Research Article

Atlas-Based Tongue Muscle Correlation Analysis From Tagged and High-Resolution Magnetic Resonance Imaging

Fangxu Xing,^a Maureen Stone,^b Tessa Goldsmith,^c Jerry L. Prince,^d
Georges El Fakhri,^a and Jonghye Woo^a

Purpose: Intrinsic and extrinsic tongue muscles in healthy and diseased populations vary both in their intra- and intersubject behaviors during speech. Identifying coordination patterns among various tongue muscles can provide insights into speech motor control and help in developing new therapeutic and rehabilitative strategies.

Method: We present a method to analyze multisubject tongue muscle correlation using motion patterns in speech sound production. Motion of muscles is captured using tagged magnetic resonance imaging and computed using a phase-based deformation extraction algorithm. After being assembled in a common atlas space, motions from multiple subjects are extracted at each individual muscle location based on a manually labeled mask using high-resolution magnetic

resonance imaging and a vocal tract atlas. Motion correlation between each muscle pair is computed within each labeled region. The analysis is performed on a population of 16 control subjects and 3 post-partial glossectomy patients.

Results: The floor-of-mouth (FOM) muscles show reduced correlation comparing to the internal tongue muscles. Patients present a higher amount of overall correlation between all muscles and exercise en bloc movements.

Conclusions: Correlation matrices in the atlas space show the coordination of tongue muscles in speech sound production. The FOM muscles are weakly correlated with the internal tongue muscles. Patients tend to use FOM muscles more than controls to compensate for their postsurgery function loss.

Understanding the relationship between the anatomy and functions of the human tongue has been a widely studied topic in the investigation of oromotor behaviors such as speech, swallowing, and breathing. The tongue is, however, challenging to study because of its highly complex muscular architecture and motion patterns (Abd-el-Malek, 1955; Stone et al., 2018; Takemoto, 2001), in which interdigitated muscles are responsible for the rapid yet precise deformations that form the various tongue shapes necessary in human speech (Kent & Read, 2002;

Kier & Smith, 1985). As part of the study of motor control, quantifying the cooperation and interaction between these tongue muscles has been a major focus. Furthermore, in order to improve therapeutic and rehabilitative processes, oral surgeons and speech-language pathologists have strived to understand the compensation strategies often observed in patients with unintelligible speech, such as postglossectomy patients (Bressmann, Jacobs, Quintero, & Irish, 2009; Bressmann, Sader, Whitehill, & Samman, 2004; Nicoletti et al., 2004; Pauloski et al., 1998; Rastadmehr, Bressmann, Smyth, & Irish, 2008) and patients with amyotrophic lateral sclerosis (Perry, Martino, Yunusova, Plowman, & Green, 2018). These compensation strategies are closely related to the patients' altered tongue anatomy, their muscle functions, and the phonological rules of the target language. The aim of this work was to assess and compare muscle coordination patterns between controls and patients in a normalized space to shed light on normal and compensated motion patterns in speech sound production.

In the past decades, tagged magnetic resonance imaging (tMRI) has been widely used in the structural and functional studies of the human tongue (Parthasarathy, Prince, Stone, Murano, & NessAiver, 2007; Stone et al., 2001).

^aGordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston

^bDepartment of Neural and Pain Sciences, University of Maryland Dental School, Baltimore

^cDepartment of Speech, Language and Swallowing, Massachusetts General Hospital, Boston

^dDepartment of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD

Correspondence to Fangxu Xing: fxing1@mgm.harvard.edu

Editor-in-Chief: Julie Liss

Editor: Jack Jiang

Received December 14, 2018

Revision received March 25, 2019

Accepted May 10, 2019

https://doi.org/10.1044/2019_JSLHR-S-18-0495

Disclosure: The authors have declared that no competing interests existed at the time of publication.

The key technique of tMRI is to overlay a gridded or striped tag pattern in the image by magnetic field manipulation (NessAiver & Prince, 2003; Zerhouni, Parish, Rogers, Yang, & Shapiro, 1988). The tag pattern deforms with the tongue tissue in speech. Its information can be extracted in post-processing to recover four-dimensional (4D) motion fields (three-dimensional [3D] space and time; Liu et al., 2012; Xing et al., 2017). Based on this technique, besides improving the accuracy of tongue motion calculation, most previous works have focused on analyzing a population's collected motion characteristics. For example, to study tongue cancer patients' unique speech patterns, principal components have been extracted (Stone, Liu, Chen, & Prince, 2010) and multiple analyses have been carried out to compare the motion fields of a group of post-partial glossectomy patients to that of a group of healthy controls (Stone, Langguth, Woo, Chen, & Prince, 2014; Xing et al., 2016). To localize internal tongue motion, previous studies have divided the tongue into a smaller set of coordinated motion patterns—that is, functional units—to reveal parts of the tissue that are most likely to work in coordination (Woo, Stone, et al., 2017; Woo, Prince, et al., 2018). In terms of studying individual tongue muscles, there have been works that analyzed strain in tongue muscle fiber directions, showing active or passive contractions of individual muscles during speech (Xing et al., 2018; Xing, Ye, Woo, Stone, & Prince, 2015).

However, there exists a common limitation in previous reported analyses: the difficulty of statistically achieving multisubject motion field analysis without spatial and temporal alignment among different subjects. First, since the acquired tMRI slices only contain in-plane information, 3D motion is estimated by combining slices from multiple orientations in each individual subject's space. This process does not guarantee a spatial alignment between different subjects due to their varying tongue shapes and their positions in the scanner. Second, the variation of speaking rate between different subjects brings in more inconsistency, causing a temporal misalignment between estimated motion instants. To deal with the inconsistency, most previous analyses worked around this problem by focusing on either subject-specific analysis or global quantities without alignment. For example, the postsurgery principal component analysis used the mean motion in each quadratic section of the tongue without aligning all subjects (Xing et al., 2016), the functional unit analysis used subject-specific point locations instead of 4D motion fields (Woo, Stone, et al., 2017), and the strain analysis studied each individual subject's motion field instead of the combined multisubject statistics (Xing et al., 2015). As a result, to study the coordination between internal tongue muscles statistically, the most essential challenge is to construct a common atlas space in which alignment of multisubject motion fields can be found.

Previously, a 3D structural atlas of the vocal tract from a collection of high-resolution magnetic resonance images has been developed and validated in order to statistically represent the anatomy of the tongue (Stone et al., 2018; Woo et al., 2015). The vocal tract atlas also provides a normalized space in which intrinsic and extrinsic tongue muscles are

identified and a variety of healthy subjects and diseased populations can be mapped and compared. Following a similar rationale, in this work, we propose the construction of an atlas of 4D motion fields that statistically represents the function of the tongue. We use this dynamic atlas to achieve a correlation analysis between the motion of various internal tongue muscles in speech sound production in order to reveal their cooperation patterns. The method is established on top of an existing anatomical cine tongue atlas and a set of 4D tongue motion estimates from a number of healthy control subjects. With the aid of the deformation fields obtained during the cine atlas construction, individual subject's motion fields are deformed into the atlas space and aligned with the manually labeled high-resolution magnetic resonance imaging (hMRI) muscle masks. Finally, correlation matrices of muscle motion are computed between all labeled regions. We compared the result of 16 healthy controls and three post-partial glossectomy patients. Besides revealing the unique motion patterns of the floor-of-mouth (FOM) muscles from the other tongue muscles, patients' unique muscle motion patterns are also studied and discussed. The method shows the capability of quantitative muscle behavior analysis starting with simple speech sound production, with the ultimate goal of extending the scope of study to cover more complex muscle structures and help the understanding of natural speech production.

Method

The overall process of the proposed analysis is illustrated in Figure 1. The input is multisubject tMRI and hMRI data. The output is a collection of correlation matrices between muscles aligned in the atlas space. The details for each specific method are described below.

Figure 1. Workflow of the proposed method. High-res MRI = high-resolution magnetic resonance imaging; Constrn. = construction; Corr. = correlation.

Data Acquisition

A data set of 16 healthy controls and three post-partial glossectomy patients was used in the analysis. All subjects read and signed a consent form and the Health Insurance Portability and Accountability Act form before the study, and the entire protocol was approved by the University of Maryland Baltimore Internal Review Board. All three patients had T1N0M0 tumors that were removed with partial glossectomy and wounds closed by sutures (T1 primary). Using the TNM Classification of Malignant Tumors system (Union for International Cancer Control), T1 means the tumor was small and not greater than 2 cm in the largest dimension, though resection includes a 1- to 1.5-cm margin of clean tissue around the tumor. There were no active nodes or metastasis. The first two patients had tumors on the right side of the tongue, and the third patient had a tumor on the left side. No radiation therapy or chemotherapy was performed on any of the patients, and their tongue volume is minimally altered due to surgery impacts. All subjects were instructed to perform a speech task by pronouncing the phrase “a souk” (International Phonetic Alphabet: /əˈsuk/). The phrase was specifically designed to start with a centralized tongue position /ə/, moving prominently forward into /s/, and ending with a prominent upward motion into /k/. In the MRI scanner, all subjects repeated the speech phrase to a metronome. There were three repetitions per image slice (synchronized by the repeating rhythm of the metronome), each repetition for a second of data collection. Pauses were made between different slice orientations. The magnetic resonance tagging sequence (NessAiver & Prince, 2003) was triggered by the metronome at every speech cycle, precisely synchronizing the acquisition of the tongue motion. All scanning sessions were carried out on a Siemens 3.0T Tim Trio system (Siemens Medical Solutions), with a 12-channel head coil and a four-channel neck coil. Other imaging parameters are listed in Table 1. An example of an acquired sagittal slice tagged in both horizontal and vertical directions is shown in Figures 2b and 2c.

4D Motion Estimation

To estimate 4D motion from tMRI data, we used the phase vector incompressible registration algorithm (PVIRA; Xing et al., 2017). PVIRA is a phase-based deformation extraction algorithm under a diffeomorphic image registration framework. Specifically, with the input being a set of two-dimensional (2D) slices from three cardinal orientations (axial, sagittal, and coronal), PVIRA uses cubic B-spline to interpolate their intensity values onto a denser

3D grid. Then, a harmonic phase (HARP) filter is applied to yield phase volumes at each time frame (Osman, McVeigh, & Prince, 2000). Finally, PVIRA applies a demon-based image registration (Mansi, Pennec, Sermesant, Delingette, & Ayache, 2011) on these phase volumes to find the motion estimate while preserving both incompressibility and inverse consistency. For each subject at each time frame, PVIRA yields a dense 3D motion field in the subject’s independent space (see Figure 2d).

For any subject labeled by s , we denote the PVIRA estimate as $\mathbf{u}_{s,t}(\mathbf{X})$ at time frame t , $t = 1, 2, \dots, 26$. At the undeformed time frame, $\mathbf{X}$ is the tissue point coordinates in the voxel grid. Therefore, each subject yields a 4D motion sequence quantified by 26 such motion fields. Examples of motion fields of a control and a patient at two time frames /s/ and /k/ are shown in Figures 3a, 3c, 3e, and 3g. We use $\phi_{s,t}$ to denote the corresponding deformation between time frame t and the first (undeformed) time frame. Mathematically, we have

$$\phi_{s,t}(\mathbf{X}) = \mathbf{X} + \mathbf{u}_{s,t}(\mathbf{X}). \quad (1)$$

This equation shows the motion of tissue point coordinates $\mathbf{X}$. At the time frame t , they move to a new location through motion field $\mathbf{u}_{s,t}(\mathbf{X})$. This material frame definition is called a *Lagrangian framework* (Sedov, 1997). It roots every motion field $\mathbf{u}_{s,t}(\mathbf{X})$ in the undeformed frame $\mathbf{X}$, as opposed to the *Eulerian framework* where the motion fields are measured in a deformed frame. In this work, we choose to apply the Lagrangian framework over the Eulerian framework, because any quantity computed over time is always mapped and displayed on the undeformed frame so that the tongue appears “motionless” (Woo, Xing, et al., 2017). Since hMRI will be used later for muscle segmentation and it was acquired in a static space, it is necessary to also use a motionless framework for the estimated motion fields. In such a way, a connection between the tMRI space and the hMRI space can be found by matching the two motionless data sets with image registration (Vercauteren, Pennec, Perchant, & Ayache, 2009).

4D Motion Atlas Construction

Now, we seek spatial and temporal alignment between all subjects’ motion fields. First, we address the time alignment issue caused by inconsistent speaking rates between different subjects. Since PVIRA’s estimation results have certain physical properties, such as inverse consistency and incompressibility, it is not physically meaningful to

Table 1. Tongue motion magnetic resonance imaging scan parameters.

Region of interest	Field of view (mm ²)	Resolution (mm ²)	Slice thickness	Frame rate	No. slices (subject dependent)		
					Axial	Sagittal	Coronal
Tongue and surrounding tissues	240 × 240	1.88 × 1.88	6 mm	26 frames/s	10–14	5–9	10–14

Figure 2. (a) High-resolution magnetic resonance imaging (MRI) of sagittal tongue. (b, c) Tagged MRI in two directions. (d) Estimated three-dimensional tongue motion in a sagittal view.

interpolate between or beyond the estimated motion fields. Therefore, we manually specify the time indices t_o , t_s , t_u , and t_k of the four critical time frames /ə/, /s/, /u/, and /k/ from “a souk” for each subject by checking all MRI data. Specifically, the two consonants are defined as the first time frame in which the tongue (tip for /s/ and body for /k/) makes contact with the palate. The vowels are defined as the last frame before the tongue begins moving toward the consonant. For the schwa, the tip starts to extend, and the body starts to move forward and up. For /u/, the motion change is when the tongue starts moving directly up, instead of back and (maybe) slightly up. After all subjects are specified, we directly align these four time frames to a benchmark subject’s corresponding critical time indices T , T_s , T_u , and T_k by reassigning indices. For example, to reassign critical time frame /ə/ in the benchmark’s common space, we have

$$\mathbf{u}_{s,T_o}(\mathbf{X}) = \mathbf{u}_{s,t_o}(\mathbf{X}). \quad (2)$$

After reassigning these four instants, the remaining time indices between these critical frames are reassigned with the field closest to its linearly interpolated time index in the original subject’s timeline. For example, for any $T_o < T < T_s$, the corresponding time instant t in the original timeline is found by

$$\mathbf{u}_{s,T}(\mathbf{X}) = \mathbf{u}_{s,t}(\mathbf{X}), \text{ where } t = \text{round} \left(t_o + \frac{t_s - t_o}{T_s - T_o} (T - T_o) \right). \quad (3)$$

Next, we address the spatial alignment issue between subjects. In previous work, we have reported a method to construct an intensity tongue atlas using cine MRI data (Woo, Xing, Lee, Stone, & Prince, 2018). We regard this cine atlas from the same subject group as pre-existing data and use it as the basis of the common space that we relocate the 4D motions in. During the intensity atlas creation process, the deformation field to warp each subject to the atlas space is found by diffeomorphic image registration (Vercauteren et al., 2009) between each undeformed first time frame and the atlas space. We denote these time deformation fields as ψ_s ($1 \leq s \leq N$) between subject s and the atlas space. N is the number of subjects, which equals 14

in this study. In this model, an assumption was made that a global normalization method is capable of accounting most of the speaker variability in its operation. In this case, diffeomorphic image registration was used as the key method, which is used in many medical image analysis applications to account for anatomic variability efficiently. If we want to deform all subjects’ PVIRA motion estimates to the atlas space, according to Ehrhardt, Werner, Schmidt-Richberg, and Handels (2011), this can be achieved by composition of a sequence of motion fields, that is,

$$\phi'_{s,T}(\mathbf{X}) = \psi_s \circ \phi_{s,T} \circ \psi_s^{-1}(\mathbf{X}). \quad (4)$$

This equation can be understood in such a way: In the atlas space, a grid of points $\mathbf{X}$ deforms to a new location in time. This process is equivalent to the same grid deforming first to a subject space through a forward deformation field, then to a new location in time in this subject space, and finally back to the atlas space through a backward deformation field. This is as if the same grid $\mathbf{X}$ is retroprojected in the subject space. In practice, the composition is computed by interpolation of 3D vector fields, and ψ_s and ψ_s^{-1} are both available through cine atlas construction (details in Woo, Xing, et al., 2018). Note that $\phi'_{s,T}(\mathbf{X}) = \mathbf{X} + \mathbf{u}'_{s,T}(\mathbf{X})$. We have therefore acquired the warped motion fields $\mathbf{u}'_{s,T}(\mathbf{X})$ in the common atlas space. Examples of warped motion fields of a control and a patient at two time frames /s/ and /k/ are shown in Figures 3b, 3d, 3f, and 3h. Since all fields from different subjects are already temporally aligned, we average the fields among all subjects, and the mean is considered as a statistical 4D motion atlas (see Figures 3i and 3j).

$$\bar{\mathbf{u}}_T(\mathbf{X}) = \frac{1}{N} \sum_s \mathbf{u}'_{s,T}(\mathbf{X}). \quad (5)$$

Muscle Correlation Analysis

To locate internal muscle locations, the 3D anatomical vocal tract atlas was used to provide anatomical information for manual segmentation. Its resolution reaches $0.9 \times 0.9 \text{ mm}^2$, doubling that of tMRI. The same intensity atlas construction method was used on the hMRI data set to create

Figure 3. (a–d) Three-dimensional tongue motion of a control subject at time /s/ and /k/ in the original space and warped atlas space. (e–h) Three-dimensional tongue motion of a patient in the two spaces. (i, j) Four-dimensional motion atlas of /s/ and /k/ combining 16 control subjects. Note that color-coded cones are used to visualize motion, where anterior–posterior vectors are coded green, inferior–superior are coded blue, and left–right are coded red.

the vocal tract atlas (Woo et al., 2015) that contains clear internal muscle structures (see Figure 2a). A feasible manual segmentation of the internal muscles in the atlas space was taken from a previous study performed, evaluated, and discussed by speech scientists. The validity of the manual segmentation was analyzed and studied, with results summarized in the work of Stone et al. (2018). Especially, in their

methods section, detailed muscle identification criterion, related anatomy studies, and limitations on current segmentation are described with more properties of the segmentation result evaluated in the results section. In general, the delineation of all labels was carried out on each 2D slice and later combined into a 3D rendering to reveal muscle locations. For a muscle labeled by L , we denote its masked

region by $M_L(X)$. Its value is 1 for voxels inside L and 0 otherwise. Thus, each muscle's motion in the atlas space at time frame T is

$$\bar{u}_{L,T}(X) = M_L(X) \bar{u}_T(X). \quad (6)$$

Since the 4D atlas $\bar{u}_T(X)$ is already an average motion field, $\bar{u}_{L,T}(X)$ can be considered as an average “single muscle atlas” labeled by L .

In general, the 4D motion atlas is a tool that provides a common space for statistical investigation of muscle activities. Besides mean motion field, if we regard each subject's individual muscle motion $u'_{s,L,T}(X) = M_L(X)u'_{s,T}(X)$ as a sample of that muscle's general motion $U'_{L,T}(X)$ (as a random variable) in the atlas space, the correlation coefficient between any two muscle pairs L_1 and L_2 can be found by

$$\begin{aligned} c_{L_1L_2,T} &= \text{corr}(U'_{L_1,T}, U'_{L_2,T}) \\ &= \frac{E[(U'_{L_1,T} - \bar{u}_{L_1,T})(U'_{L_2,T} - \bar{u}_{L_2,T})]}{\sigma_{U'_{L_1,T}} \sigma_{U'_{L_2,T}}}. \end{aligned} \quad (7)$$

Here, σ is the standard deviation of the corresponding random variable. The range of $c_{L_1L_2,T}$ spans over $[-1, 1]$, where a high positive coordination between muscles L_1 and L_2 at time frame T yields a value close to 1 and a low coordination yields 0. After computing all correlation coefficients between all pairs in M number of muscles, the muscle correlation matrix at time frame T can be denoted as $C_T = (c_{L_iL_j,T}) \in \mathbb{R}^{M \times M}$, which reflects all muscles' coordination patterns over time.

In this particular application, the correlation matrices have specific properties. Although the range of $c_{L_1L_2,T}$ spans over $[-1, 1]$ in theory, it is expected to be positive due to the general trend of the displacement fields in the whole tongue. As illustrated in Figure 5f, at each muscle's location, its vector field in the atlas space as a random variable $U'_{L_1,T}$ has multiple sample vectors from all subjects (red dashed arrows) with a mean (red solid arrow). These sample vectors generally point to a similar direction because they are samples of the same muscle. Similarly, another muscle and its random variable $U'_{L_2,T}$ has multiple sample vectors as well (orange arrows). When correlation was computed between these two muscles using Equation 7, due to the sample vectors generally pointing to the same directions in the numerator part, the inner product operation yields a positive number. On the other hand, when the two sets of vectors generally point to opposite directions such as red $U'_{L_1,T}$ and blue $U'_{L_3,T}$, the inner product operation yields a negative number. However, Figure 3 shows the vector fields of the whole tongue following the same trend almost everywhere, no matter what muscles they are in. Since the tongue is an incompressible object with smooth displacement fields, a negative correlation with opposite displacement directions is almost nonexistent.

Results

Internal Tongue Muscle Labeling

The proposed workflow was implemented using MATLAB-based functions and in-house user interfaces for processing tMRI images (MathWorks). The muscle segmentation work was performed using the ITK-SNAP software (Yushkevich et al., 2006). Since the vocal tract atlas constructed using hMRI is a static image, it contains only one volume. The labeling result is shown in Figure 4.

Biomedically, the muscles of the tongue are classified as intrinsic and extrinsic, depending on their attachment to bones (Maton, 1997; Warwick, Williams, & Gray, 1973). The extrinsic muscles (genioglossus, hyoglossus, and styloglossus in Figure 4) are attached to the bone structure. The intrinsic muscles (superior longitudinal, inferior longitudinal, verticalis, and transverse muscles in Figure 4) are not attached to any bones. Both muscle types are responsible for the deformation and movement of the tongue. Moreover, although not often considered part of the tongue muscle group as their nominal function is to move the hyoid or jaw, the FOM muscles (mylohyoid, geniohyoid, and digastric muscles in Figure 4) are also important for positioning the whole tongue in the vocal tract and aiding in the tongue's position change, especially in its elevation. Therefore, we included all of them in the correlation study as well.

Control's Correlation Analysis

With the manually labeled muscle masks, we computed the correlation matrices using the 4D motion atlas from 16 healthy controls. Since correlation matrices can be computed using an arbitrary number of time frames (depending on investigation focus), we included one previous time frame and one later time frame around each critical frame and formed four brief time intervals of interest: around /ə/, /s/, /u/, and /k/, respectively. The correlation matrices are plotted in Figures 5a–5d. The color scheme ranges from -1 as dark blue to $+1$ as dark red.

One immediate observation is that the muscle motions are all positively correlated (> 0), as expected in the Method section. This is a direct result from using displacement fields to represent each muscle's location. Although each muscle contracts or expands in their unique way, their combined deformation drives the tongue to a single general orientation, yielding smooth vector fields flowing at positively correlated directions. Over the 26 time frames, the muscles get more positively correlated (from light red to dark red), indicating an increasing amount of coherence between motion vectors when pronouncing “a souk.” If we consider a correlation value of .7 and above to be a worthwhile correlation that is meaningful, Figure 5a shows three blocks of highly correlated muscles. These are (a) the FOM (lower right corner squares of MH, GH, D), (b) the SG and HG, and (c) the GG, T, V, and SL. Figure 5a also shows low correlation between the FOM and the other tongue muscles (yellow-orange), with the exception of IL. This correlation changes, however, during the /u/ and

Figure 4. Manual labeling of the internal tongue muscles using high-resolution magnetic resonance images. Muscles overlay each other so that they are shown in two separate figures.

Muscle labels

1. Genioglossus 2. Transverse 3. Verticalis 4. Superior Longitudinal 5. Inferior Longitudinal
6. Styloglossus 7. Hyoglossus 8. Mylohyoid 9. Geniohyoid 10. Digastric

Figure 5. Correlation pattern of 10 tongue muscles from 16 healthy controls in different time periods pronouncing “a souk.” (a) Around time frame /ə/. (b) Around time frame /s/. (c) Around time frame /u/. (d) Around time frame /k/. (e) General correlation pattern of all time frames. (f) Explanation of directions and correlation sign.

especially the /k/ gesture, when all the muscles of the tongue and FOM highly correlate. Also, if we consider all time frames together pronouncing “a souk” as one task, the general correlation pattern is plotted in Figure 5e. The FOM muscles also show less cooperation with top muscles.

Patient’s Correlation Analysis

With the three post-partial glossectomy patients, we analyzed their results separately, because each patient had his or her unique surgical treatment and motion pattern. It is not reasonable to treat each patient’s unique motion as one sample from a “general patient motion pattern.” Therefore, we computed each patient’s muscle correlation pattern independently. The result of each patient is shown in Figures 6–8. Compared to the controls (see Figure 5), Patients 1 and 2 have much higher correlations between the FOM muscles and the tongue dorsal muscles (darker red). Patients show more tendency to exercise en bloc movements compared to the controls.

Discussion

The data set used in this article used only patients with a unilateral tumor occurring posterior to the tongue

tip. Since the resections were all small and similar in size ($2.4 \times 2.1 \times 1.8$, $2.8 \times 2.4 \times 1.8$, $2.2 \times 2.1 \times 1.4 \text{ cm}^3$), the effects of tumor size on motion pattern are not considered. However, reduced control in the tip on the resected side may contribute to motion differences between patients and controls. Tongue-tip fricatives such as /s/ are challenging for these patients (Heller, Levy, & Sciubba, 1991), so the speech task includes the sound /s/.

Results showed an interesting organizational strategy for the controls and differences with the patients. To begin, the controls had three blocks of muscles that worked as coordinated units: the FOM; SG and HG; and GG, T, V, and SL. Figure 5 shows the FOM muscles to be a group unto themselves. This is not surprising, as they are not considered to be internal tongue muscles, and their function is linked with swallowing more than speech; they pull the hyoid forward during swallowing and lower the jaw during speaking. However, when the FOM muscles are shortened, they thicken, which elevates the tongue. Thus, they are available for use as tongue elevators and especially may be used by patients with weakened or damaged tongues to augment tongue elevation gestures.

Figure 5 also shows the SG and HG to be a correlated unit throughout the word. When these two muscles activate together, they pull the tongue straight back. The word *souk*

Figure 6. Correlation pattern of 10 tongue muscles from post-partial glossectomy Patient 1 pronouncing “a souk.” (a–d) Four critical time frames. (e) All time frames.

Figure 7. Correlation pattern of 10 tongue muscles from post-partial glossectomy Patient 2 pronouncing “a souk.” (a–d) Four critical time frames. (e) All time frames.

was chosen because its primary direction of motion is front to back. The correlation of these two muscles shows that they are controlling the anterior-to-posterior position of the tongue together. The third block of muscles, GG, V, T, and SL, contains the four largest tongue muscles and the ones controlling the four directions of deformation (Stone et al., 2018). T controls tongue width, V controls tongue height, SL controls tongue length, and GG controls the radial shortening of the entire tongue. These muscles are highly innervated and likely to be activated in multiple locations for small local motions (Parthasarathy et al., 2007; Sokoloff, 2000). The coordination of these muscles indicates that they are controlling the overall deformation of the tongue during these sounds. The remaining, smaller muscles most likely are fine-tuning tongue shape across sounds and subjects and thus show slightly less correlation.

The IL muscle is a curious case. Its role, based on location and fiber direction, is to shorten/elevate the tongue and depress the tip. Anatomically, the intrinsic tongue muscles are completely interdigitated, the FOM muscles are completely bundled, and the extrinsic muscles are bundled at their origin but interdigitate when they enter the tongue body (Stone et al., 2018). The IL is an exception, as it is an intrinsic muscle that is separated from the other

muscles in the anterior tongue by a triangular boundary of septa (Abd-El-Malek, 1939). This separation creates a bundled muscle within the tongue similar to the FOM muscles below. For the controls and Patient 3, the IL correlates with the FOM at least as often as the tongue muscles (see Figures 5 and 8). This separation reduces friction and co-contraction with other muscles when it is active and may facilitate its role in elevating the tongue.

Examining the patients, it can be seen that the consonants cause a high muscle correlation. Patient 2 reaches complete correlation during /s/, Patient 1 reaches complete correlation during /u/, and Patient 3 retains independence between the tongue and FOM muscles but, otherwise, has complete correlations by /s/ as well. This is consistent with a reduction in degrees of freedom. All the muscles are engaged. En bloc movements seem to be the way that patients compensate for their loss of elegance in speech sound production.

Since sensation loss of the tongue often occurs after glossectomy, there could be an impact on muscle correlation. Before our study, the patients’ oral sensation was tested in two ways. First, a von Frey filament was applied to multiple locations on the tongue surface and lateral margins to test tactile awareness. Second, two-point discrimination was tested using curve-tip fine forceps with

Figure 8. Correlation pattern of 10 tongue muscles from post-partial glossectomy Patient 3 pronouncing “a souk.” (a–d) Four critical time frames. (e) All time frames.

rigid distances of 0, 3, and 6 mm. For Patient 1, there was no awareness of touch or two-point discrimination in the tongue’s tumor side’s body or back. However, both tests showed good sensation throughout the entire tongue tip and the tongue’s native side’s body and back. For Patient 2, Von Frey showed no awareness of touch surrounding the tumor side, but normal sensation in the tip. Two-point discrimination was good everywhere on the native side but was reduced on the tumor side throughout the tongue’s length. For Patient 3, Von Frey showed no sensation immediately anterior to the tumor region but showed good sensation throughout the entire tongue tip and on the entire native side. Since all three patients showed similar conditions of sensation loss, its impact on muscle correlation is difficult to determine. Future study could focus on varying patient’s sensation loss type for more in-depth analyses.

A limitation of this article is the use of velocity fields as the basis for muscle coordination analysis, causing smoothness of the motion fields and an all positive correlation that limits the range of comparison. If internal fiber directions could be learned through additional imaging techniques, strain may be calculated along these muscle fiber directions that reflects the activation pattern (contraction or expansion) of the internal muscles. Despite the fact that, in a volume-preserving structure, such as the tongue,

some muscles must be shortening orthogonal to those that are lengthening, a strain field would reflect this shortening and lengthening with positive and negative strain across the muscle antagonists. Also, correlation matrices computed from strain along the muscle fiber directions may serve as a more insightful indicator for the muscle coordination patterns, where negative correlation results are expected and could be more informative. The 3D velocity fields, however, display higher dimensional data. The arrows reflect a point’s 3D motion, not the 2D components of that motion. Thus, motion that is both forward and inward is represented by a single oblique purple arrow. The speech task, “a souk,” was chosen for its use of fairly simple motions, which also happen to be primarily in the anterior-posterior and superior-inferior direction: forward into /s/ and backward/upward into /uk/. The simultaneous out-of-plane component (medial/lateral) is minimal and is subsumed by the main motion in the velocity field. As a result, more methods to enhance this analysis will be investigated in the future.

Another limitation is that there was little previous work on the same topic, and conclusions drawn from these results are hardly supported by further evidence. We note that current findings were only derived from this data set, and these data cover limited ground. As further research

develops, more insights will be gained on this topic, and further comparison can be made.

Conclusion

In this article, we presented a method to analyze multi-subject tongue muscle correlation using motion patterns in speech sound production. Correlation between each two muscle pairs is computed within each labeled region. The analysis is performed on a population of 16 healthy subjects and three post-partial glossectomy patients. Correlation matrices in the atlas space show the coordination of tongue muscles during speech. The FOM muscles are less coordinated from the internal tongue muscles. Patients tend to use more FOM muscles to compensate for their postsurgery function loss.

Acknowledgments

This project was supported by National Institute on Deafness and Other Communication Disorders Grants R21DC016047 (awarded to PI: Woo), R00DC012575 (awarded to PI: Woo), and R01DC014717 (awarded to PI: Prince) and National Cancer Institute Grant R01CA133015 (awarded to PI: Stone).

References

- Abd-El-Malek, S. (1939). Observations on the morphology of the human tongue. *Journal of Anatomy*, 73(Pt. 2), 201–210.
- Abd-El-Malek, S. (1955). The part played by the tongue in mastication and deglutition. *Journal of Anatomy*, 89(Pt. 2), 250–254.
- Bressmann, T., Jacobs, H., Quintero, J., & Irish, J. C. (2009). Speech outcomes for partial glossectomy surgery: Measures of speech articulation and listener perception. *Canadian Journal of Speech-Language Pathology and Audiology*, 33(4), 204.
- Bressmann, T., Sader, R., Whitehill, T. L., & Samman, N. (2004). Consonant intelligibility and tongue motility in patients with partial glossectomy. *Journal of Oral and Maxillofacial Surgery*, 62(3), 298–303.
- Ehrhardt, J., Werner, R., Schmidt-Richberg, A., & Handels, H. (2011). Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. *IEEE Transactions on Medical Imaging*, 30(2), 251–265.
- Heller, K. S., Levy, J., & Sciubba, J. J. (1991). Speech patterns following partial glossectomy for small tumors of the tongue. *Head & Neck*, 13(4), 340–343.
- Kent, R. D., & Read, C. (2002). *Acoustic analysis of speech*. San Diego, CA: Singular.
- Kier, W. M., & Smith, K. K. (1985). Tongues, tentacles and trunks: The biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnean Society*, 83(4), 307–324.
- Liu, X., Abd-Elmoniem, K. Z., Stone, M., Murano, E. Z., Zhuo, J., Gullapalli, R. P., & Prince, J. L. (2012). Incompressible deformation estimation algorithm (IDEA) from tagged MR images. *IEEE Transactions on Medical Imaging*, 31(2), 326–340.
- Mansi, T., Pennec, X., Sermesant, M., Delingette, H., & Ayache, N. (2011). iLogDemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues. *International Journal of Computer Vision*, 92(1), 92–111.
- Maton, A. (1997). *Human biology and health*. Englewood Cliffs, NJ: Prentice Hall.
- NessAiver, M., & Prince, J. L. (2003). Magnitude image CSPAMM reconstruction (MICSr). *Magnetic Resonance in Medicine*, 50(2), 331–342.
- Nicoletti, G., Soutar, D. S., Jackson, M. S., Wrench, A. A., Robertson, G., & Robertson, C. (2004). Objective assessment of speech after surgical treatment for oral cancer: Experience from 196 selected cases. *Plastic and Reconstructive Surgery*, 113(1), 114–125.
- Osman, N. F., McVeigh, E. R., & Prince, J. L. (2000). Imaging heart motion using harmonic phase MRI. *IEEE Transactions on Medical Imaging*, 19(3), 186–202.
- Parthasarathy, V., Prince, J. L., Stone, M., Murano, E. Z., & NessAiver, M. (2007). Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. *The Journal of the Acoustical Society of America*, 121(1), 491–504.
- Pauloski, B. R., Logemann, J. A., Colangelo, L. A., Rademaker, A. W., McConnel, F., Heiser, M. A., . . . Esclamado, R. (1998). Surgical variables affecting speech in treated patients with oral and oropharyngeal cancer. *The Laryngoscope*, 108(6), 908–916.
- Perry, B. J., Martino, R., Yunusova, Y., Plowman, E. K., & Green, J. R. (2018). Lingual and jaw kinematic abnormalities precede speech and swallowing impairments in ALS. *Dysphagia*, 1–8.
- Rastadmehr, O., Bressmann, T., Smyth, R., & Irish, J. C. (2008). Increased midsagittal tongue velocity as indication of articulatory compensation in patients with lateral partial glossectomies. *Head & Neck*, 30(6), 718–726.
- Sedov, L. I. (1997). *Mechanics of continuous media (in 2 volumes)*. Singapore: World Scientific.
- Sokoloff, A. J. (2000). Localization and contractile properties of intrinsic longitudinal motor units of the rat tongue. *Journal of Neurophysiology*, 84(2), 827–835.
- Stone, M., Davis, E. P., Douglas, A. S., NessAiver, M., Gullapalli, R., Levine, W. S., & Lundberg, A. (2001). Modeling the motion of the internal tongue from tagged cine-MRI images. *The Journal of the Acoustical Society of America*, 109(6), 2974–2982.
- Stone, M., Langguth, J. M., Woo, J., Chen, H., & Prince, J. L. (2014). Tongue motion patterns in post-glossectomy and typical speakers: A principal components analysis. *Journal of Speech, Language, and Hearing Research*, 57(3), 707–717.
- Stone, M., Liu, X., Chen, H., & Prince, J. L. (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.
- Stone, M., Woo, J., Lee, J., Poole, T., Seagraves, A., Chung, M., . . . Blemker, S. S. (2018). Structure and variability in human tongue muscle anatomy. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(5), 499–507.
- Takemoto, H. (2001). Morphological analyses of the human tongue musculature for three-dimensional modeling. *Journal of Speech, Language, and Hearing Research*, 44(1), 95–107.
- Vercateren, T., Pennec, X., Perchant, A., & Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1), S61–S72.
- Warwick, R., Williams, P. L., & Gray, H. (1973). *Gray's anatomy*. London, United Kingdom: Longman.
- Woo, J., Lee, J., Murano, E. Z., Xing, F., Al-Talib, M., Stone, M., & Prince, J. L. (2015). A high-resolution atlas and statistical model of the vocal tract from structural MRI. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 3(1), 47–60.
- Woo, J., Prince, J. L., Stone, M., Xing, F., Gomez, A., Green, J. R., . . . Fakhri, G. E. (2018). A sparse non-negative matrix factorization framework for identifying functional units of tongue

- behavior from MRI. Retrieved from <https://arxiv.org/abs/1804.05370>
- Woo, J., Stone, M., Xing, F., Green, J., Gomez, A., Wedeen, V., ... El Fakhri, G. (2017). Discovering functional units of the human tongue during speech from cine- and tagged-MRI. *The Journal of the Acoustical Society of America*, 141(5), 3647.
- Woo, J., Xing, F., Lee, J., Stone, M., & Prince, J. L. (2018). A spatio-temporal atlas and statistical model of the tongue during speech from cine-MRI. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(5), 520–531.
- Woo, J., Xing, F., Stone, M., Green, J., Reese, T. G., Brady, T. J., ... El Fakhri, G. (2017). Speech map: A statistical multimodal atlas of 4D tongue motion during speech from tagged and cine MR images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 1–13. <https://doi.org/10.1080/21681163.2017.1382393>
- Xing, F., Prince, J. L., Stone, M., Reese, T. G., Atassi, N., Wedeen, V. J., ... Woo, J. (2018). Strain map of the tongue in normal and ALS speech patterns from tagged and diffusion MRI. *Proceedings of SPIE—The International Society for Optical Engineering*, 10574, pii: 1057411.
- Xing, F., Woo, J., Gomez, A. D., Pham, D. L., Bayly, P. V., Stone, M., & Prince, J. L. (2017). Phase vector incompressible registration algorithm for motion estimation from tagged magnetic resonance images. *IEEE Transactions on Medical Imaging*, 36(10), 2116–2128.
- Xing, F., Woo, J., Lee, J., Murano, E. Z., Stone, M., & Prince, J. L. (2016). Analysis of 3-D tongue motion from tagged and cine magnetic resonance images. *Journal of Speech, Language, and Hearing Research*, 59(3), 468–479.
- Xing, F., Ye, C., Woo, J., Stone, M., & Prince, J. (2015). Relating speech production to tongue muscle compressions using tagged and high-resolution magnetic resonance imaging. *Proceedings of SPIE—The International Society for Optical Engineering* (Vol. 9413), pii: 94131L.
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31(3), 1116–1128.
- Zerhouni, E. A., Parish, D. M., Rogers, W. J., Yang, A., & Shapiro, E. P. (1988). Human heart: Tagging with MR imaging—A method for noninvasive assessment of myocardial motion. *Radiology*, 169(1), 59–63.