Speech patterns in a muscular hydrostat: normal and glossectomy tongue movement

Maureen Stone[1,2], Xiaofeng Liu[4], Hideo Shinagawa[1], Emi Murano[1], Rao Gullapalli[3], Jiachen Zhuo[3], Jerry Prince[5]

[1]Department of Biomedical Sciences, [2]Department of Orthodontics,.
University of Maryland Dental School, Baltimore, MD, USA.
[3]Department of Diagnostic Radiology,
University of Maryland School of Medicine, Baltimore, MD USA
[4]Dept. of Computer Science, [5]Dept. of Electrical and Computer Engineering,
Johns Hopkins University, Baltimore, MD, USA

1. INTRODUCTION

This work is intended to explore the relationship between tongue muscle activity and motion in glossectomy speakers. We are motivated by the need for a greater understanding of the mechanisms that underlie tongue motion, in order to better interpret clinical observations, and to provide data that can help predict optimal surgical outcomes. In order to speak intelligibly, glossectomy patients learn to compensate in varying degrees to the changes in tongue morphology caused by their surgery. However, it is presently difficult to relate postoperative clinical observations about speech quality back to the particulars of the glossectomy surgery. Clearly, there is a need for a greater understanding of the biomechanical and muscular mechanisms that underlie compensatory tongue motion in order to provide outcome data that can assist surgeons in planning and evaluating reconstructive surgery. The goal of this project is to apply a mechanical analysis to motions of the inferior longitudinal muscle (IL), which is often cut during partial and hemiglossectomy surgery. We have looked at IL's mechanical deformations during speech for a normal speaker and a patient with a partial glossectomy and a radial forearm free flap (RFFF) to examine whether differences in tongue motion post glossectomy could be explained by changes in muscle mechanics.

The MRI techniques used in this study have not been used often with patients, because they require abilities not available to most patients. For example, high-resolution MRI and Diffusion Tensor MRI require the ability to hold the tongue still, without swallowing, for up to three minutes. Cine-MRI and tagged Cine-MRI require many precise repetitions of speech utterances because the final data set is a summation of multiple repetitions. It is still not possible to use these techniques with neurologically disordered patients, who have large inter-repetition variability. However, glossectomy patients do not have central processing deficits or motor control problems that preclude precise repetitions and long sustained postures. For these type behaviors the glossectomy's abilities are normal. Therefore, they are almost uniquely able, as patients, to undergo these procedures and benefit from their interpretation. Although glossectomy patients fatigue more rapidly than normal speakers (normal speakers never experience tongue fatigue), these procedures are nonetheless within their abilities.

Diffusion Tensor Imaging (DTI) captures muscle fiber orientation from diffusion weighted images (Gilbert and Napadow, 2005). DTI quantifies the tissue orientation by imaging the orientation-dependent diffusion processes that are associated with fibers such as myocytes and axons and their bundles (Basser et al, 1994). Initially DTI was used to image excised tissues [porcine tongue, for example, in (Napadow et al., 2001)], but now in vivo human imaging in the resting state is possible (Gilbert and Napadow, 2005, Shinagawa et al, 2006). Diffusion orientation is typically characterized by a tensor representation (Basser et al, 1994) where the dominant orientation of the tensor represents the orientation of a fiber (muscle in our case) passing through the voxel of interest. DTI works better with bundled muscles (such as inferior longitudinalis) rather than interdigitated muscles

Tagged Cine MRI (tMRI) tracks the tongue tissue – point-by-point within the volume of the tongue – in order to help identify the compensatory patterns of motion. tMRI originally was pioneered for use in the heart to measure motion and internal tissue characteristics (Axel et al, 1989, Zerhouni et al, 1988). Although EMG is the best method to study muscle activity directly, tongue EMG is difficult to interpret due to interdigitated, orthogonal muscles, such as Transversus and Verticalis (Takemoto, 2001). Tagged cine-MRI provides data complementary to EMG. Tag motion shows tissue motion, which means that the direction of fiber compression is accurately determined in orthogonal muscles that cannot be differentiated by EMG. Muscle compression may parallel muscle activation, though tMRI cannot measure activation of specific muscles in the way that accurately placed fine-wire EMG electrodes can. tMRI has been used by several groups to study internal tongue deformation. Most of this work used snapshot tMRI to capture two moments in time from a continuous motion: a "reference" image and a "deformed" image; the second occurs later in time (Niitsu et al, 1994). A modification of this technique is to create a movie or image-sequence by having the subject repeat the task multiple times. During each successive repetition the deformed image is taken at a later moment in time. The deformed images are then reconstructed into a movie (Napadow et al, 1999). Our group has pursued an alternate approach in which tagged data is continuously acquired throughout a single task repetition, yielding tagged Cine-MRI. The signals acquired during a single repetition are too weak to provide good image quality. Therefore, the cine images for each specific time frame of each repetition are assembled in Fourier space in order to generate an MRI movie of the task (Parthasarathy et al, 2007). This is the equivalent of ensemble averaging.

2 METHODS

2.1 Subjects and Task

Two subjects were studied in this experiment. One normal speaker (hereafter NL) and one partial glossectomy with flap (hereafter PT). NL was a 26 y.o. non-native speaker of English (native language: Japanese) who had no difficulty saying the task word as all the sounds are similar to those found in Japanese. PT was a 43 year old American English male who was seen 1 yr after a partial-glossectomy that removed a left lateral SCC T1N0M0 tumor of the lateral oral tongue, preserved the tongue tip and included a microvascular forearm free flap reconstruction. His tumor was on the tongue dorsum and it was decided to reconstruct with a flap to replace the resected tissue, increase bulk, and maintain the mobility of the tip. The goal was to provide enough tongue volume to touch the palate and facilitate a functional swallow. PT's speech was mostly within normal limits, but with a slightly effortful pronunciation. He reported marked effort when tired. His /s/ was the only impaired sound acoustically and was produced with a noisy emission (lateral lisp).

The speech task was "a souk." The word was chosen because the motion from "s" to "u" moves the tongue backwards and retrudes the tip, the motion from "u" to "k" elevates the tongue body. Tongue backing in "su" might engage inferior longitudinalis to retrude the tip; this muscle was cut on the resected side.

2.2 Recording Procedures.

MRI recordings.

Diffusion Tensor MRI (DTI) and high resolution MRI (hMRI) scans were made to provide anatomical information. The hMRI data was used to establish the region to seed for the DTI data collection. For these recordings the operator scanned the oral cavity area while the subject held still in a quiet breathing task. HMRI was collected three times, once for all the slices in a single direction with a 3mm slice thickness and a spatial resolution of 0.94 mm/pixel (see Figure 1). The DTI scan was in the same axial plane as the hMRI. Each hMRI scan took between one and two minutes; DTI (done once) took from 3-5 minutes. Tagged cine-MRI (tMRI) images were collected in multiple planes, while the subject repeated speech tasks to the beat of the auditory rhythm cue. Cine-MRI was collected along with tMRI to allow better

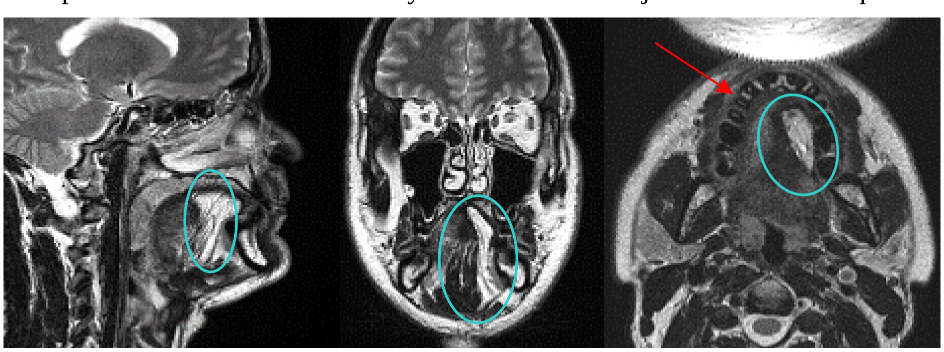

Figure.1. High resolution images of the tongue for PT in three planes that intersect the flap. Flap is white due to high fat content.

determination of tissue edges and had a resolution of 1.875 mm/pixel. These techniques had a 6 mm slice thickness and were aligned so that one 6mm slice contained 2 of the 3mm slices. One tMRI dataset required four sequential data collections. That is, two CSPAMM image sequences, each containing one tag direction, were repeated twice with phase shifted tag patterns. To collect this data, the subject repeated each speech task 4 times per slice per collection, resulting in 80 to 100 repetitions with four pauses.

Acoustic cues to trigger MRI and synchronize speech repetitions.

The speaker must repeat the utterance precisely for good image quality, because multiple repetitions are averaged during MICSR data collection. Recently we have augmented the UMD MRI facilities by developing an MRI trigger system that uses acoustic cues to synchronize speech utterance repetitions with MRI acquisition. The protocol for synchronized auditory cueing is based on the method of Masaki and colleagues (Masaki et al.1999, Shimada et al., 2002). This synchronization technology improves the precision of the subject's repetitions. In this method, the audio system of the MRI console delivers short white-noise pulses through headphones at predetermined intervals to the subject, and triggers the MRI acquisition synchronously. The subject utters syllables and breathes in time with the pulses. The number and timing of pulses is varied to accommodate the number of syllables and the final breath. A 15-minute training protocol, with feedback from the experimenter, was developed using nine normal subjects. This protocol was used successfully with the glossectomy patients. In the MRI machine subjects wore headphones to reduce noise, and to hear the experimenter and the acoustic cue.

2.3 Data Analysis.

Determine 3D location of IL. DTI was used to identify the fiber bundles within the inferior longitudinal (IL) muscle on the preserved side of the tongue. Therefore, IL is very well visualized on DTI. IL is often cut during glossectomy surgery, even in partial resections, thus the muscle on the intact side has no pair to work with and must create new motion patterns. The seeding process for IL is shown in Figure 2. The muscle location is estimated from known anatomy (a) and visually identified on the hMRI image (b). A seed is planted within the bundled muscle, in the DTI data and the region is 'grown' along the tissue fibers that emanate from the seed point. For patients seeds are planted at several locations of the tongue (c) to account for breakage or disturbance due to surgery. DTI data are collected from axial slices and reconstructed into a 3D volume, which can be sliced in any plane. Fiber directions for IL extracted from DTI are in Figure 3 for NL and PT. PT has severed tissue on the left side and the fiber direction for the intact IL is green posteriorly and becomes blue anteriorly, because the fibers angle

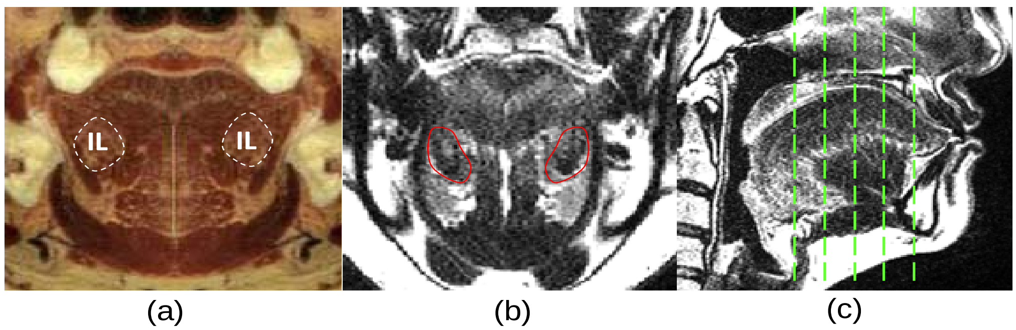

Figure 2. The seeding process used in DTI to extract IL tongue fibers. IL location was estimated from anatomical and histological data (a). The bundled fibers were identified on high-resolution coronal MR images (b) and the seed was planted in the proper location, in spatially aligned axial DTI slices. For PT, seeds were planed in multiple locations based on the coronal images due to large changes in fiber direction.

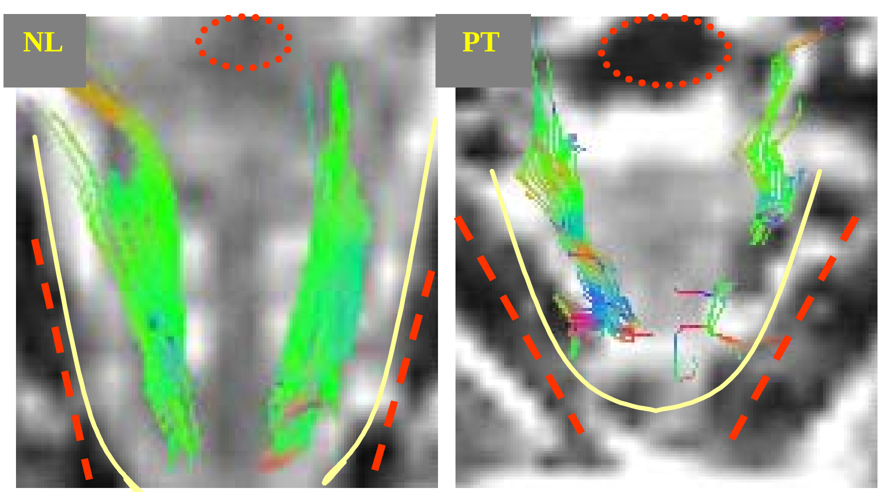

Figure 3. Axial DTI images show IL muscle (green) for the two subjects NL and PT. The tongue is outlined in yellow, the pharynx is circled in red, and the mandibular bones are tracked in red.

upward in the front of the tongue. This deviation of IL may have been caused by surgical reconstruction or scar formation.

Motion Tracking of IL. An 'average fiber' for IL was extracted from the bundled muscle in the intact side for both subjects. This fiber was overlaid on the tMRI data in order to track its motion during the word "asouk." The fiber was overlaid onto the sagittal tMRI plane most closely corresponding to the fiber plane in the DTI data. For NL it was plane 2 of 7, with plane 4 being the midline. For PT it was plane 3 of 7. HARP-3D automatically tracks the movement of the fiber nodes. The initial HARP phase values of the grid points of this fiber are computed in space based on its position in the DTI data. The first sound of the utterance is a schwa, a fairly neutral tongue position, to optimize fiber placement by simulating the rest position used in the DTI data. 2D-HARP tracking is applied on each tissue slice of the image sequences, thereby giving partial knowledge of the components of 3D motion on all image planes. The points of intersection between the fiber and the image planes are then computed, which give a sparse set of motion observations of the fiber (from the previous 2D HARP computations). Each component of motion is then interpolated using a 3D thin plate spline so that the motion on all nodes of the fiber is known. The fiber position is updated until no motion is implied anywhere on the fiber (within a small tolerance) and then time is incremented, and the steps are repeated (Liu et. al., 2006).

Global Motion and Fiber Elongation.
Based on the position of the component points of the fiber at each moment in time, HARP-3D computes the biomechanical properties of the muscle between later time frames and the first time frame. These properties, shown in Table 1, are global, or rigid, motion (x, y translation and rotation), change in fiber length, and bending energy .

Bending Energy
To measure the global bending of the tongue muscle fiber, we used the bending energy model proposed in Duncan et al (1991). For a thin straight rod, let s be the length along the rod of any points on the rod, and R(s) be the radius of curvature. The elastic bending energy of the rod after deformation can be written as:

$$E_{abs} = \frac{1}{2} EI \int_{s_1}^{s_2} \frac{1}{R(s)^2} ds$$

where E is the Young's modulus, and I is the inertial component. The integral is over the entire length of the rod. Here we assume E and I are constant, and so can be ignored in the equation. We call this the absolute bending energy. For a muscle fiber that is not straight even at the rest position, the bending energy after it deforms is expressed:

$$E_{rel} = \frac{1}{2} EI \int_{s_1}^{s_2} \left(\frac{1}{R_t(s)} - \frac{1}{R_0(s)}\right)^2 ds$$

where $R_t(s)$ and $R_0(s)$ are the radii of curvature of the deformed and undeformed muscle fiber respectively. This is called the relative bending energy. A nice property of the bending energy is that it is invariant to rigid transformation. Therefore it can measure the shape change of the fiber.

3 RESULTS

Internal Tongue Motion: Velocity Fields.

The internal tongue motion seen in the midsagittal tongue during production of "a souk" is shown in Figure 4 at the onset of motion from /s/ to /u/, at the /u/ and the /k/. For NL, the transition from /s/ to /u/ reflects a soft tissue, volume preserving structure. The two ends of the tongue compress inward and the middle of the tongue elevates. The motion from /s/ to /u/ is quite different for PT, who rotates the tongue backward around a central core. This pattern was previously seen (and reported last year) in a patient with primary closure. Her central core corresponded to the scar left by the primary closure. For the present flap patient it is not possible to determine the location of any scars as they are beneath the flap. It does seem, however, as if a scar may have formed at the center of rotation seen here. Another unusual pattern in this image is that, the tip of the tongue is pulled backward and downward, despite the tongue body direction of rotation, which is upward and backward. Although a small depression is also seen in NL, the extent is lesser. Moreover, PT has a noisy /s/, which could result from lateral emission of air through the gap. The motion at /u/ for NL indicates the addition of backward motion consistent with a pull from styloglossus muscle. PT now uses the inward compression and upward expansion seen earlier in NL. At /k/ both subjects are beginning to pull the tongue forward into the inhalation.

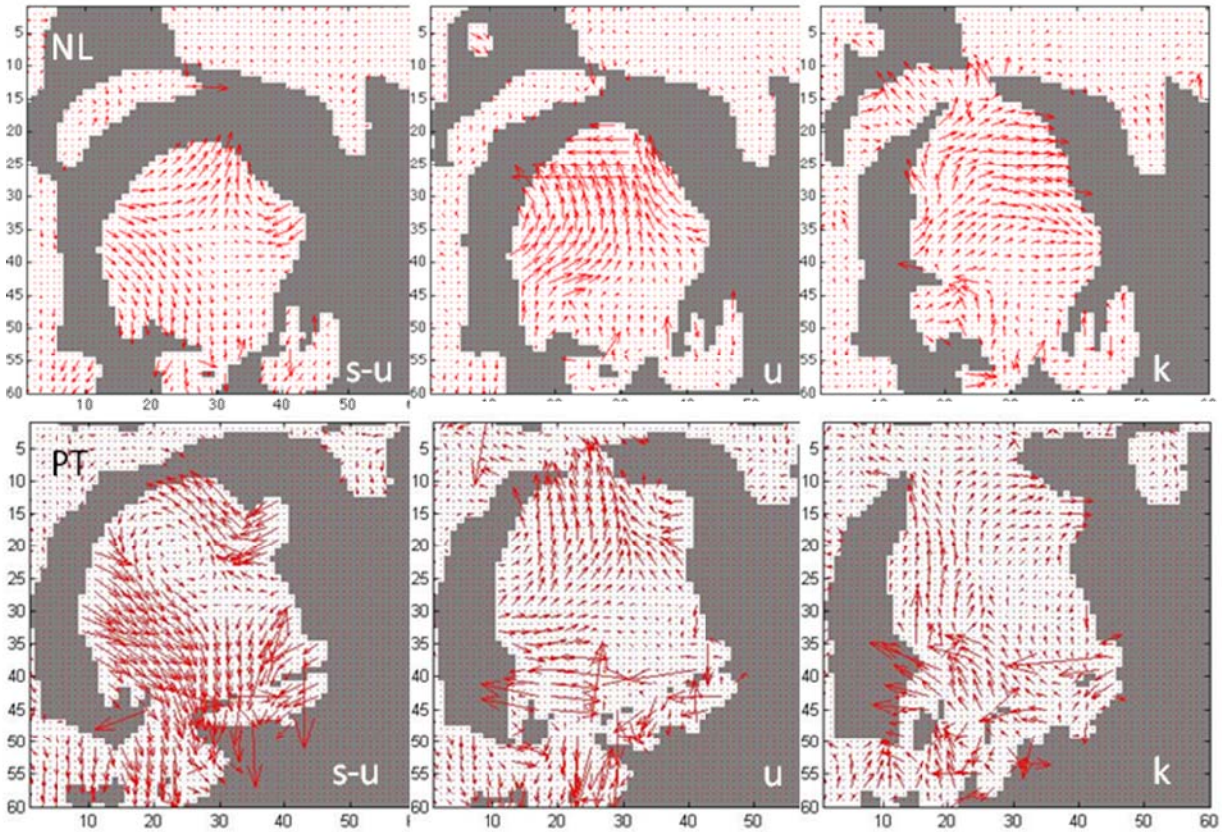

Figure 4. Velocity fields for NL (top) and PT (bottom) for the motion between /s/ and /u/ (a) the /u/ (b) and the /k/ (c) in the word "a souk."

Fiber Position and Elongation

Table 1 presents the mechanical changes in the IL fiber during some time-frames of "a souk". The times at which each speech sound occurred (column 1) were determined from Cine-MRI images of the tongue surface, not the IL fiber lengths. The asterisks refer to the velocity field images in Figure 4. The C1-3 values represent the global motion of the fiber and the elongation represents homogeneous stretch. The general changes in global motion and elongation for the IL fiber were fairly similar for the two subjects. C1 shows that both subjects moved the fiber forward into the /s/. NL then moved it backwards, whereas PT moved it back, then forward into /k/. This is not unlikely however, since the words were followed by an inhalation and the tongue could have anticipated that forward motion. C2 shows that the fiber moved down into /s/ and then upward into /u/ for both subjects. Again PT moved downward at the end, unlike NL. C3 shows backward rotation occurred for both subjects, with considerably less rotation for PT later in the word. Instead PT showed more elongation into the /s/ and more shortening thereafter, eventhough he never shortened to resting length (tf- 1).

Table 1. Mechanical changes in IL fiber for NL (top) and PT (bottom). Negative x-translation is backward and negative y-translation is upward. Negative rotation is backward. Time-frames with * and ** refer to the velocity fields in Figure 4 (b) and (c) respectively.

Time-frame **(NL)**	Speech Sound	C1 (x-translation) (mm)	C2 (y-translation) (mm)	C3 (rotation angle) (degree)	Fiber elongation relative to first phase
1	schwa	0	0	0	0
3	start /s/	2.4 (+2.4)	-0.8	0.5	5.1%
4	max /s/	2.4 (0.0)	-0.9	-1.3	0.5%
8 *		2.4 (0.0)	-0.1	-20.2	1.9%
11 **	/u/	2.7 (+0.3)	-0.1	-20.2	-2.7%
14 ***	/k/ palate	3.3 (+0.6)	-0.3	-18.1	-9.7%
15	/k/ velum	3.3 (0.0)	-0.4	-19.0	-10.3%

Time-frame **(PT)**	Speech Sound	C1 (x-translation) (mm)	C2 (y-translation) (mm)	C3 (rotation angle) (degree)	Fiber elongation relative to first frame
3	schwa	0	0	0	0
10	start /s/	3.4 (+3.4)	-0.7	-5.3	18.9%
14	max /s/	2.8 (-0.6)	-0.2	-9.1	17.3%
16 *		2.4	0.1	-6.2	14.1%
18 **	/u/	1.9 (-0.5)	0.4	-4.0	7.0%
20 ***	/k/ palate	1.7 (+0.1)	0.5	-3.3	3.3%
21	/k/ velum	1.6 (-0.1)	0.4	-1.6	5.0%

Figure 5 depicts the changes in fiber length, position, and height during “a souk” for NL (right) and PT (left). Time is on the y-axis, with time-frame 1 at the bottom. The back of the tongue is on the left. Over time (bottom-to-top) the fibers shorten and lengthen, move anteriorly and posteriorly, and change somewhat in bend (color). Table 1 indicates continuous shortening of IL throughout the word, suggesting activation for both subjects. Both subjects translate the muscle forward and upward during /s/ (NL:1-4; PT:11-14). They then compress the fiber as they move into /u/ (NL-11; PT-19). NL continues compressing into /k/ (15). The changes in AP position and compression are not very different between the two subjects. However, fiber bend is quite different between the subjects. For NL, the compression causes upward arching in the front and center (red) (time-frames 10-18). For PT it causes the most elevation anteriorly (yellow), as can be seen in Figure 5.

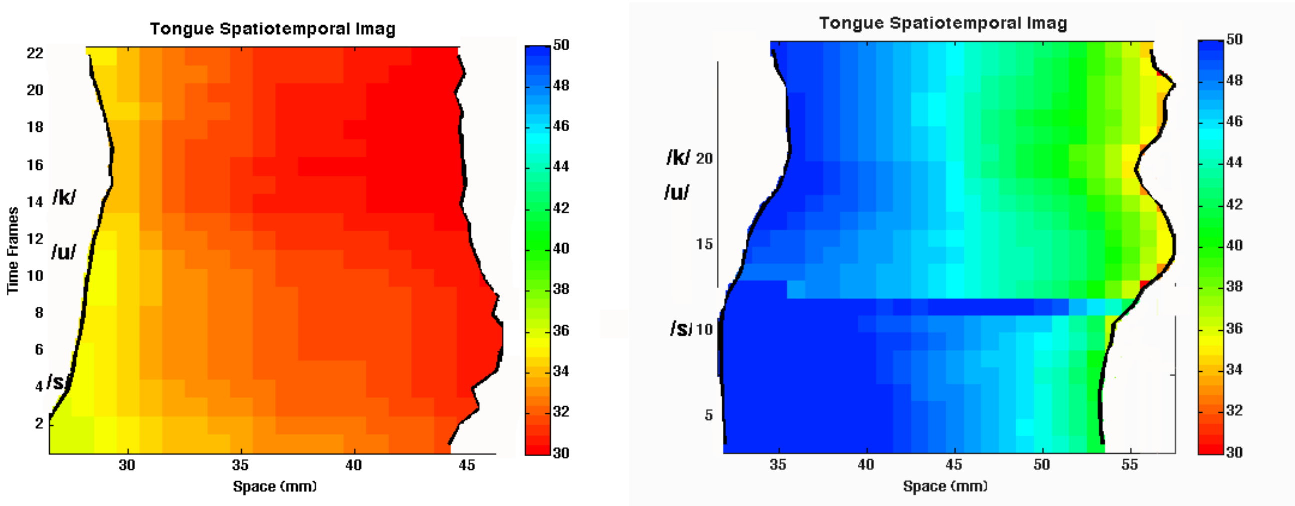

Figure 5. IL fibers displayed over time for NL (left) and PT (right). The frames corresponding to /s/, /u/, and /k/ have been identified at left. Color refers to height with red highest and blue lowest. PT had a lower position overall. Shapes can also be seen in Figure 6.

Fiber Shape and Bending Energy.

Figure 6 displays the general position and shape for the IL fibers at all time-frames overlaid. The big difference between the two subjects is the direction of the bend. NL fiber shape is slightly convex; PT shape is concave. The NL fiber rotates after the /s/; PT rotates during the /s/ (see Table 1). Two badly tracked fibers are seen at time-frames 11 and 12 for PT. They occur just after maximum /s/ is reached.

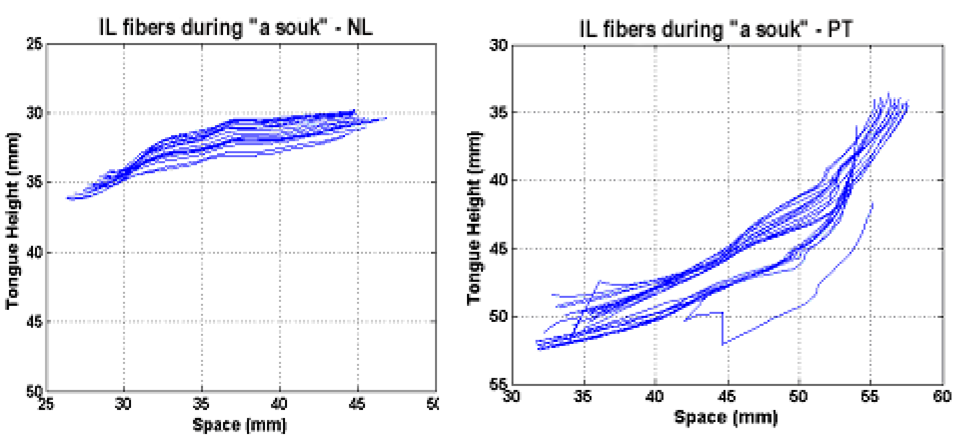

Figure 6. Overlay of all IL positions during the utterance "a souk" for NL (left) and PT (right). NL is flat, convex in shape and rotates backward during motion. PT is steeply sloped, concave and does not rotate. Two time frames are mistracked for PT: #11,12.

Bending energy is shown for NL (left) and PT (right) in Figure 7. The badly tracked time-frames of PT, 11 and 12, have been omitted. The blue line indicates change in bend relative to the first time-frame. The red line indicates change in bend relative to a straight line. For PT the first contour is close to a straight line, therefore the red and blue lines track each other. For NL the initial shape is slightly convex. NL's largest bends occur during /k/ (tf 14), the sound with the highest tongue body, and the subsequent inhalation (tf 17). For PT the largest bends occur during /u/ and /k/, and also /s/. As bend increases, NL becomes more convex, and PT becomes more concave.

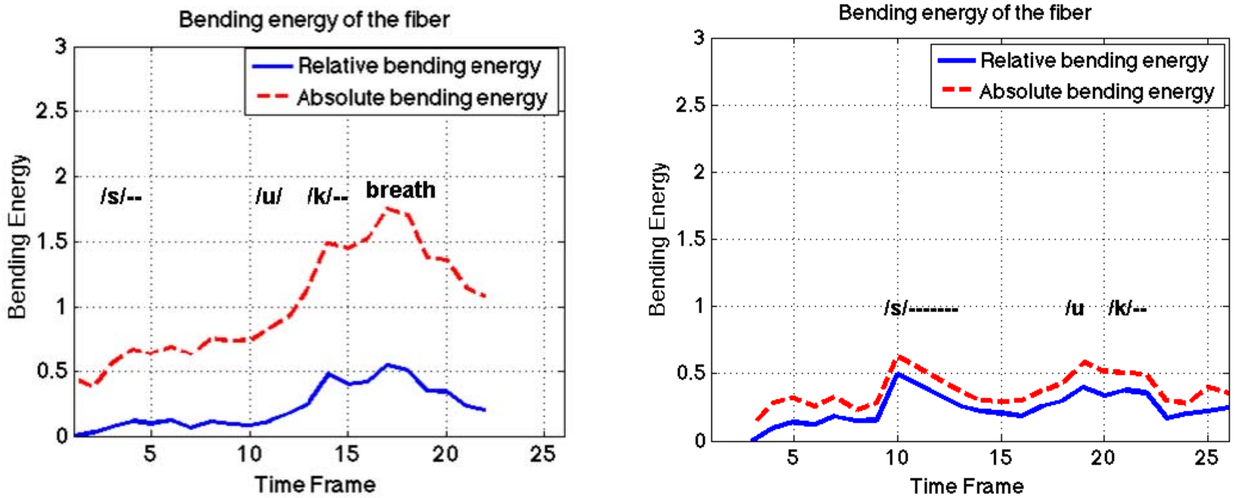

Figure 7. Bending energy for the IL fiber over time for NL (left) and PT (right).

4. DISCUSSION

There are some features of the IL that are quite similar across the two subjects. The first is the timing of the events. They both shorten and lengthen the fiber at the same speech sounds, consistent with tongue body extension and compression, especially for the latter part of the word. Second, the high-to-low orientation of both IL fibers are comparable; they are highest at the tip and lowest in back. Third, large changes in bend occur during the high body sounds /u/ and /k/.

IL's behavior during the /u/ to /k/ motion is very similar for both subjects. The tongue is a volume preserving, soft tissue, structure and motion is closely associated with deformation. The elevation of the tongue body is usually attributed to' pulling' with PG or SG. In these subjects among others, we have also seen anterior-to-posterior compression resulting in upward expansion, which seems to be a normal volume preserving deformation used for this gesture. Both subjects may activate IL for this compression as it would pull the tip and root together, and the muscle does shorten. For NL, the IL length during /k/ is 5% shorter than during /s/, and 10% shorter than during schwa. The /s/ is the IL's most expanded sound. For PT the IL is 13 % shorter during /k/ than it was during /s/, but it is longer than during the schwa by 5% (Table 1). Either other muscles are more active in elevating the tongue, or the relative IL shortening from /s/ is more important than the absolute length in these movements.

The /s/ to /u/ motion is more complicated. PT has a change in bend during /s/, which is not seen in NL. He also expands the fiber three times as much as NL (Table 1). The tongue body rotates backward at the start of the /s/-to-/u/ motion (Figure4), but the IL muscle does not rotate. It seems to be pulling in the tongue tip. However, due to scarring or the flap itself, the inward compression/upward expansion seen in NL is replaced by a rigid backward rotation for a few frames. Partway through the gesture, the normal motion resumes. NL does not rotate the tongue during /s-u/, but his IL muscle does rotate. These differences may have to do with the counter effects of the scar/flap.

The depression behind the tongue tip during /s/ is the most interesting feature of PT's speech, since an audible noise is heard during his /s/. The orientation of IL may have to do with this. The concave shape in NL indicates that IL contraction will pull the tongue tip back. The concave orientation in PT is more consistent with pulling the tip downward. It may be that the orientation is an important feature in the acoustic quality of the /s/. The flap/scar also create rigidities that are not easily overcome, irrespective of muscle orientation. Since a slight depression behind the tongue tip is not abnormal during /s/, it is also possible that normal subjects have better control of the degree to which the tip is pulled down than glossectomies. More subjects will help elucidate this issue. Comparison of the motions with the amount and type of speech therapy received by a patient may also be helpful.

5. Conclusion

The mechanics of the fiber shed some light on the control difficulties faced by glossectomy speakers. The IL fiber orientation was concave for PT and convex for NL. More subjects will need to be studied to determine whether the concavity is due to surgery or a variant shape also seen in normal subjects. The scar also affects the motion during some tongue tip sounds. The /s/ to /u/ motion involved rigid rotation about the scar in a manner that was not observed in tongue body sounds. Changes in fiber length, bend, elongation were fairly similar between NL and PT during the word. Differences that did occur were most typical during the /s/.

6. References

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

Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy andimaging. Biophys J 1994: 66(1):259-267.

Duncan, J.S., Lee, F.A., Smeulders, A.W.M., and Zaret, B.L. “A bending energy model for measurement of cardiac shape deformity,” IEEE Trans. Med. Imag., pp 307-332, vol. 10, no. 3, 1991.

Gilbert RJ, Napadow VJ. Three-dimensional muscular architecture of the human tongue determined in vivo with diffusion tensor magnetic resonance imaging. Dysphagia 2005: 20(1):1-7.

Napadow VJ, Chen Q, Wedeen VJ, Gilbert RJ. Intramural mechanics of the human tongue in association with physiological deformations. J Biomech 1999: 32(1):1-12.

Napadow V.J., Chen Q., Mai V., So P.T., and Gilbert R.J.(2001) Quantitative analysis of three-dimensional-resolved fiber architecture in heterogeneous skeletal muscle tissue using nmr and optical imaging methods. Biophys J. 80(6), 2968-2975.

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

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

Shinagawa H, Murano EZ, Zhuo J, Gullapalli RP, Landman B, Prince JL, Stone M. Tongue muscle fiber tracking during tongue protrusion and rest. J Acoust Soc Am 2006: 120(5 (pt 2)):3354.

Shinagawa , H., Murano , EZ, Zhuo , J., Gullapalli , R., Landman , B., Prince, JL, and Stone, M. (2007) Human Tongue Myoarchitecture with Oral Appliance, Using Diffusion Tensor Imaging. International Assn of Dental Research. New Orleans, LA, March 2007.

Takemoto H. Morphological analyses of the human tongue musculature for three-dimensional modeling. J Speech Lang Hear Res 2001: 44(1):95-107.

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