JSLHR JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH

Atlas-based Tongue Muscle Correlation Analysis from Tagged and High-resolution MRI

Journal:	*Journal of Speech, Language, and Hearing Research*
Manuscript ID	JSLHR-S-18-0495.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	25-Mar-2019
Complete List of Authors:	Xing, Fangxu; Massachusetts General Hospital, Radiology; Harvard Medical School, Radiology Stone, Maureen; University of Maryland Dental School, Neural and Pain Sciences Goldsmith, Tessa; Massachusetts General Hospital, Speech, Language and Swallowing Disorders Prince, Jerry; Johns Hopkins University, Dept. Electrical and Computer Engineering El Fakhri, Georges; Massachusetts General Hospital, Radiology; Harvard Medical School, Radiology Woo, Jonghye; Massachusetts General Hospital, Radiology; Harvard Medical School, Radiology
Keywords:	Speech motor control, Speech-language-pathology, Neuroimaging, Speech

Atlas-based Tongue Muscle Correlation Analysis from Tagged and High-resolution MRI

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

[a] Gordon Center for Medical Imaging, Dept. of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA

[b] Dept. of Neural and Pain Sciences, University of Maryland Dental School, Baltimore, MD

[c] Dept. of Speech, Language and Swallowing, Massachusetts General Hospital, Boston, MA

[d] Dept. of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD

This research was supported by NIH/NIDCD R21DC016047 (PI: Woo), R00DC012575 (PI: Woo), R01DC014717 (PI: Prince), and NIH/NCI R01CA133015 (PI: Stone).

Correspondence concerning this article should be addressed to Fangxu Xing, Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114. E-mail: fxing1@mgh.harvard.edu. Tel: (617) 726-6904.

Abstract

Purpose: Intrinsic and extrinsic tongue muscles in healthy and diseased populations vary both in their intra- and inter-subject behaviors during speech. Identifying coordination patterns among various tongue muscles can provide insights into speech motor control and help developing new therapeutic and rehabilitative strategies. *Method:* We present a method to analyze multi-subject tongue muscle correlation using motion patterns in speech sound production. Motion of muscles is captured using tagged magnetic resonance imaging (MRI) and computed using a phased-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 MRI 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 sixteen control subjects and three 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 post-surgery function loss.

Keywords: tongue, muscle, correlation, MRI, speech-language pathology, atlas, motion, glossectomy

Introduction

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, 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 post-glossectomy patients (Bressmann, Sader, Whitehill, & Samman, 2004; Bressmann, Jacobs, Quintero, & Irish, 2009; Nicoletti et al., 2004; Pauloski et al., 1998; Rastadmehr, Bressmann, Smyth, & Irish, 2008) and amyotrophic lateral sclerosis (ALS) patients (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 is 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). 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 normal 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—i.e., *functional units*—to reveal parts of the tissue that are most likely to work in coordination (Woo et al., 2017; Woo 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, Ye, Woo, Stone, & Prince, 2015; Xing et al., 2018).

However, there exists a common limitation in previous reported analyses: the difficulty of statistically achieving multi-subject 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 post-surgery 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 et al., 2017), and the strain analysis studied each individual subject's motion field instead of the combined multi-subject 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 multi-subject motion fields can be found.

Previously, a 3D structural atlas of the vocal tract from a collection of high-resolution magnetic resonance images (hMRI) 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 normal 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 hMRI muscle masks. Finally, correlation matrices of muscle motion are computed between all labeled regions. We compared the result of sixteen normal controls and three post-partial glossectomy patients. Besides revealing the unique motion patterns of the floor-of-mouth 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 multi-subject 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.

Data Acquisition

A dataset of sixteen normal controls and three post-partial glossectomy patients was used in the analysis. All subjects read and signed a consent form and the HIPAA 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 the wound 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 2cm in the largest dimension, though resection includes a 1-1.5cm 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 size. 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” (IPA: /ə’suk/). 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 MR imaging 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 MR 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, Malvern, PA) with a 12-channel head coil and a 4-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 2(b) and 2(c).

Table 1

Tongue Motion MRI Scan Parameters

Region of Interest	Field of View (mm^2)	Resolution (mm^2)	Slice Thickness	Frame Rate	# Slices (Subject dependent)		
					Axial	Sagittal	Coronal
Tongue & surrounding tissues	240×240	1.88×1.88	6 mm	26 frm/sec	10-14	5-9	10-14

4D Motion Estimation

To estimate 4D motion from tMRI data, we used PVIRA—the phase vector incompressible registration algorithm (Xing et al., 2017). PVIRA is a phased-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 (Figure 2(d)).

For any subject labeled by s, we denote the PVIRA estimate as $\boldsymbol{u}_{s,t}(\boldsymbol{X})$ at time frame t, $t = 1,2,..., 26$. At the undeformed time frame, $\boldsymbol{X}$ 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 3(a), 3(c), 3(e), and 3(g). We use $\phi_{s,t}$ to denote the corresponding deformation between time frame t and the first (undeformed) time frame. Mathematically, we have

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

This equation shows the motion of tissue point coordinates $\boldsymbol{X}$. At the time frame t, they move to a new location through motion field $\boldsymbol{u}_{s,t}(\boldsymbol{X})$. This material frame definition is called a *Lagrangian* framework (Sedov, 1997). It roots every motion field $\boldsymbol{u}_{s,t}(\boldsymbol{X})$ in the undeformed frame $\boldsymbol{X}$, 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 Eulerian, because any quantity computed over time is always mapped and displayed on the undeformed frame so that the tongue appears *motionless* (Woo 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 datasets 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 interpolate between or beyond the estimated motion fields. Therefore, we manually specify the time indices $t_{\text{ə}}$, 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 towards 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 $T_{\text{ə}}$, 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

$$\boldsymbol{u}_{S,\,T_{\text{ə}}}(\boldsymbol{X}) = \boldsymbol{u}_{S,\,t_{\text{ə}}}(\boldsymbol{X})\ . \tag{2}$$

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 $T_ə < T < T_s$, the corresponding time instant t in the original timeline is found by

$$\boldsymbol{u}_{s,T}(\boldsymbol{X}) = \boldsymbol{u}_{s,t}(\boldsymbol{X}) \text{ , where } t = \text{round}\left(t_ə + \frac{t_s - t_ə}{T_s - T_ə}(T - T_ə)\right) . \tag{3}$$

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 $(1 \le s \le N)$ 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, i.e.,

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

This equation can be understood in such a way: In the atlas space, a grid of points $\boldsymbol{X}$ 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 $\boldsymbol{X}$ is retro-projected 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 et al. (2018)). Note that $\phi'_{s,T}(\boldsymbol{X}) = \boldsymbol{X} + \boldsymbol{u}'_{s,T}(\boldsymbol{X})$. We have therefore acquired the warped motion fields $\boldsymbol{u}'_{s,T}(\boldsymbol{X})$ 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 3(b), 3(d), 3(f), and 3(h). 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 (Figures 3(i) and 3(j)).

$$\overline{\boldsymbol{u}}_T(\boldsymbol{X}) = \frac{1}{N}\sum_s \boldsymbol{u}'_{s,T}(\boldsymbol{X}) \ . \tag{5}$$

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 × 0.9 mm², doubling that of tMRI. The same intensity atlas construction method was used on the hMRI dataset to create the vocal tract atlas (Woo et al., 2015) that contains clear internal muscle structures (Figure 2(a)). 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 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 $M_L(\boldsymbol{X})$. 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

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

Since the 4D atlas $\bar{\boldsymbol{u}}_T(\boldsymbol{X})$ is already and average motion field, $\bar{\boldsymbol{u}}_{L,T}(\boldsymbol{X})$ 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 $\boldsymbol{u}'_{s,L,T}(\boldsymbol{X}) = M_L(\boldsymbol{X})\boldsymbol{u}'_{s,T}(\boldsymbol{X})$ as a sample of that muscle's general motion $\boldsymbol{U}'_{L,T}(\boldsymbol{X})$ (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

$$c_{L_1L_2,T} = \mathrm{corr}\left(\boldsymbol{U}'_{L_1,T}, \boldsymbol{U}'_{L_2,T}\right) = \frac{\mathrm{E}\left[\left(\boldsymbol{U}'_{L_1,T} - \bar{\boldsymbol{u}}_{L_1,T}\right)\left(\boldsymbol{U}'_{L_2,T} - \bar{\boldsymbol{u}}_{L_2,T}\right)\right]}{\sigma_{\boldsymbol{U}'_{L_1,T}}\sigma_{\boldsymbol{U}'_{L_2,T}}}. \quad (7)$$

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 muscle's 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 5(f), at each muscle's location, its vector field in the atlas space as a random variable $\boldsymbol{U}'_{L_1,T}$ 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 $\boldsymbol{U}'_{L_2,T}$ has multiple sample vectors as well (orange arrows). When correlation was computed between these two muscles using Eq. (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 $\boldsymbol{U}'_{L_1,T}$ and blue $\boldsymbol{U}'_{L_3,T}$, 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 non-existent.

Results

Internal Tongue Muscle Labeling

The proposed workflow was implemented using Matlab-based functions and in-house user interfaces for processing tMRI images (MathWorks, Natick, MA). 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 floor-of-mouth (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 normal 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 5(a) to 5(d). 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 0.7 and above to be a worthwhile correlation that is meaningful, Figure 5(a) shows three blocks of highly correlated muscles. These are 1) the FOM (lower right corner squares of MH, GH, D), 2) the SG-and-HG, and 3) the GG, T, V, and SL. Figure 5(a) 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 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 5(e). 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/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 (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 dataset used in this paper used only patients with a unilateral tumor occurring posterior to the tongue tip. Since the resections were all small and similar in size (2.4×2.1×1.8, 2.8×2.4×1.8, 2.2×2.1×1.4 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 as well as 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 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 (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 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 rigid distances of 0mm, 3mm, and 6mm. For Patient 1, there was no awareness of touch or two-point discrimination in the tongue's tumor side's body or back. But 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 paper 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 like 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. And 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 AP and SI direction: forward into /s/ and backward/upwards 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 dataset and this data covers limited ground. As further research develops, more insights will be gain on this topic and further comparison can be made.

Conclusion

In this paper, 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 sixteen normal subjects and three post-partial glossectomy patients. Correlation matrices in the atlas space show the coordination of tongue muscles during speech. The floor-of-mouth muscles are less coordinated from the internal tongue muscles. Patients tend to use more floor-of-mouth muscles to compensate for their post-surgery function loss.

Acknowledgements

This project was supported by NIH/NIDCD R21DC016047 (PI: Woo), R00DC012575 (PI: Woo), R01DC014717 (PI: Prince), and NIH/NCI R01CA133015 (PI: Stone).

Conflict of Interest

The authors declare no conflict of interest of this article.

References

Abd-El-Malek, S. (1939). Observations on the morphology of the human tongue. *Journal of anatomy*, *73*(Pt 2), 201.

Abd-El-Malek, S. (1955). The part played by the tongue in mastication and deglutition. *Journal of anatomy*, 89(Pt 2), 250.

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.

Bressmann, T., Jacobs, H., Quintero, J., & Irish, J. C. (2009). Speech Outcomes for Partial Glossectomy Surgery: Measures of speech articulation and listener perception. Indicateurs de la parole pour une glossectomie partielle: Mesures de l'articulation de la parole et de la perception des auditeurs. *Special Issue: Head and Neck Cancer*, *33*(4), 204.

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. (2002). *Acoustic analysis of speech*. Singular Publishing Group.

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. *Medical Imaging, IEEE Transactions on*, 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*. 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. *Medical Imaging, IEEE Transactions on*, 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)*.

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., 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., 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., 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.

Vercauteren, 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*. 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., 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), 3647-3647.

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.

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. *arXiv preprint arXiv:1804.05370*.

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.

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. In *Medical Imaging 2015: Image Processing* (Vol. 9413, p. 94131L). International Society for Optics and Photonics.

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., 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., 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. In *Medical Imaging 2018: Image Processing* (Vol. 10574, p. 1057411). International Society for Optics and Photonics.

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.

Atlas-based Tongue Muscle Correlation Analysis from Tagged and High-resolution MRI

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

[a] Gordon Center for Medical Imaging, Dept. of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA

[b] Dept. of Neural and Pain Sciences, University of Maryland Dental School, Baltimore, MD

[c] Dept. of Speech, Language and Swallowing, Massachusetts General Hospital, Boston, MA

[d] Dept. of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD

This research was supported by NIH/NIDCD R21DC016047 (PI: Woo), R00DC012575 (PI: Woo), R01DC014717 (PI: Prince), and NIH/NCI R01CA133015 (PI: Stone).

Correspondence concerning this article should be addressed to Fangxu Xing, Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114. E-mail: fxing1@mgh.harvard.edu. Tel: (617) 726-6904.

Abstract

Purpose: Intrinsic and extrinsic tongue muscles in healthy and diseased populations vary both in their intra- and inter-subject behaviors during speech. Identifying coordination patterns among various tongue muscles can provide insights into speech motor control and help developing new therapeutic and rehabilitative strategies. *Method:* We present a method to analyze multi-subject tongue muscle correlation using motion patterns in speech sound production. Motion of muscles is captured using tagged magnetic resonance imaging (MRI) and computed using a phased-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 MRI 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 sixteen control subjects and three 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 post-surgery function loss.

Keywords: tongue, muscle, correlation, MRI, speech-language pathology, atlas, motion, glossectomy

Introduction

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, 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 post-glossectomy patients (Bressmann, Sader, Whitehill, & Samman, 2004; Bressmann, Jacobs, Quintero, & Irish, 2009; Nicoletti et al., 2004; Pauloski et al., 1998; Rastadmehr, Bressmann, Smyth, & Irish, 2008) and amyotrophic lateral sclerosis (ALS) patients (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 is 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.

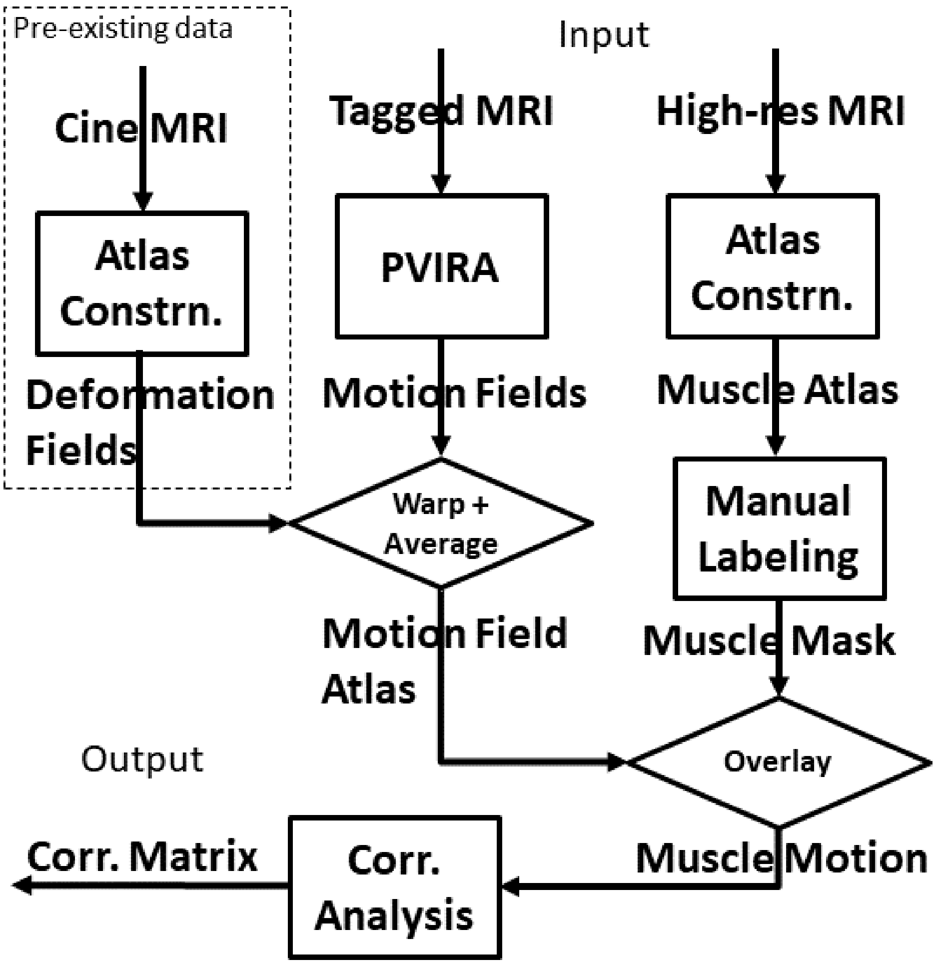
In the past decades, tagged magnetic resonance imaging (tMRI) has been widely used in the structural and functional studies of the human tongue (Parthasarathy, Prince, Stone, Murano, & NessAiver, 2007; Stone et al., 2001). The key technique of tMRI is to overlay a gridded or striped tag pattern in the 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 normal 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—i.e., *functional units*—to reveal parts of the tissue that are most likely to work in coordination (Woo et al., 2017; Woo 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, Ye, Woo, Stone, & Prince, 2015; Xing et al., 2018).

However, there exists a common limitation in previous reported analyses: the difficulty of statistically achieving multi-subject 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 post-surgery 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 et al., 2017), and the strain analysis studied each individual subject's motion field instead of the combined multi-subject 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 multi-subject motion fields can be found.

Previously, a 3D structural atlas of the vocal tract from a collection of high-resolution magnetic resonance images (hMRI) 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 normal 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 hMRI muscle masks. Finally, correlation matrices of muscle motion are computed between all labeled regions. We compared the result of sixteen normal controls and three post-partial glossectomy patients. Besides revealing the unique motion patterns of the floor-of-mouth 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

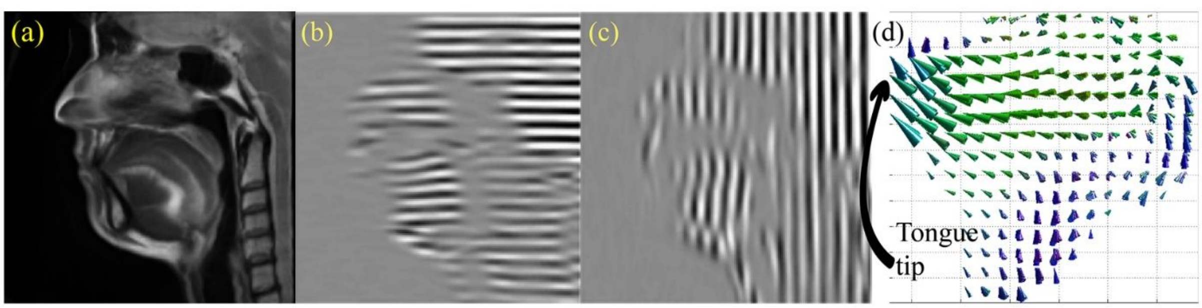
The overall process of the proposed analysis is illustrated in Figure 1. The input is multi-subject tMRI and hMRI data. The output is a collection of correlation matrices between muscles aligned in the atlas space. The details for each specific method are described below.



~~Figure 1. Workflow of the proposed method.~~

Data Acquisition

A dataset of sixteen normal controls and three post-partial glossectomy patients was used in the analysis. All subjects read and signed a consent form and the HIPAA 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 the wound 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 2cm in the largest dimension, though resection includes a 1-1.5cm



~~Figure 2. (a) High-resolution MRI of sagittal tongue. (b)(c) Tagged MRI in two directions. (d) Estimated 3D tongue motion in a sagittal view.~~

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 size. 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” (IPA: /ə’suk/). 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 MR imaging 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 ~~with~~ a second ~~for~~ of data collection. Pauses were made between different slice orientations. The MR 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, Malvern, PA) with a 12-channel head coil and a 4-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 2(b) and 2(c).

Table 1

Tongue Motion MRI Scan Parameters

Region of Interest	Field of View (mm^2)	Resolution (mm^2)	Slice Thickness	Frame Rate	# Slices (Subject dependent)		
					Axial	Sagittal	Coronal

Tongue & surrounding tissues	240×240	1.88×1.88	6 mm	26 frm/sec	10-14	5-9	10-14

4D Motion Estimation

To estimate 4D motion from tMRI data, we used PVIRA—the phase vector incompressible registration algorithm (Xing et al., 2017). PVIRA is a phased-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 (Figure 2(d)).

For any subject labeled by s, we denote the PVIRA estimate as $\boldsymbol{u}_{s,t}(\boldsymbol{X})$ at time frame t, $t = 1,2,\ldots, 26$. ~~.~~At the undeformed time frame, $\boldsymbol{X}$ 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 3(a), 3(c), 3(e), and 3(g). We use $\phi_{s,t}$ to denote the corresponding deformation between time frame t and the first (undeformed) time frame. Mathematically, we have

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

This equation shows the motion of tissue point coordinates X. ~~. At the undeformed time frame, $\boldsymbol{X}$ is the tissue point coordinates in the voxel grid.~~ At the time frame t, they move to a new location through motion field $\boldsymbol{u}_{s,t}(\boldsymbol{X})$. This material frame definition is called a *Lagrangian* framework (Sedov, 1997). It roots every motion field $\boldsymbol{u}_{s,t}(\boldsymbol{X})$ in the undeformed frame $\boldsymbol{X}$, 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 Eulerian, because any quantity computed over time is always mapped and displayed on the undeformed frame so that the tongue appears *motionless* (Woo 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 datasets 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 interpolate between or beyond the estimated motion fields. Therefore, we manually specify the time indices $t_ə$, 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 towards 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. ~~Each of these time instants marks a maximum tongue position for any individual subject.~~ After all subjects are specified, w~~W~~e directly align these four time frames to a benchmark subject's corresponding critical time indices $T_ə$, 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

$$\boldsymbol{u}_{s,T_ə}(\boldsymbol{X}) = \boldsymbol{u}_{s,t_ə}(\boldsymbol{X}) \,. \qquad (2)$$

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 $T_ə < T < T_s$, the corresponding time instant t in the original timeline is found by

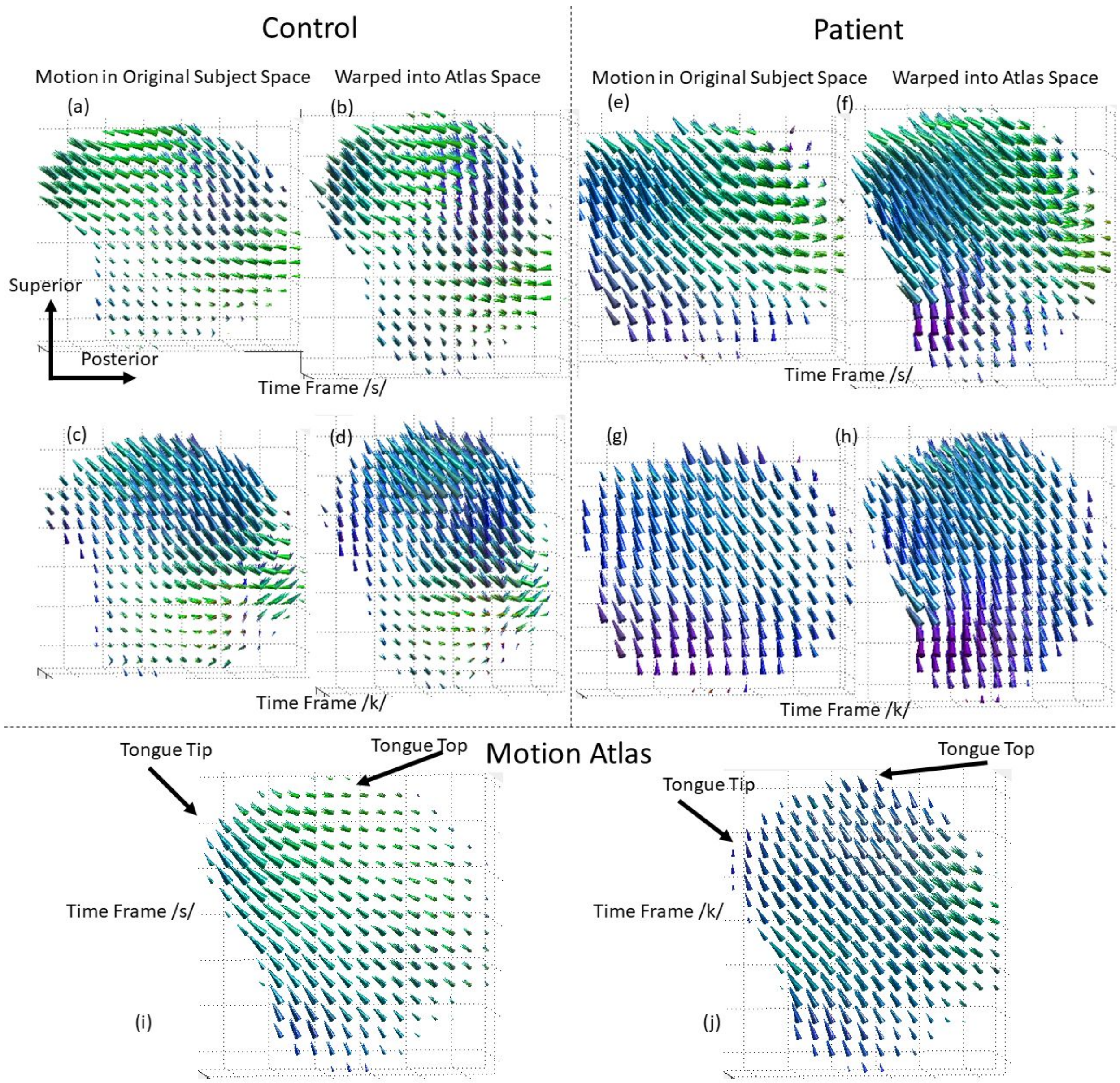
$$\boldsymbol{u}_{s,T}(\boldsymbol{X}) = \boldsymbol{u}_{s,t}(\boldsymbol{X}) \text{ , where } t = \mathrm{round}\left(t_ə + \frac{t_s - t_ə}{T_s - T_ə}(T - T_ə)\right) . \quad (3)$$

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 $(1 \le s \le N)$ 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, i.e.,

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

This equation can be understood in such a way: In the atlas space, a grid of points $\boldsymbol{X}$ 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 $\boldsymbol{X}$ is retro-projected in the subject space. I~~I~~n 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 et al. (2018)). Note that $\phi'_{s,T}(\boldsymbol{X}) = \boldsymbol{X} + \boldsymbol{u}'_{s,T}(\boldsymbol{X})$. We have therefore acquired the warped motion fields $\boldsymbol{u}'_{s,T}(\boldsymbol{X})$ 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 3(b), 3(d), 3(f), and 3(h). 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 (Figures 3(i) and 3(j)).

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



~~Figure 3. (a)(b)(c)(d) 3D tongue motion of a control subject at time /s/ and /k/ in the original space and warped atlas space. (e)(f)(g)(h) 3D tongue motion of a patient in the two spaces. (i)(j) 4D 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.~~

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 0.9×0.9 mm^2, doubling that of tMRI. The same intensity atlas construction method was used on the hMRI dataset to create the vocal tract atlas (Woo et al., 2015) that contains clear internal muscle structures (Figure 2(a)). ~~Manual segmentation of internal muscles was performed by a speech scientist.~~ 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 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~~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 $M_L(\boldsymbol{X})$. 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

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

Since the 4D atlas $\bar{\boldsymbol{u}}_T(\boldsymbol{X})$ is already and average motion field, $\bar{\boldsymbol{u}}_{L,T}(\boldsymbol{X})$ 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 $\boldsymbol{u'}_{s,L,T}(\boldsymbol{X}) = M_L(\boldsymbol{X})\boldsymbol{u'}_{s,T}(\boldsymbol{X})$ as a sample of that muscle's general motion $\boldsymbol{U'}_{L,T}(\boldsymbol{X})$ (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

$$c_{L_1L_2,T} = \mathrm{corr}\left(\boldsymbol{U'}_{L_1,T}, \boldsymbol{U'}_{L_2,T}\right) = \frac{\mathrm{E}\left[\left(\boldsymbol{U'}_{L_1,T} - \bar{\boldsymbol{u}}_{L_1,T}\right)\left(\boldsymbol{U'}_{L_2,T} - \bar{\boldsymbol{u}}_{L_2,T}\right)\right]}{\sigma_{\boldsymbol{U'}_{L_1,T}}\sigma_{\boldsymbol{U'}_{L_2,T}}}. \quad (7)$$

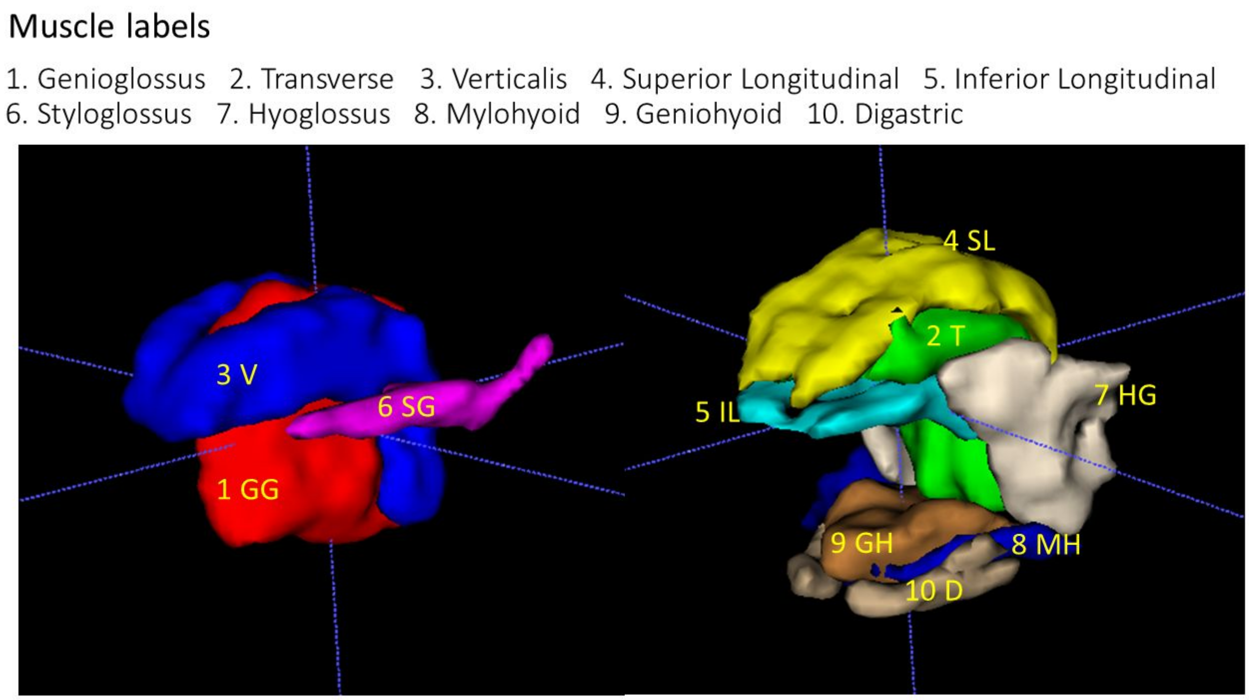
Here σ is the standard deviation of the corresponding random variable. The range of $c_{L_1L_2,T}$ spans over [-1, 1], where a high positive coordination between muscles L_1 and L_2 at time frame T yields a value close to 1 and a low coordination yields 0. After computing all correlation coefficients 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 muscle's 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 5(f), at each muscle's location, its vector field in the atlas space as a random variable $\boldsymbol{U'}_{L_1,T}$ 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 $\boldsymbol{U}'_{L_2,T}$ has multiple sample vectors as well (orange arrows). When correlation was computed between these two muscles using Eq. (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 $\boldsymbol{U}'_{L_1,T}$ and blue $\boldsymbol{U}'_{L_3,T}$, 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 non-existent.

Results

Internal Tongue Muscle Labeling

The proposed workflow was implemented using Matlab-based functions and in-house user interfaces for processing tMRI images (MathWorks, Natick, MA). 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.



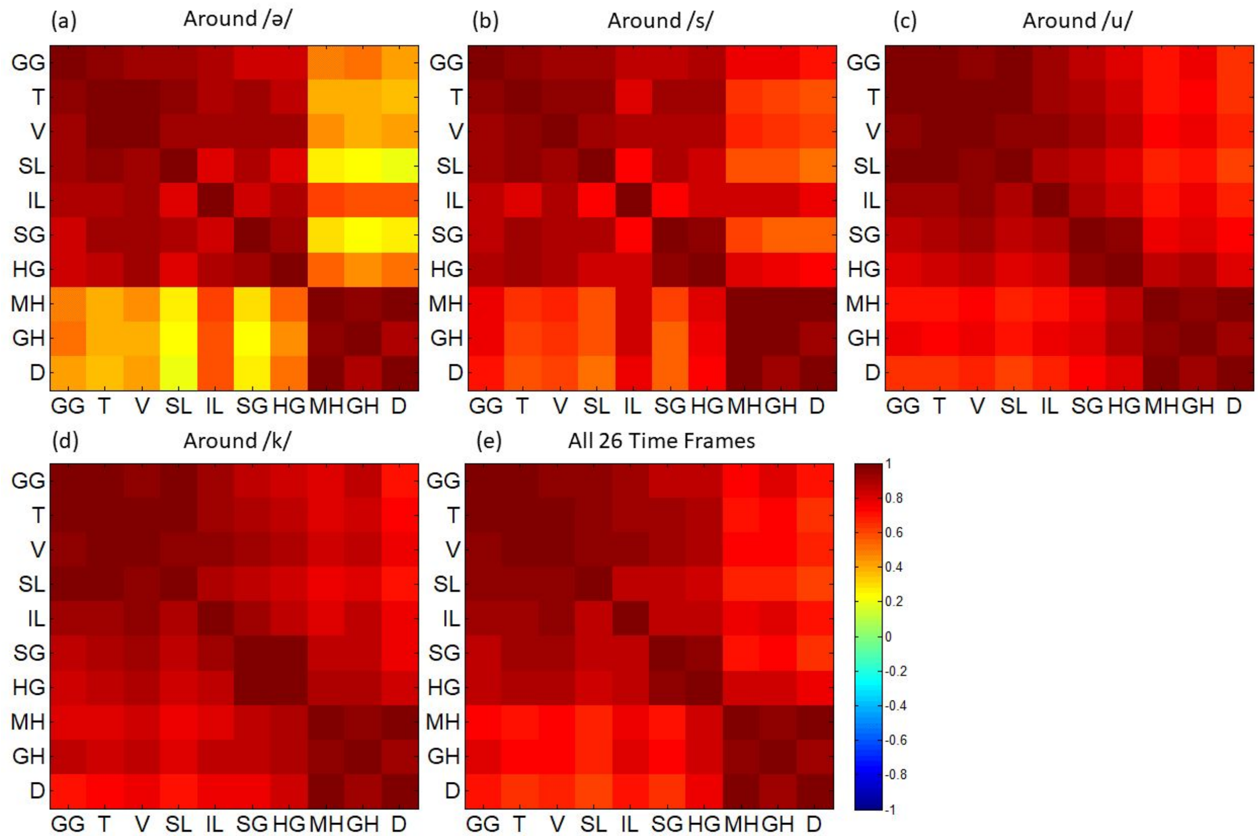
~~Figure 4. Manual labeling of the internal tongue muscles using hMRI. Muscles overlay each other so that they are shown in two separate figures.~~

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 floor-of-mouth (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 normal 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 5(a) to 5(d). The color scheme ranges from -1 as dark blue to +1 as dark red.

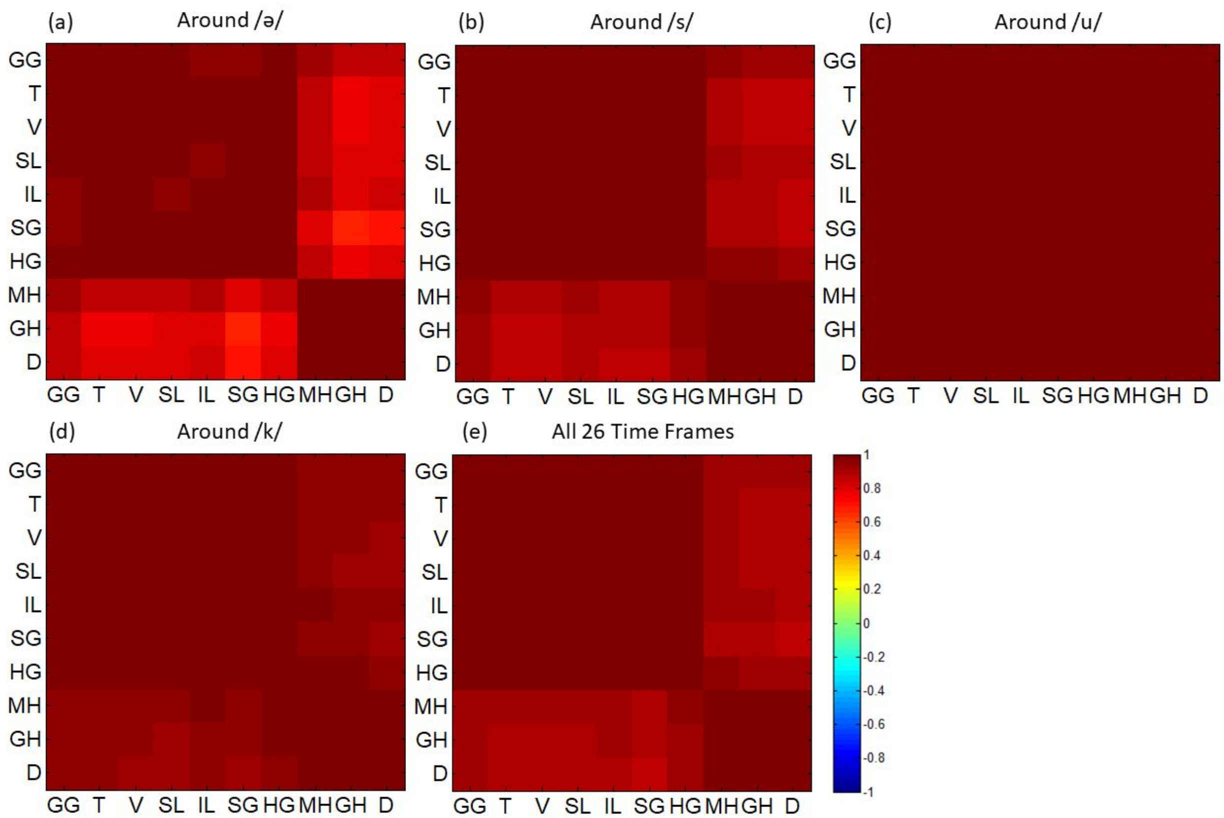
One immediate observation is that the muscle motions are all positively correlated (> 0) as expected in the Method section.~~, because t~~ This is a direct result from ~~the~~ using ~~vector~~ displacement fields ~~located at~~to represent each muscle's location. Although each muscle contracts or expands 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 0.7 and above to be a worthwhile correlation that is meaningful, Figure 5(a) shows three blocks of highly correlated muscles. These are 1) the FOM (lower right corner squares of MH, GH, D), 2) the SG-and-HG, and 3) the GG, T, V, and SL. Figure 5(a) 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 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 5(e). The FOM muscles also show less cooperation with top muscles.



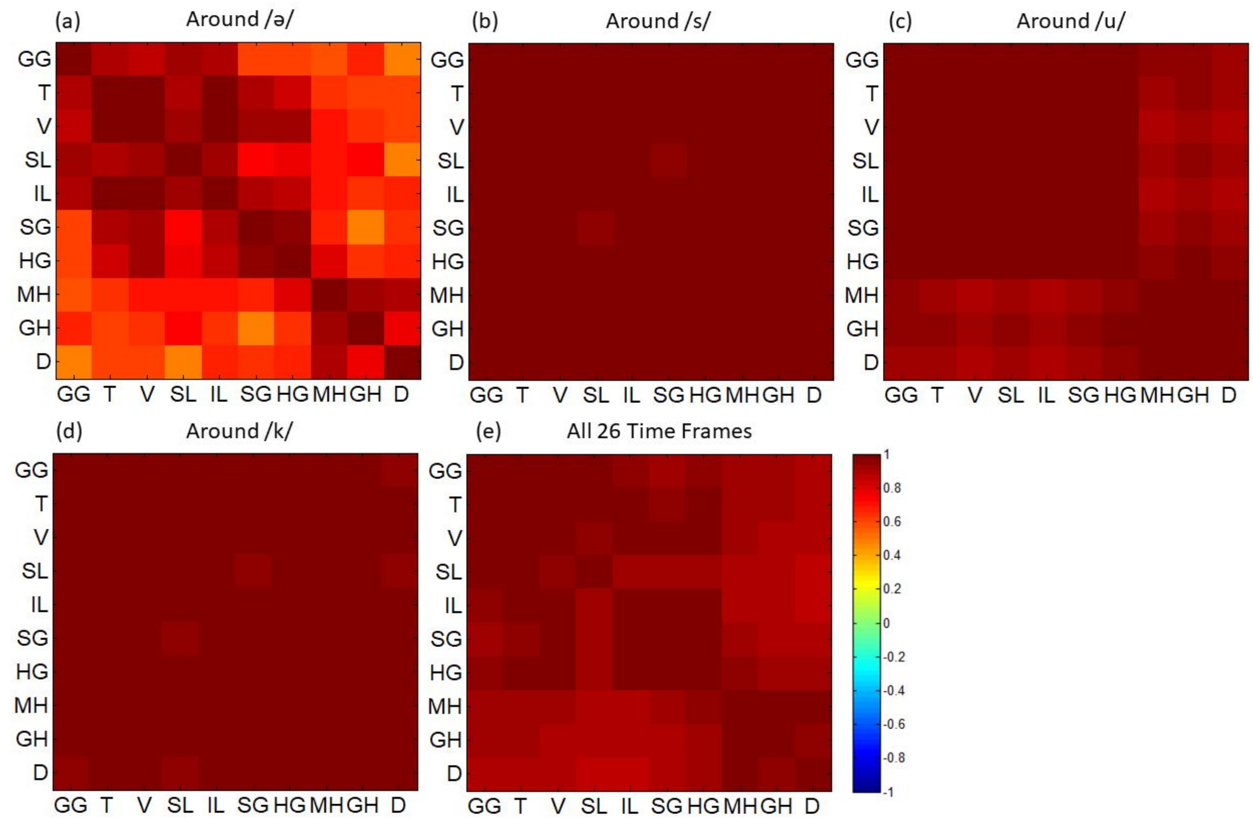
~~Figure 5. Correlation pattern of 10 tongue muscles from 16 normal controls in different time periods pronouncing “a souk”. (a) Around time frame /ə/. (b) Around time frame /s/. (c) Around time frame /u/. (d) Around time frame /k/. (e) General correlation pattern of all time frames.~~

Patient’s Correlation Analysis

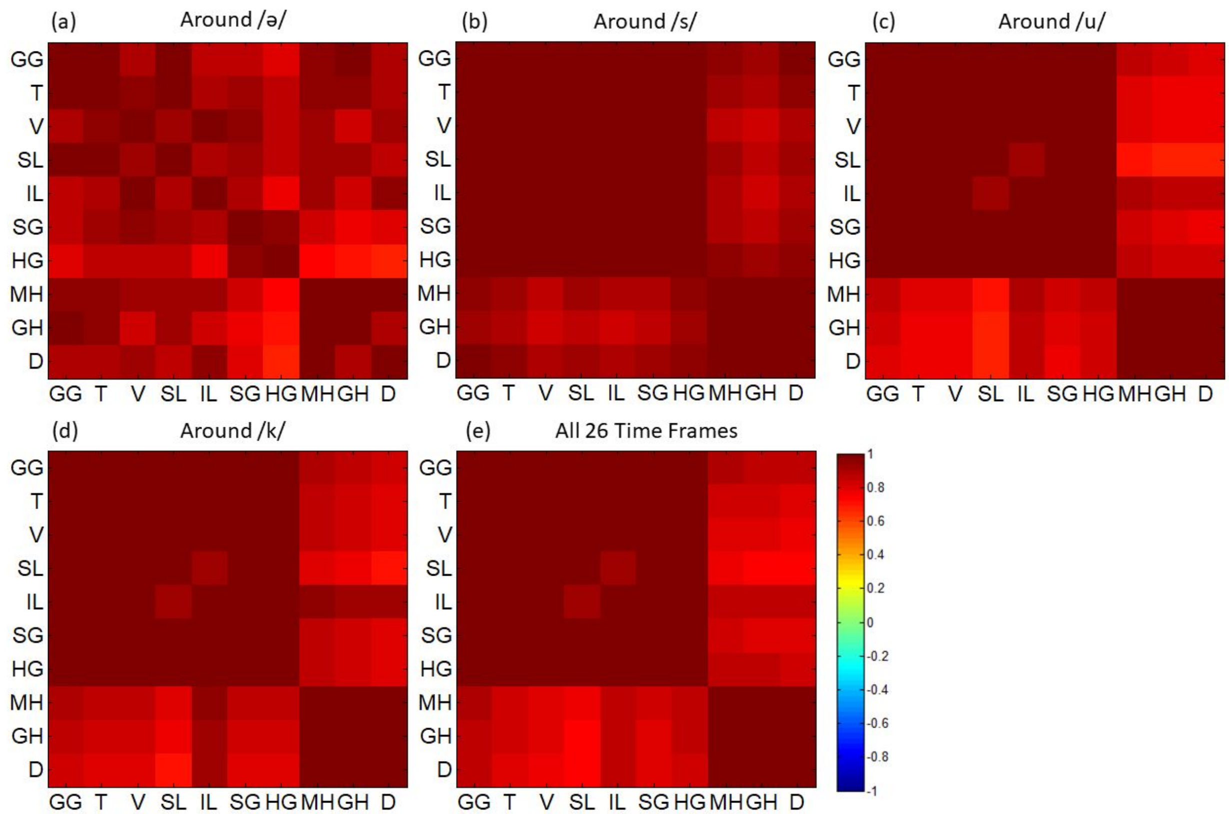
With the three post-partial glossectomy patients, we analyzed their results separately, because each patient had his/her unique surgical treatment and motion pattern. It is not reasonable to treat each patient’s unique motion as one sample from a “general patient motion pattern”. Therefore, we computed each patient’s muscle correlation pattern independently. The result of each patient is shown in Figures 6-8. Compared to the controls (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.



~~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.~~



~~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.~~

Discussion

The dataset used in this paper used only patients with a unilateral tumor occurring posterior to the tongue tip. Since the resections were all small and similar in size (2.4×2.1×1.8, 2.8×2.4×1.8, 2.2×2.1×1.4 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 as well as 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 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~~ 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 (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 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 rigid distances of 0mm, 3mm, and 6mm. For Patient 1, there was no awareness of touch or two-point discrimination in the tongue's tumor side's body or back. But 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 paper 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 like 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. And correlation matrices computed from ~~these~~ strain ~~results~~ along the muscle fiber directions may serve as a more insightful indicator for the muscle coordination patterns~~.~~, where ~~In this case,~~ 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 AP and SI direction: forward into /s/ and backward/upwards into /uk/. The simultaneous out-of-plane component (medial/lateral) is minimal and is subsumed by the main motion in the velocity field. ~~And correlation matrices computed from these strain results along the muscle fiber directions may serve as a more insightful indicator for the muscle~~

~~coordination patterns. In this case, negative correlation results are expected and could be more informative.~~ As a result, m~~M~~ore methods to enhance ~~the~~ 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 ~~emphasize~~note that current findings were only derived from this dataset and this data ~~also~~ covers limited ground. As further research develops, more insights will be gain on this topic and further comparison can be made.

Conclusion

In this paper, 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 sixteen normal subjects and three post-partial glossectomy patients. Correlation matrices in the atlas space show the coordination of tongue muscles during speech. The floor-of-mouth muscles are less coordinated from the internal tongue muscles. Patients tend to use more floor-of-mouth muscles to compensate for their post-surgery function loss.

Acknowledgements

This project was supported by NIH/NIDCD R21DC016047 (PI: Woo), R00DC012575 (PI: Woo), R01DC014717 (PI: Prince), and NIH/NCI R01CA133015 (PI: Stone).

Conflict of Interest

The authors declare no conflict of interest of this article.

References

Abd-El-Malek, S. (1939). Observations on the morphology of the human tongue. *Journal of anatomy*, *73*(Pt 2), 201.

Abd-El-Malek, S. (1955). The part played by the tongue in mastication and deglutition. *Journal of anatomy*, 89(Pt 2), 250.

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.

Bressmann, T., Jacobs, H., Quintero, J., & Irish, J. C. (2009). Speech Outcomes for Partial Glossectomy Surgery: Measures of speech articulation and listener perception. Indicateurs de la parole pour une glossectomie partielle: Mesures de l'articulation de la parole et de la perception des auditeurs. *Special Issue: Head and Neck Cancer*, *33*(4), 204.

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. (2002). *Acoustic analysis of speech*. Singular Publishing Group.

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. *Medical Imaging, IEEE Transactions on*, 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*. 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. *Medical Imaging, IEEE Transactions on*, 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)*.

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., 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., 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., 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.

Vercauteren, 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*. 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., 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), 3647-3647.

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.

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. *arXiv preprint arXiv:1804.05370*.

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.

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. In *Medical Imaging 2015: Image Processing* (Vol. 9413, p. 94131L). International Society for Optics and Photonics.

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., 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., 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. In *Medical Imaging 2018: Image Processing* (Vol. 10574, p. 1057411). International Society for Optics and Photonics.

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.

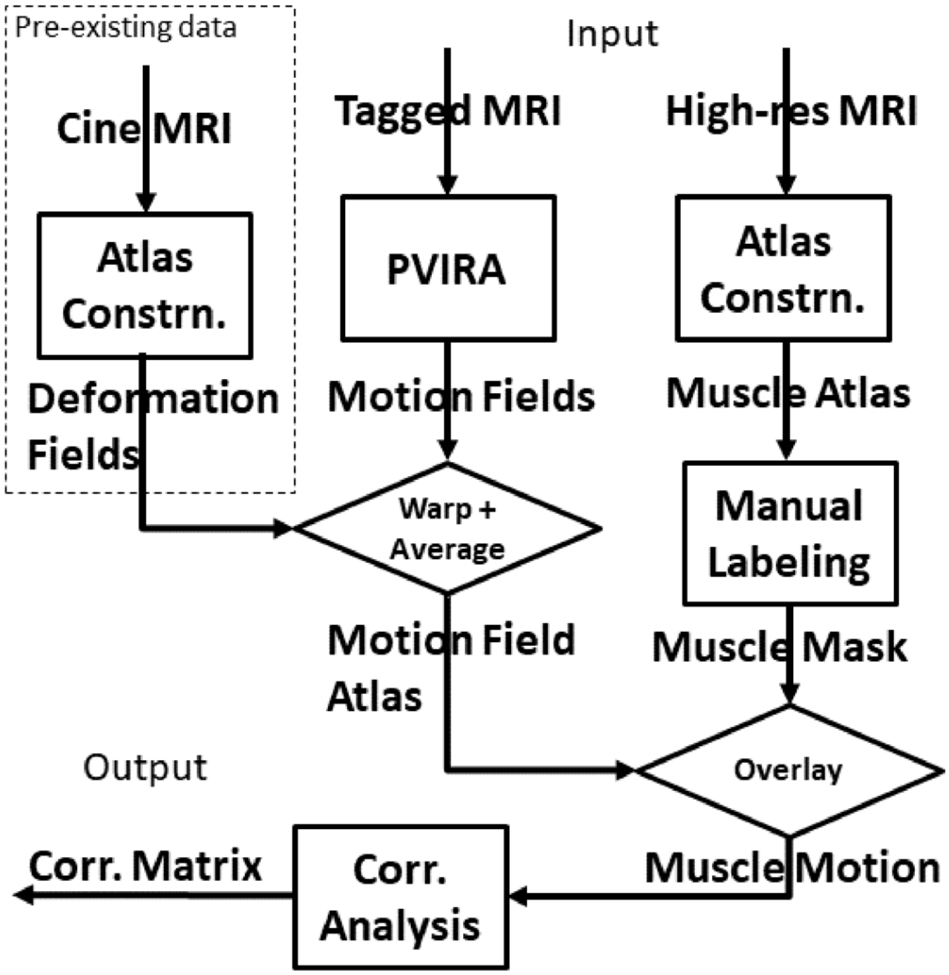


Figure 1. Workflow of the proposed method.

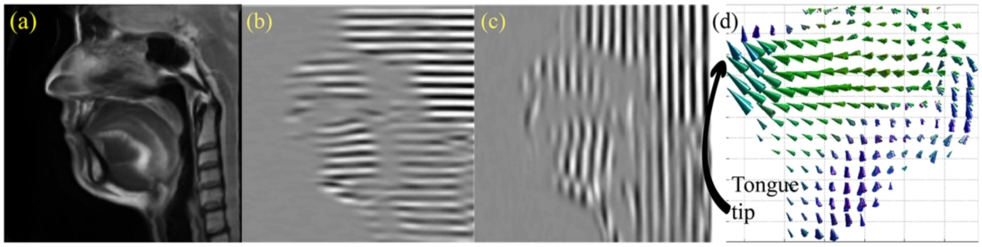


Figure 2. (a) High-resolution MRI of sagittal tongue. (b)(c) Tagged MRI in two directions. (d) Estimated 3D tongue motion in a sagittal view.

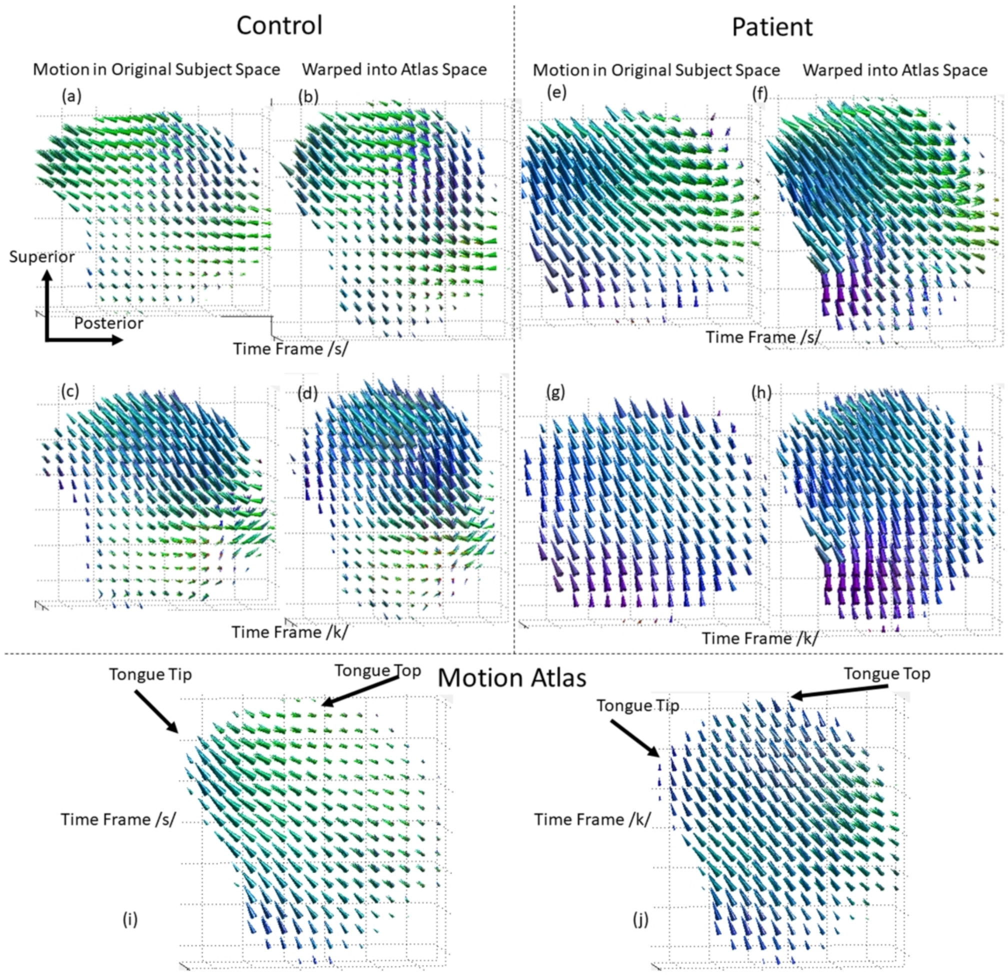


Figure 3. (a)(b)(c)(d) 3D tongue motion of a control subject at time /s/ and /k/ in the original space and warped atlas space. (e)(f)(g)(h) 3D tongue motion of a patient in the two spaces. (i)(j) 4D motion atlas of /s/ and /k/ combining 16 control subjects. Note that color-coded cones are used to visualize motion, where anterior-posterior vectors are coded green, inferior-superior are coded blue, and left-right are coded red.

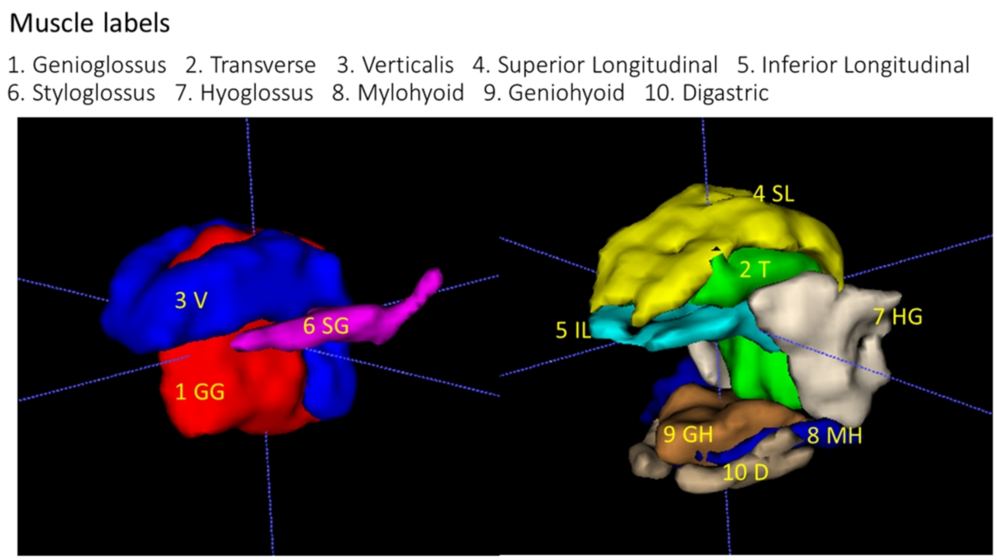


Figure 4. Manual labeling of the internal tongue muscles using hMRI. Muscles overlay each other so that they are shown in two separate figures.

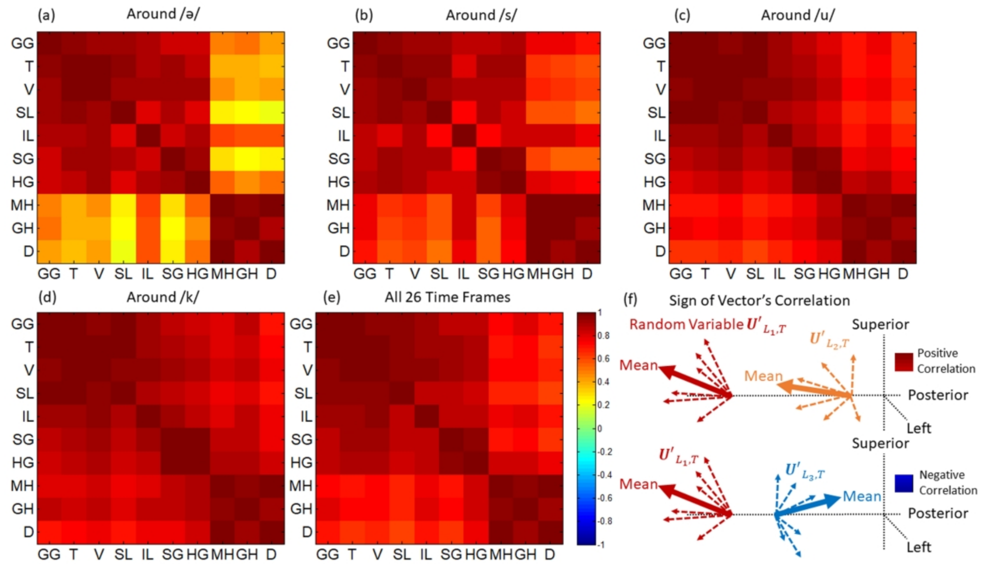


Figure 5. Correlation pattern of 10 tongue muscles from 16 normal controls in different time periods pronouncing "a souk". (a) Around time frame /ə/. (b) Around time frame /s/. (c) Around time frame /u/. (d) Around time frame /k/. (e) General correlation pattern of all time frames. (f) Explanation of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

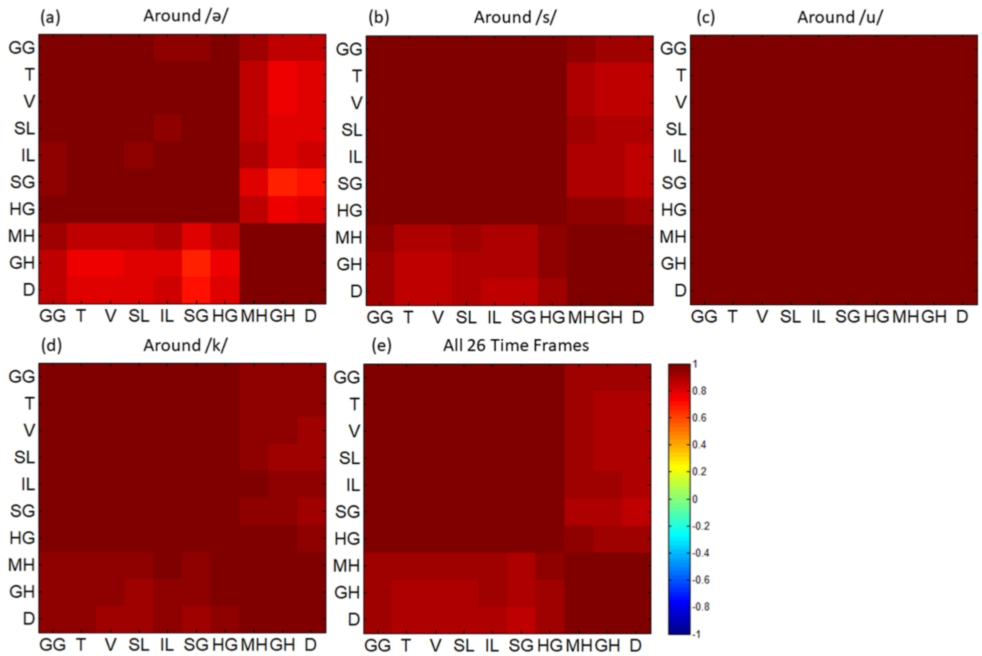


Figure 6. Correlation pattern of 10 tongue muscles from post-partial glossectomy Patient 1 pronouncing "a souk". (a)-(d) Four critical time frames. (e) All time frames.

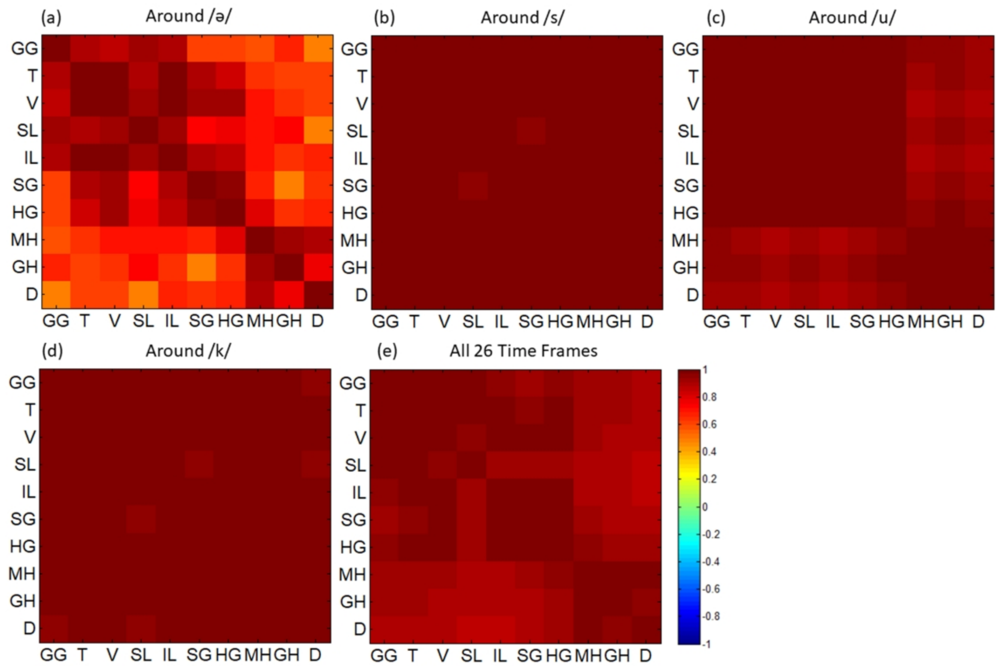


Figure 7. Correlation pattern of 10 tongue muscles from post-partial glossectomy Patient 2 pronouncing “a souk”. (a)-(d) Four critical time frames. (e) All time frames.

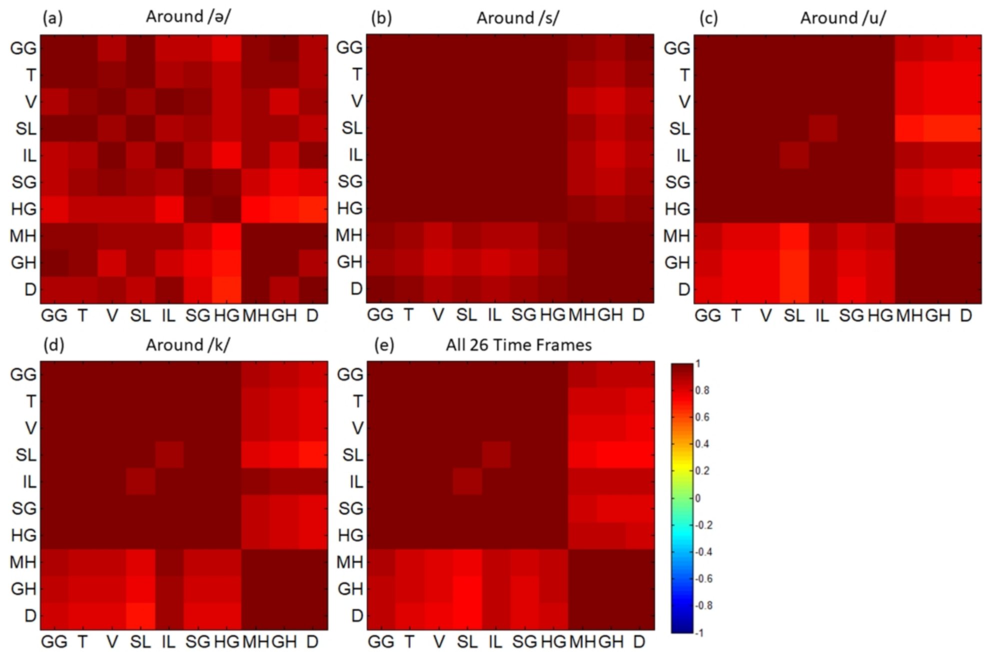


Figure 8. Correlation pattern of 10 tongue muscles from post-partial glossectomy Patient 3 pronouncing "a souk". (a)-(d) Four critical time frames. (e) All time frames.

Table 1

Tongue Motion MRI Scan Parameters

Region of Interest	Field of View (mm^2)	Resolution (mm^2)	Slice Thickness	Frame Rate	# Slices (Subject dependent)		
					Axial	Sagittal	Coronal
Tongue & surrounding tissues	240×240	1.88×1.88	6 mm	26 frm/sec	10-14	5-9	10-14

Table 1. Tongue Motion MRI Scan Parameters

Response to Review Comments

Title: Atlas-based Tongue Muscle Correlation Analysis from Tagged and High-resolution MRI

Journal: Journal of Speech, Language, and Hearing Research

Initial reference number: JSLHR-S-18-0495

We thank the editor and reviewers for the careful examination of our previous manuscript. We have carefully addressed all issues raised by the reviewers and have modified the manuscript accordingly. In light of the new material that were added for clarification or more explanation, some parts of the original manuscript have been rewritten for both clarity and brevity. The details of our review and modifications are described below in a numbered list; the referees' original comments are in boldface. Accordingly, those changes can be quickly found in the marked version of the manuscript.

REVIEWER COMMENTS

1. Overall Strengths: / 2. Importance: / 3. Justification/Rationale:

There are no comments to address in these sections. We thank the reviewers for appreciating the technical contribution and our muscle-level investigation using the multi-subject analysis technology described in this work. Especially, the study of floor-of-mouth muscles was a major investigation that we decided to include when designing the experiments since these muscles' importance was often less studied as the reviewers have remarked. Plus, breaking down speech and muscle studies into simpler analyses was often difficult and our work presented one feasible way to achieve this. We thank the reviewers for acknowledging our motivation/justification and this study's importance for future speech work. We will address all detailed problems below.

4. Methods/Approach:

Reviewer 1: One suggestion would be to add some detail on how the manual segmentation of the tongue muscles was validated. I also wonder if there was any sensory loss to the tongue due to the glossectomy. If so this could impact coordination of motor function. Perhaps I missed it, but I don't see any reference to ethics permission.

1) The paper *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* extensively explains how the muscles were segmented

independently by two speech experts for the 3D atlas. Especially, in Section 2.2 of this reference paper, there are 3 paragraphs explaining all related anatomy studies, how the muscles are segmented, and identifying muscles that couldn't be measured well with current technology. More evaluation and properties of the segmentation result were also presented in the results section of the paper. This important reference paper was cited in our original manuscript in line 25 of page 3, line 19 of page 6, and line 52 of page 25. Since the reviewer specifically raised this validation question, we now emphasize the above arguments and direct the readers more to this reference paper in the revised text (Method, Muscle Correlation Analysis, before Eq. (6)).

2) 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 rigid distances of 0mm, 3mm, and 6mm. For Patient 1, there was no awareness of touch or two-point discrimination in the tongue's tumor side's body or back. But 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. We added this information as a new third-to-last paragraph in Discussion. We note that this could also be added to the method section in Data Acquisition describing the patients. But since these three patients showed similar conditions of sensation loss, its impact on muscle correlation is difficult to determine for now. So we kept the paper more focused and only added this part in Discussion for the future.

3) We guess what the reviewer meant by "ethics permission" is related to the subjects' consent info and HIPAA regulatory to participate in the study. And yes, all subjects read and signed a consent form and the HIPAA form before the study, and the entire protocol was approved by the University of Maryland Baltimore internal review board. We thank the reviewer for pointing this out and now added this clarification in the Method section in Data Acquisition.

Reviewer 2: The work relies on a complex process intended to align all the data of several speakers (time and space) onto a reference tongue. The realignment process is reported in other papers of the authors and only a small number of equations summarizes the different stages. The outline of the spatial alignment becomes a little bit obscure due to the notations which are not explained (prime in Eq. 4 and following probably denotes the time and space aligned data). It would be easier to explain Eq. 4 it with a sentence, and to say that the movement of a point in the reference tongue is obtained by considering the movement of that point retro-projected in the current image (if I understand well).

We thank Reviewer 2 for pointing out the math problem. We understand the equations are kind of "vague" for being simply re-summarized in this paper to keep its conciseness. In this revision, we now explained the deformation process related to Eq. 4 in a more detailed way. We hope now the mathematical process would be simpler to understand. And yes, "the movement of a point in the reference tongue is obtained by considering the movement of that point retro-projected in the current image" is the correct understanding. And retro-project is a great word for this. We borrowed it in the new explanation following Eq. (4).

Similarly, the sentence corresponding to Eq. 6 should be completed by saying that this is the average motion. The accumulation of small details (see also other remarks below) eventually results in the loss of the reader.

Similarly, we further explained Eq. (6) with more details and clarified the meaning of each variable.

The segmentation of the tongue into muscles is interesting, but it is difficult to know on what basis it was obtained ("Manual segmentation of internal muscles was performed by a speech scientist" line 22 page 16). Much of the work is based on the assumption that all the data can be projected onto a single decomposition of the tongue into muscles. Is this segmentation speaker independent? This point should be addressed because the computation of the correlation between the muscles' motion depends on this assumption. The diffeomorphic deformation probably does not suffice to compensate for this likely speaker variability. The second point is critical and should be answered by the authors in the paper.

This concern is very similar to what Reviewer 1 had also mentioned: the validity of the manual segmentation. First, we now clarified the segmentation process (see response to Reviewer 1). Second, the segmentation is in the atlas space only so it can be considered the segmentation of a "general average tongue anatomy" (i.e., speaker independent), rather than speaker-specific. If we accept its validity, only this one set of segmentation is necessary. And this is the whole point of constructing an atlas space. Third, regarding speaker variability, since variability is almost very difficult to measure independently, we have to rely on an assumption that some global method is capable of accounting most of the variability in its operation. And in this case, we considered it being diffeomorphic deformation, which is mostly used in many medical image analysis applications to account for anatomic variability. Of course, diffeomorphic deformation cannot be the only perfect model for all possible variability and we acknowledge this situation, but given current image processing technology, it is likely to be the most efficient model to apply. We clarified all of these points in the revision (Method, 4D Motion Atlas Construction, before Eq. (4)) since this is the basis of all of our proposed methods. We thank Reviewer 2 for raising its importance.

5. Results/Findings:

Reviewer 1:

No comments to address in this section.

Reviewer 2: The critical point, which could be related to the previous remark (in the approach) is the correlation analysis. All the correlations between muscles are positive. The explanation provided by the authors (page 19, bottom) is not very convincing. This point is also addressed in the discussion (page 26, bottom). The fact that all correlations are positive is counter-intuitive, because the tongue is an object whose volume does not change. Therefore, there should be motions which are negatively correlated. The fact that velocity fields is considered should not change the results. The authors should provide a better explanation.

We understand that an all-positive correlation result might be difficult to imagine, and simply explaining the reason in words might be even harder. As a result, we made a new illustration just to clarify this issue (see updated Figure 5(f)). Essentially, this is a mathematical problem regarding the computation and definition of correlation matrices using vectors. Eq. (7) is the key. At each muscle's location, we consider its vector field in the atlas space as a random variable $\boldsymbol{U}'_{L_1,T}$. Because it is random, it has many sample vectors from all subjects pointing to many directions (red dashed arrows in Figure 5(f)) with a mean (red solid arrow). These sample vectors generally point to a similar direction because they are samples of the same muscle, which generally moves to the same direction despite different subjects. Now we check another muscle and its random variable $\boldsymbol{U}'_{L_2,T}$. It has many sample vectors too (orange arrows). When we compute the correlation between these two muscles using Eq. (7), we see that there is essentially a "dot product" operation between these sample vectors in the numerator part. When the two sets of vectors generally point to the same direction such as red and orange, the dot product yields a positive number. On the other hand, when the two sets of vectors generally point to opposite directions such as red $\boldsymbol{U}'_{L_1,T}$ and blue $\boldsymbol{U}'_{L_3,T}$, the dot product yields a negative number. Now, if we look back to Figure 3 for some examples of tongue vector fields, we see that these vectors follow a same trend almost everywhere in the whole tongue, no matter what muscles they are in. In other words, there are no two muscles where one has forward displacement and the other has backward displacement. That would tear the tongue apart. As the reviewer pointed out, the tongue is an object whose volume does not change. Its muscles will smoothly follow a similar general trend to *not* tear the tongue apart. Therefore, the existence of any negative correlation is not only unexpected, but also disobeys the tongue's incompressible property, and signals a potential mistake in the pipeline.

Again, we emphasize that this positive result is unavoidably caused by using the vector fields and the mathematical property of an "inner product". If we could develop a way to compute principal strains (which is a set of scalars), that would describe a coordination pattern between the simultaneous *activation* of muscles. And we surely expect negative correlation between their *activations*. 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 AP and SI direction: forward into /s/ and backward/upwards into /uk/. The simultaneous out-of-plane component (medial/lateral) is minimal and is subsumed by the main motion in the velocity field. We have added all these explanations to the end of the method section following Eq. (7) as a new paragraph and also enhanced the related paragraph in the discussion section. We hope that with the help of this new Figure 5(f) and the updated/more detailed explanation, this issue can now be clarified.

page 5 – lines 25-31 and pages 9-10. It is not clear whether there is only one acquisition for the three directions, or three acquisitions synchronized with a metronome. It could be clear for tagged MRI specialists, but it is not clear for others.

There were three acquisitions for each image slice, synchronized with the repeating rhythm of the metronome. We have clarified this in the revision (Method, Data Acquisition).

page 11 – line 31. The reader has to wait until line 55 the explanation for X which appears at line 31.

We have now addressed this notation sooner this in the revision.

page 12 – lines 55. Time points used to define critical time frames for each of the four sounds of “souk” and are defined as a “maximum tongue position”. Which are these maxima? Tongue top for /s/, tongue dorsum contact with the palate for k probably, but what for the two vowels? This should be described precisely. How robust against speaker variability are these time points, especially for the three patients?

We agree that “maximum tongue position” is a vague concept. We thank the reviewer for pointing this out. These positions were determined like this: the two consonants were defined as the first time frame in which the tongue (tip for /s/ and body for /k/) made contact with the palate. The vowels were defined as the last frame before the tongue began moving towards the consonant. For the schwa, the tip started to extend and the body started to move forward and up. For /u/, the motion change was when the tongue started moving directly up, instead of back and (maybe) slightly up. We added this explanation to the method section “4D Motion Atlas Construction”, replacing the vague concept. Since the process for determining these critical time frames was completely manual from visually checking the raw MRI data and was performed on a subject-by-subject basis, all speaker variability was individually one-by-one accounted for and manually checked for every subject, including any patient. This manual process was described in the same section before the “maximum tongue position” in the original manuscript, so we did not re-iterate this in the revision.

6. Discussion/Conclusions:

Reviewer 1: One topic that could strengthen the discussion is whether the /e/ sound requires correlated modules since this sounds requires less tongue muscle activation that other muscles. I would have liked a comment about the seeming independence of the inferior longitudinal muscle.

We believe the reviewer is referring to the schwa "uh" sound. Actually, there is no indication that the schwa requires less tongue muscle activation than the other muscles. It is called a 'neutral vowel' because the vocal tract shape during schwa is similar to a uniform cross-sectional tube. The tongue, however, still needs to deform into the proper shape from its previous one. Therefore, it is likely that the tongue uses correlated regions when making this sound. In this study, its occurrence at the beginning of the task makes it difficult to identify what it has deformed from, however.

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, 8). This separation reduces friction and co-contraction with other muscles when it is active and may facilitate its role in elevating the tongue.

Abd-El-Malek, S. (1939). Observations on the morphology of the human tongue. Journal of anatomy, 73(Pt 2), 201.

We have added the IL discussion and the new Abd El Malek, 1939 reference.

Reviewer 2: There is very little work to compare this paper to. Therefore the interpretation of the results made by the authors can hardly be supported by others results.

We understand the limitation of existing literature and comparability. We could only do our best to explain these findings derived from our current data. When further research comes, we believe more insights can be gain on this topic. In this revision, we mentioned this limitation in the discussion part to make it more rigorous.

The conclusion about the correlation between muscle motions is really surprising (always positive correlations). Explanations are neither sufficient nor convincing.

We have addressed this comment in previous response for part 5.

Authors should be more specific about the limitations and perhaps add others, such as the sequence chosen.

Besides the use of displacement fields rather than strain as a major limitation (which is also related to the positiveness of correlation), we have added more limitations such as the narrow variety of patient type, incapability of sensation loss check, lack of comparisons of the current work in the discussion section. The tagged MRI sequence used in this study is pretty standard for dynamic imaging (derived from cardiac imaging technology). Therefore, it is very difficult to further increase the speed or efficiency by the sequence itself. So we chose not to add a comment about the selection of sequence.

7. Additional Remarks: / 8. Final Summary:

No comments to address in these sections.

EDITORIAL COMMENTS

*** On the manuscript title page:**

- For the Corresponding Author, please include an appropriate telephone number.

We added a contact number.

- Please include a Conflict of Interest statement, with header Conflict of Interest. If there are any relevant conflicts of interest please list them or if none state that there are no relevant conflicts of interest.

We added the statement.

- In the Funding statement, where possible, please state the name of the recipient of the funding.

We updated this information accordingly.

*** Please do not embed figures in your manuscript file. Each figure must be submitted only as a separate graphics file. Acceptable file types include JPEG (.jpg, .jpeg), PNG (.png), TIF (.tif, .tiff), PDF (.pdf), Excel (.xls, .xlsx), or PowerPoint (.ppt, .pptx) file.**

We fixed the embedding.

- Be sure to include figure callouts in the manuscript.

We double-checked this.

- Please place the figure legends after the References section of your manuscript (not embedded in the text).

We fixed the embedding.

- Please note that all figures must have a minimum resolution of 200 dpi (preferably 300 dpi or greater).

We double-checked the dpi.

*** Please also note - Dr. El Fakhri has not completed the digital copyright form. Please ask them to go to the Author dashboard in ScholarOne, find the correct manuscript, and complete the copyright forms.**

Dr. El Fakhri has now completed the form.