A comparative acoustic study on speech of glossectomy patients and normal subjects

Xinhui Zhou[1], Maureen Stone[2], Carol Espy-Wilson[1]

[1]Department of Electrical and Computer Engineering, University of Maryland, College Park, USA
[2]Departments of Neural and Pain Sciences and Orthodontics, University of Maryland Dental School, Baltimore, USA
[1]{zxinhui@umd.edu, espy@umd.edu}, [2]mstone@umaryland.edu

Abstract

Oral, head and neck cancer represents 3% of all cancers in the United States and is the 6th most common cancer worldwide. Tongue cancer patients are treated by glossectomy, a surgical procedure to remove the cancerous tumor. As a result, the tongue properties such as volume, shape, muscle structure, and motility are affected. As a result, the vocal tract acoustics are affected too. This study compares the speech acoustics between normal subjects and partial glossectomy patients with T1 or T2 tumors. The acoustic signal of four vowels (/iy/, /uw/, /eh/, and /ah/) and two fricatives (/s/ and /sh/) were analyzed. Our results show that, while the average formants (F1-F3) for the four vowels between the normal subjects and the glossectomy patients are very similar, the average centers of gravity for the two fricatives differ significantly. These differences in fricatives can be explained by the more posterior constriction in patients due to the glossectomy (or the cancer tumor) and its resulting longer front cavity.

Index Terms: glossectomy, speech production, acoustic phonetics, formants, center of gravity, and skewness.

1. Introduction

Oral, head and neck cancer represents 3% of all cancers in the United States and is the 6th most common cancer worldwide [1]. Recent studies have shown a five-fold increase of cancer incidence on the oral portion of the tongue among young men and a six-fold increase among young women [2]. Tongue cancer patients are usually treated by glossectomy, a surgical procedure to remove the cancerous tumor plus about 1 cm of tissue around it. After glossectomy, the tongue will be sutured closed or a flap will be added to reconstruct the tongue volume. As a result, the properties of tongue such as volume, shape, and muscle mechanics are more or less affected by the surgery. As an example, Figure 1 shows an asymmetrical tongue with a hole on its left side for a glossectomy patient. Due to the change in the tongue properties, the patients' critical function of speech might be impaired.

There are many studies that have assessed the speech quality after glossectomy [3][4][5][6]. It has been found that the tumor size and location, the type of reconstruction, and the affected muscles have roles in affecting the speech intelligibility. But the main determinants of good speech after glossectomy are not well established. In general, it is important to maintain the tongue motility for good speech, and the intelligibility of consonants is deteriorated more than the vowels. Instead of intelligibility, Savariaux et. al [7] did a longitudinal study of the speech acoustics of a group of patients in the pre-surgery and post-surgery conditions. However, overall, the relationship between the speech acoustics and the glossectomy procedure was not well explored.

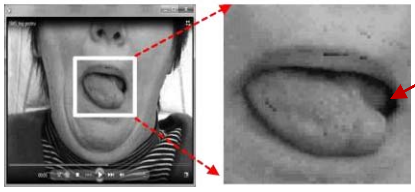

Figure 1: *A glossectomy patient with tongue protruded (resection location is on her left side of the face)*

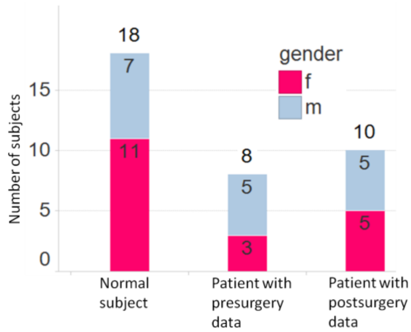

Figure 2: *Subject number information in the database*

The main objective of this paper is to assess the acoustic differences in speech production between the normal subjects and the glossectomy patients. Such a study will help us with our long-term goal, which is to model the vocal tract acoustics of the glossectomy patients and to provide clinical guidance to the surgeon for a better speech outcome. This effort is part of a larger ongoing project which aims at understanding the tongue muscle mechanics, tongue motion pattern, and the vocal tract acoustics of glossectomy patients as well.

In the rest of this paper, we describe our database and also methodologies for measuring vowel format patterns and fricative spectral properties. In addition, we present our measurement results along with data from classical literature [8][9][10]. Some explanations in terms of articulation and vocal tract modeling are provided for our observations. Finally, a summary along with our plans for future work are given.

2. Database and Methodologies

2.1. Database

Our database consists of 18 normal control subjects and 16 glossectomy patients. All are native American English adult speakers. The patients are in either pre-surgery or post-surgery condition and their ages range from 29 to 61 years with a

28 – 31 August 2011, Florence, Italy

mean of 44. The tumors are sized T1 (largest dimension <=2 cm) or T2 (2-4 cm) and located in the lateral part of the tongue. Figure 2 shows the three groups (normal subjects, patients for whom we had pre-surgery data, and patients for whom we had post-surgery data). So far, there are only two patients (one male and one female) from whom we have both pre-surgery and post-surgery speech data.

In addition to MRI scanning and other acoustic data, the subjects were instructed to pronounce a 16 VCV nonsense words using 4 vowels (/iy/, /ah/, /eh/, /uw/) x 4 consonants (/s/, /sh/, /l/, /g/). These combinations maximize the vowel space and provide a large assortment of consonant positions and manners of production. Each VCV was repeated at least three times. The audio data was acquired by a miniature digital recorder (Olympus 300M) and the signal was downsampled at 22 kHz. We did acoustic analyses of the vowels and the fricatives /s/ and /sh/.

2.2. Acoustic measurement

We measured the formants F1-F3 of the vowels using the WaveSurfer formant tracker with manual correction of formant trajectory. The LPC order was 12 and the analysis window size was 50 ms. For each vowel realization, the formants in the middle frame were measured.

The center of gravity and skewness were used to describe the spectral properties of the fricatives as in [10]. The center of gravity is defined as in Equation (1). It is the mean frequency of the power spectral density (PSD) *S(f)* and it is strongly correlated to the spectral peak location. The skewness is defined as in Equation (2), which indicates the asymmetry of the PSD. We used Welch's method to estimate the PSD.

$$f_0 = \int_0^{\infty} f|S(f)|^2 df \Big/ \int_0^{\infty} |S(f)|^2 df \qquad (1)$$

$$skewness = \int_0^{\infty} (f - f_0)^3 |S(f)|^2 df \Big/ \left(\int_0^{\infty} |S(f)|^2 df\right)^{1.5} \qquad (2)$$

3. Results

3.1. Formants of vowels (/iy/, /eh/, /ah/, and /uw/)

Figures 3 and 4 show the F1-F2 and F1-F3 plots of the four vowels (/iy/, /ah/, /eh/, and /uw/) for the three subject groups, respectively. For clarity, only the means of F1, F2 and F3 are plotted. The classic vowel formant data from Peterson, et al. (1952) [8] and Hillenbrand et al. (1995) [9] are also included in the plot. It can be seen that our data of the four vowels are well separated in the F1-F2 plot, and they form the regular vowel quadrilaterals as the classic data in the literature. The F3 patterns of our data are also consistent to the classic data in literature, with a higher F3 for /iy/ and a lower F3 for /uw/ in each subject group.

It can be also seen that there are some differences in the average formant values among the three subject groups. However, no consistent formant difference pattern across gender and/or across phonemes has been found in our data, except the average F1 from the patients with pre-surgery data. In this case, the average F1 is always smaller than the averages from the other two groups. However, the F1 differences are within 50 Hz. Moreover, among all the cases, the largest average F1 difference (in female /eh/) among the three groups is about 80 Hz. Except a 520 Hz large difference in F2 (in /uw/ for the females) between the post-surgery data and the pre-surgery data, the largest F2 difference (in /iy/ for the females) is about 240 Hz. The largest F3 difference (in /iy/ for the females) among the three groups is less than 400 Hz. So, on

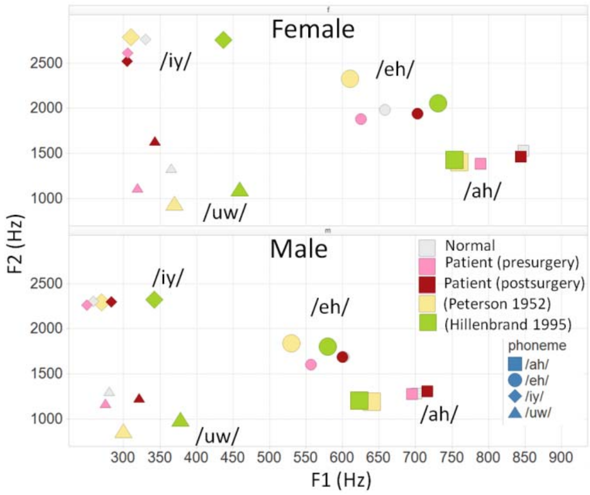

Figure 3: *F1-F2 plots for vowels /iy/, /eh/, /ah/, and /uw/ (upper: female, bottom: male, different colors stand for different subject groups or studies in literature)*

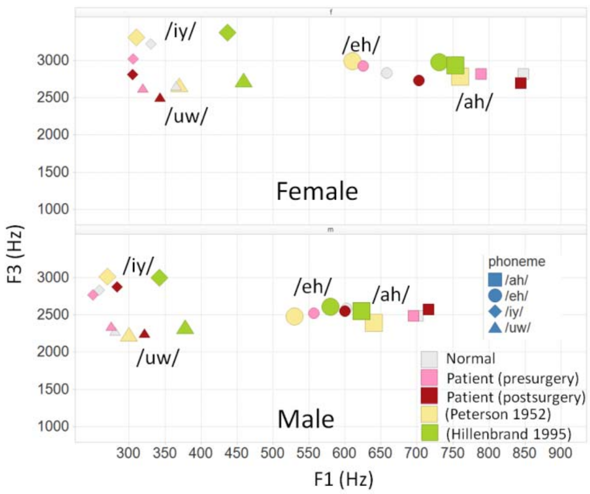

Figure 4: *F1-F3 plots for vowels /iy/, /eh/, /ah/, and /uw/ (upper: female, bottom: male, different colors stand for different subject groups or studies in literature)*

average, the formant patterns of F1, F2 and F3 among the three subject groups are very similar to each other. And there is no evidence of a consistent formant difference pattern between the normal subjects and the patients in our database.

3.2. Center of gravity and skewness of the fricatives (/s/ and /sh/)

Figure 5 shows the scatter plots of the center of gravity and skewness for /s/ and /sh/, respectively. It can be seen that, on average, /s/ has a higher center of gravity and a smaller kewness than /sh/. Also the female subjects tend to have a larger center of gravity than the male subjects. It can be seen from the scatter plots that the data for the pre-surgery patients and the data for the post-surgery patients are always more overlapped with each other than with the data for the normal subjects. This result implies some consistent difference in /s/ and /sh/ between the normal subjects and the patients.

Figure 6 shows box plots [11] of the center of gravity for /s/ and /sh/, respectively. The whisker length w is 1 and the data were drawn as outliers if they are larger than q3+w(q3–

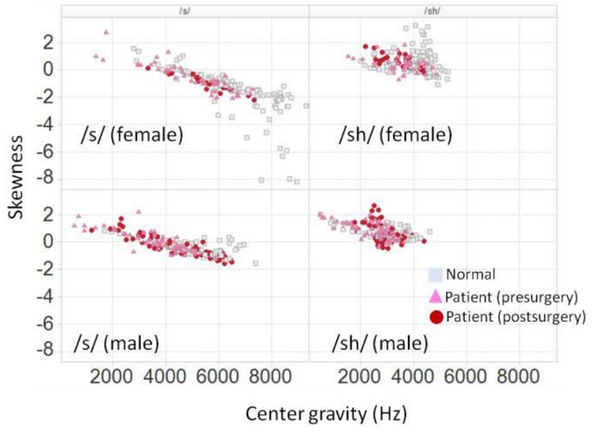

Figure 5: *Scatter plots of center of gravity and skewness for fricatives /s/ (left panels) and /sh/ (right panels). (upper: female, bottom: male, different colors/shapes stand for different subject groups)*

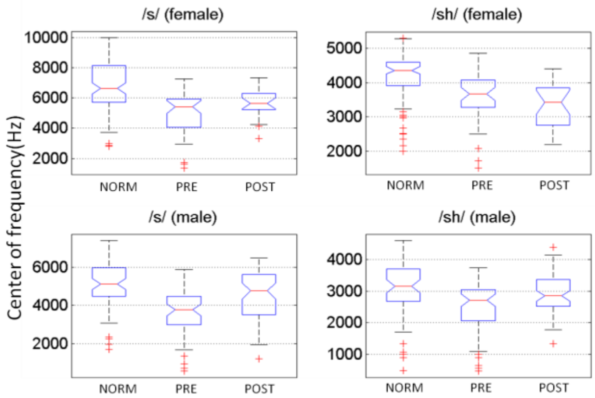

Figure 6: *Box plots of center of gravity for /s/ (left panels) and /sh/ (right) produced by normal subjects (NORM), patients with pre-surgery data (PRE) and patients with post-surgery data (POST)*

q1) or smaller than q1– w(q3 –q1), where q1 and q3 are the 25th and 75th percentiles, respectively. The notch size in the box plot represents the 95% confidence interval of the median. The two medians are significantly different at the 5% significance level if their confidence intervals do not overlap. It can be seen in Figure 6 that the medians of the center of gravity for the normal subjects are consistently higher than the medians of the patients' data. The notches for the normal subject data are not overlapped with the notches for the patient subject data, except in the case of the female /s/ in the post-surgery data. This result means that the median differences between them are significant.

There are also some differences in the center of gravity between the pre-surgery data and the post-surgery data. However, the confidence intervals of both are overlapped for most of the cases, which means the significant levels for those differences are smaller than 5%.

Figure 7 shows the skewness box plots of /s/ and /sh/. The center of gravity and the skewness are somewhat correlated. Skewness tends to be more negative when the center of gravity becomes higher. For the female subjects in our data, the order of the skewness medians among the three subject groups is reverse to the order for the center of gravity medians. However, for the male subjects, the median skewness in the normal subjects is not the smallest among the three subject groups. The notches between the normal subjects and the

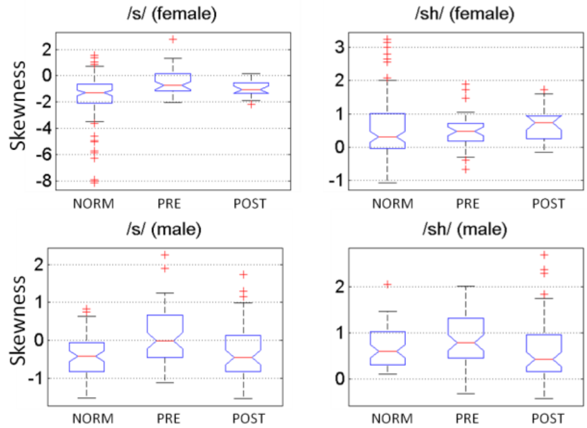

Figure 7: *The box plots of skewness for /s/ (left side) and /sh/ (right side) produced by normal subjects (NORM), patients with pre-surgery data (PRE) and patients with post-surgery data (POST)*

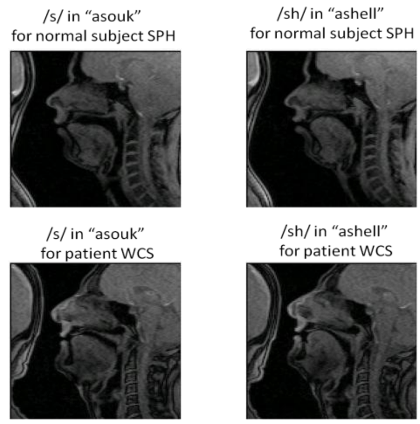

Figure 8: *The midsagittal MR images of /s/ (left) and /sh/ (right) production by the normal subject SPH (upper) and the patient WCS (bottom) after surgery*

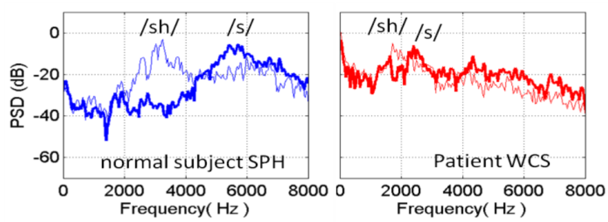

Figure 9: *Spectra of /s/ and /sh/ produced by the normal subject SPH (left) and the patient WCS (right)*

patients are more overlapped than in the case of the center of gravity. So, the median skewness differences between the data for the normal subjects and the data for the patients are not as consistent and significant as it is for the center of gravity.

4. Discussion

4.1. Vowels

Our results for the four vowels (/iy/, /uw/, /eh/, and /ah/) show that the format patterns between the normal subjects and the glossectomy patients in our database are similar, and there is no evidence that there is any consistent formant difference pattern between them. There are several factors contributing to this result. First of all, all of our patients had T1 or T2 tumors which are not as large as T3, T4 and T5 tumors. Only a partial glossectomy is needed for these patients, so the perturbation to the vocal tract shape is relatively small. Second, even though the tongue shape after glossectomy is changed, the relative change in terms of the vocal tract area function in the front cavity may not change much, at least in the cases of /eh/, /ah/, and /uw/, where the area function in the front cavity is pretty large in general. Furthermore, through auditory feedback, the subjects can adapt their articulation strategies to compensate for the vocal tract shape perturbation [12]. However, patients with T3-T5 should have more difficulty in vowel production than the patients we studied.

4.2. Fricatives

The production of /s/ and /sh/ involves a narrow anterior constriction in the vocal tract which generates a turbulence noise source. The resulting spectral peak is located at the first resonant frequency of the front cavity [13]. The constriction for /s/ is usually formed more anterior to the constriction for /sh/. Correspondingly, the front cavity length for /s/ is shorter than for /sh/. Therefore, the frequency of the spectral peak in /s/ is higher than it is in /sh/. Intuitively a spectral peak at a higher frequency leads to a higher center of gravity. Our results show that the average centers of gravity for /s/ and /sh/ for normal subjects is consistently higher than it is for the patient data. This difference might be explained by the constriction differences in fricative production between normal subjects and patients.

Figure 8 shows the midsagittal cine-MR (magnetic resonance) images of /s/ and /sh/ produced by one normal subject (SPH) and one patient (WCS, after surgery) from our database. It can be seen that, for subject SPH, the constriction for /s/ is formed at the teeth, whereas the constriction is formed at the alveolar for patient WCS. So, the constriction is more posterior in /s/ of WCS. For /sh/, the constriction also looks more backward for WCS than for SPH. The more backward constriction makes the front cavity longer and lowers its resonance frequency. Figure 9 shows the corresponding spectra of /s/ and /sh/ for these two subjects. It can be seen that the frequencies of the spectral peaks of /s/ and /sh/ for SPH are much higher than they are for WCS. For WCS, the frequencies of the spectral peaks for /s/ and /sh/ are closer and this result is consistent with the similarity of the corresponding tongue shapes in the MR images. In fact, the frequency of the spectral peak for /s/ produced by WCS is in the same range of /sh/ produced by SPH.

There may be two reasons why the glossectomy patients have a more backward constriction (or a longer front cavity) than the normal subjects in the production of /s/ and /sh. First, the surgery or even only the existence of the tumor may change the motility of the tongue, which makes it more difficult to form a constriction by using the tongue tip as compared to a constriction using the tongue blade. The midsagittal cine-MR images in our database show that none of our glossectomy patients used the apical tongue shape SPH used for /s/ (shown in Figure 8). Second, the resection of the tongue may effectively shorten the tongue and the constriction will be more backward when the shortened tongue is raised to form the constriction. However, a 3-D vocal tract shape analysis is needed for further proof.

It has been believed, as in [7], that the speech data before surgery can be regarded as a reference of normal speech. However, our study shows that there are significant differences in the fricatives /s/ and /sh/ between normal subjects and patients before surgery. One possible explanation for these differences is that the pain or discomfort caused by the cancer tumor may affect fricative production.

5. Summary and future work

This study compares the speech acoustics between normal subjects and partial glossectomy patients with T1 or T2 tumors. The acoustic signal of four vowels (/iy/, /uw/, /eh/, and /ah/) and two fricatives (/s/ and /sh/) were analyzed. Our results show that, while the average formants (F1- F3) for the four vowels between the normal subjects and the glossectomy patients are very similar, the average centers of gravity of the two fricatives differ significantly. These differences in fricatives can be explained by the more posterior constrictions in patients due to the glossectomy (or the cancer tumor) and its resulting longer front cavity.

Data collection for many more subjects is ongoing. Our future work will include an ANOVA analysis of the acoustic measurements for a more rigorous statistical analysis, a detailed acoustic and articulatory analysis for each subject to account for speaker-specific differences, and also vocal tract modeling to interpret the acoustics for the glossectomy patients.

6. Acknowledgements

This work is supported by NIH grant R01CA133015.

7. References

[1] American Cancer Society, facts and figures, available at http://www.cancer.org/Research/CancerFactsFigures/CancerFactsFigures/cancer-facts-and-figures-2010.

[2] K. Annertz, H. Anderson, A. Biorklund A, T. Moller, S. Kantola, J. Mork, J. H. Olsen, and J. Wennerber, "Incidence and survival of squamous cell carcinoma of the tongue in Scandinavia, with special reference to young adults", International Journal of Cancer, vol. 101, pp95-99, 2002.

[3] S. G. Fletcher, "Speech Production Following Partial Glossectomy", Journal of Speech and Hearing Disorders, vol. 53, pp. 232-238, 1988.

[4] C.L. Furia, L.P. Kowalski, M.R. Latorre, E.C. Angelis, N.M. Martins, A.P. Barros and K.C. Ribeiro, "Speech intelligibility after glossectomy and speech rehabilitation", *Arch. Otolaryngol. Head Neck Surg., v*ol. 127, pp. 877-883, 2001.

[5] T. Bressmann, R. Sader, T.L. Whitehill and N. Samman, "Consonant intelligibility and tongue motility in patients with partial glossectomy", International Journal of Oral and Maxillofacial Surgery, vol. 62, pp. 298-303, 2004.

[6] Y. Matsui, T. Shirota, Y. Yamashita and K. Ohno, "Analyses of speech intelligibility in patients after glossectomy and reconstruction with fasciocutaneous/myocutaneous flaps", Journal of Oral and Maxillofacial Surgery, vol. 38, no. 4, 339-345, 2009.

[7] C. Savariaux, P. Perrier, D. Pape, J. Lebeau, "Speech production after glossectomy and reconstructive lingual surgery: a longitudinal study", Proceedings of the 2nd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), Firenze, Italy.2001.

[8] G. Peterson and H. L. Barney, "Control methods used in a study of the vowels", Journal of the Acoustical Society of America, vol. 24, no. 2, 175-184, 1952.

[9] J. Hillenbrand, L. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels", J. of the Acoustical Society of America, vol. 97, no. 5, pp. 3099-3111, 1995.

[10] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of English fricatives", Journal of the Acoustical Society of America, vol. 108, no. 3, pp. 1252-1263, 2000.

[11] M. Robert, T. W. John, L. A. Wayne, "Variations of box plots", The American Statistician, Vol. 32, no. 1, pp12-16, 1978.

[12] J. A. Jones and K. G. Munhall, "Learning to produce speech with an altered vocal tract: The role of auditory feedback", Journal of the Acoustical Society of America, vol. 113, no. 1, pp. 532-543.

[13] K. N. Stevens, *Acoustic phonetics*. Cambridge, Mass.: MIT Press, 1998.