. Thus, these patient gestures are consistent with difficulty executing sufficient tip-blade elevation. The jaw probably assists in accurately positioning the tip-blade for these subjects.

The other three laminal patients loaded positively on PC2, as did 9 of the 10 controls (see Figure 5). A positive loading on PC 2 indicated upward motion of the tip and backward motion of the blade, as can be seen in the average+PC2 model. This PC, unlike PC 1, distinguished between tip direction and blade direction (Figure 3, top centre). Interestingly, two of the 3 laminal patients who loaded positively on PC2 had moderate size tumours (2–4 cm before excision). Two the four apical patients also had moderate size tumours. Thus tumour size was not a limiting factor in the ability to elevate the tip-blade or in the use of apical vs. laminal /s/.

4.4. Limitations of the study

Several methodological issues limit generalizations made from this study. The first is the inclusion of only one velocity field per speaker. The choice of velocity field relative to the onset of the /s/ was carefully considered to optimize its choice. However, speech gestures occur over time and a single moment in time cannot fully represent the entire behaviour. The second issue is the use of a single phonetic word and context. The observations made in this paper about tongue motion cannot be conclusively determined until other directions of motion into the /s/ are examined. The third is that the effects of oral morphology on speech sounds need to be studied more comprehensively. An ongoing study is examining the effects of other oral morphological features, such as palate shape, on the phonetic realization of speech sounds.

4.5. Clinical utility of results

There are several findings in this study that have a direct application to clinical treatments. First is the clear difference between the realization of a speech gesture in a phonetic context and the expected canonical form. Since the tongue moves backwards and also changes shape between /i/ and /s/, the elevation of the tip-blade region is seen as a minor component in an essentially backward tip motion. Therapeutic strategies typically start with imitation of canonical form and progress into more complex morphological structures using /s/. These results indicate that differences in /s/ realization due to context are quite large and need to be considered in the ordering of treatment tasks. Second, the present data did not show a difference in tongue tip direction for apical vs laminal /s/. This may implicate other factors in production of /s/ type, such as oral morphology, and differences in degree of tip elevation. Third, it had previously been found (Stone et al, 2013) that glossectomy patients were more likely to produce laminal /s/ than apical /s/. This study added some additional information indicating that those patients that do produce apical /s/ are very similar and are uniformly devoid of tongue tip elevation. Therapeutic strategies should be sensitive to the patient's preference for a specific /s/-type and realize that it will be realized differently than in nonglossectomy speakers. In addition, when working with glossectomy patients, slight differences in tumor size do not create significant differences in abilito to produce speech gestures.

5.0 Conclusions

The apical and laminal /s/, following /i/, were not distinguished by motion patterns in the control subjects. They mostly used backward motion patterns with varying degrees of added upward or downward direction. This was true both for the entire tongue and the tongue tipblade area. Glossectomy patients who produced an apical /s/, however, produced very similar and specific motion patterns, which differed from laminal patients and all controls by having no upward tip component, only downward/backward motion. This strong downward direction in apical patients may reflect difficulty maintaining the tip at the proper distance from the palate.

Acknowledgments

Portions of this work were funded by NIH R01 CA133015.

References

- Atsumi T, Miyatake T. Morphometry of the degenerative process in the hypoglossal nerves in amyotrophic lateral sclerosis. Acta Neuropathologica. 1987; 73:25–31. [PubMed: 3604571]
- Axel L, Dougherty L. Heart wall motion: Improved method of spatial modulation magnetization for MR imaging. Radiology. 1989; 172:349–350. [PubMed: 2748813]
- Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version 5.3.48. 2013 from http://www.praat.org/.
- Brunner J, Fuchs S, Perrier P. On the relationship between palate shape and articulatory behaviour. J. Acoust. Soc. Am. 2009; 125:3936–3949. [PubMed: 19507976]
- Dart S. Articulatory and acoustic properties of apical and laminal articulations. UCLA Working Papers in Phonetics no. 79. 1991
- Dart SN. Comparing French and English coronal consonant articulation. J. Phonetics. 1998; 26:71–94.
- Ding, C.; Woo, J.; Stone, M.; Chen, H. Variance in tongue motion patterns during the production of /s/. Presented at the Acoustical Society of America Meeting; Oct. 2012; Kansas City, MO. 2012.
- Elfring TT, Boliek CA, Seikaly H, Harris J, Rieger J. Sensory outcomes of the anterior tongue after lingual nerve repair in oropharyngeal cancer. Journal of Oral Rehabilitation. 2012; 39:170–181. [PubMed: 21923892]
- Fischer SE, McKinnon GC, Maier SE, Boesiger P. Improved myocardial tagging contrast. Magnetic Resonance in Medicine. 1993; 30(2):191–200. [PubMed: 8366800]
- Gallagher, C.; Woo, J.; Stone, M.; Chen, H. A PCA comparison of normal and glossectomy movement patterns in the production of /s/. Presented at the Acoustical Society of America Meeting; Oct. 2012; Kansas City, MO; 2012.
- Goldstein, L.; Byrd, D.; Saltzman, E. The role of vocal tract gestural action units in understanding the evolution of phonology. In: Michael, A. Arbib, editor. Action to Language via the Mirror Neuron System. Vol. Chapter 7. Cambridge University Press; p. 215-249.Published by Cambridge University Press.
- Heller KS, Levy J, Sciubba JJ. Speech patterns following partial glossectomy for small tumours of the tongue. Head Neck. 1991; 13:340–343. [PubMed: 1869436]
- Kier W, Smith K. Tongues, tentacles and trunks: the biomechanics of movement in muscularhydrostats. Zoological Journal of the Linnean Society. 1985; 83:307–324.
- Kuruvilla MS, Green JR, Yunusova Y, Hanford K. Spatiotemporal Coupling of the Tongue in Amyotrophic Lateral Sclerosis. Journal of Speech, Language, and Hearing Research. 2012; Vol. 55:1897–1909.
- Liu, X.; Murano, EZ.; Stone, M.; Prince, JL. HARP tracking refinement using seeded region growing; Proceedings of 4th IEEE International Symposium on Biomedical Imaging; 2007. p. 372-375.
- Mackay, I. Phonetics: the science of speech production. 2nd edition. Boston: Allyn and Bacon, publishers; 1991.
- 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.
- NessAiver M, Prince JL. Magnitude Image CSPAMM Reconstruction (MICSR). Magnetic Resonance in Medicine. 2003; vol. 50:331–342.
- Narayanan SS, Alwan AA, Haker K. An articulatory study of fricative consonants using magnetic resonance imaging. J. Acoust. Soc. Am. 1995; 93:1325–1335.
- Nicoletti G, Soutar DS, Jackson MS, Wrench AA, Robertson G. Chewing and swallowing after surgical treatment for oral cancer: Functional evaluation in 196 selected cases. Plast Reconstr Surg. 2004; 114(2):329–338. [PubMed: 15277796]
- Nicoletti G, Soutar DS, Jackson MS, Wrench AA, Robertson G, Robertson C. Objective Assessment of Speech after Surgical Treatment for Oral Cancer: Experience from 196 Selected Cases. Plast. Reconstr. Surg. 2004; 113:114–125. [PubMed: 14707629]

- Osman NF, Kerwin WS, McVeigh ER, Prince JL. Cardiac motion tracking using CINE harmonic phase _HARP_ magnetic resonance imaging. Magn. Reson. Med. 1999; 42:1048–1060. [PubMed: 10571926]
- Parthasarathy V, Prince JL, Stone M, Murano E, NessAiver M. Measuring tongue motion from tagged Cine-MRI using harmonic phase (HARP) processing. J. Acoust Soc Amer. 2007; 121(1):491–504.
- Reichard, R.; Stone, M.; Woo, J.; Romberg, E.; Murano, EZ.; Prince, JL. Proceedings of Acoustics, 2012. Hong Kong, China: 2012 May 14–18. Motion of apical and laminal /s/ in normal and postglossectomy speakers. 2012
- Shapiro BL, Redman RS, Gorlin RJ. Measurement of normal and reportedly Malformed Palatal Vaults. I. Normal Adult Measurements. J. Dent. Res. 1963; 42(4):1039–1042. [PubMed: 14046509]
- Stone M, Lundberg A. Three-dimensional tongue surface shapes of English consonants and vowels. J. Acoust. Soc. Am. 1996; 99:3728–3737. [PubMed: 8655804]
- Stone M, Liu X, Chen H, Prince JL. A preliminary application of principal components and cluster analysis to internal tongue deformation patterns. Computer Methods in Biomechanics and Biomedical Engineering. 2010; 13(4):493–503. [PubMed: 20635265]
- Stone M, Rizk S, Woo J, Murano EZ, Chen H, Prince JL. Frequency of Apical and Laminal /s/ in Normal and Post-glossectomy Patients. Journal of Medical Speech Language Pathology. 2013; 20(4):106–111.
- Stone M, Langguth JM, Woo J, Chen H, Prince JL. Tongue Motion Patterns in Glossectomy and Normal Speakers: A Principal Components Analysis. J. Speech Language Hearing R. 2013 accepted.
- Unay D, Ozturk C, Stone M. Single syllable tongue motion analysis using tagged cine MRI. Computer Methods in Biomechanics and Biomedical Engineering. :1–12. (2012-ifirst copy). doi: 10.1080/10255842.2012.723697
- Ventura S, Freitas D, Tavares JMRS. Imaging of the vocal tract based on Magnetic Resonance techniques. Computer Vision, Imaging and Computer Graphics. Theory and Applications Communications in Computer and Information Science. 2010; Volume 68:146–157.
- Yunusova Y, Rosenthal JS, Rudy K, Baljko M, Daskalogiannakis J. Positional targets for lingual consonants defined using electromagnetic articulography. J. Acoust. Soc. Am. 2012; 132(2):1027– 1038. [PubMed: 22894223]
- Wozniak W, Young PA. Further observations on human hypoglossal nerve. Anatomischer Anzeiger. 1969; 125:203–205. [PubMed: 5365186]

Stone et al.



Figure 1.

PCA preprocessing. (a) Nine landmark points and their motion paths tracked through 26 time-frames. (b) Overlay and unweighted alignment of 135 time-frames with equal weight given to all 9 landmark points. Common region is central white area. (c) Overlay and weighted alignment of 251 time-frames in the three anteriormost points. The common region is more anterior (d) The common tip-blade region contains the tissue points in the common region region encircled by the blue line.

Stone et al.



Figure 2.

The average velocity field (center) and models of the average +/- loading of 1 std dev of PC1 and PC 2 for the whole-tongue. Tip on left.

Stone et al.



Figure 3.

The average velocity field and models of the average +/- loading of 1 std dev of PC1 and PC 2 for the tip-blade. Tip on left.



Figure 4.

Whole-tongue data showing velocity field loadings on PC 1 and PC2 for controls. Squares are apical /s/-types; circles are laminal.

Stone et al.



Figure 5.

Tip-blade data showing velocity field loadings on PC 1 and PC2. Controls are black; patients are gray. Squares are apical /s/-types; circles are laminal.

Table 1

Stone et al.

	% variance	58.2			21.2			3.6		2.5		1.4	9.0
	P value	0.016	0.01	0.021	0.04	0.000	0.059	0.01	0.0001	0.02	0.02	0.02	0.01
	F value	7.27	12.15	8.85	3.51	30.27	4.82	8.34	106.83	7.04	8.05	8.3	11.84
	Population	interaction	apicals	patients	interaction	apicals	controls	all /s/ data	apicals	all /s/ data	laminals	apicals	apicals
)	ΛΙ	group x /s/	dnoıß	s-type	group x /s/	groups	s-type	sdnorg	sdnorg	sdnorg	sdnorg	sdnorg	groups
	AVOVA	Two-Way	One-Way	One-Way	Two-Way	One-Way	One-Way	One-Way	One-Way	One-Way	One-Way	One-Way	One-Way
	PC	1			2			4		5		7	10