ABSTRACT

Title of Thesis: A SPECTRAL ANALYSIS OF SPEECH
PRE- AND POST- GLOSSECTOMY

Emma Griffith
Master of Science, 2025

Thesis Directed by: Professor Carol Espy-Wilson
Department of Electrical & Computer Engineering
Institute for Systems Research

This study aims to use spectral analysis to explain the changes in speech caused by a glossectomy. A single patient is observed using video recordings across 30 years at four stages: before tongue cancer, after a partial glossectomy with a radial forearm free flap, after additional surgeries including removal of the flap, and without the flap in the cold, a known area of speech difficulty for the patient. Formant frequencies and consonant spectra were analyzed to quantify the changes in speech production. Results show the greatest changes in formants occur with front vowels, indicating difficulty making a constriction with the tongue tip; however, overall change in vowel formants is still minimal. Significant spectral differences were observed in the production of the sibilant fricatives /s/ and /ʃ/, with lower spectral peaks and reduced spectral distinctiveness between the sibilant fricatives across all post-glossectomy stages, but most prominently in the cold. Other consonants are less affected, indicating the disproportionate impact a glossectomy has on speech that requires finer control of the tongue-tip. These results underscore the value of

surgical techniques that preserve tongue tip mobility where possible and have further implications in post-glossectomy targeted speech therapy.

A SPECTRAL ANALYSIS OF SPEECH PRE- AND POST- GLOSSECTOMY

by

Emma Griffith

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master's of Science
2025

Advisory Committee:
Professor Carol Espy-Wilson, Chair/Advisor
Professor Shihab Shamma
Professor Jonathan Simon

Dedication

To my dad, Dale Griffith, for inspiring this paper, motivating me to become an electrical engineer, being my role model, and always answering my questions.

To my mom, Edwige Griffith, for always supporting me. I promised I'd get this done!

Acknowledgments

First and foremost I'd like to thank my advisor, Professor Carol Espy-Wilson for giving me the opportunity to work on a project that means so much to me. Her guidance throughout the last year has been invaluable

I would also like to thank Maureen Stone, who graciously provided not only much of the data that this paper was built on, but also her time and expertise in discussing the complexities of tongue motion, anatomy, and post-surgical speech adaptation.

Christine Shadle answered my questions on fricatives and velocity fields. In turn, she asked me questions about my research that made my paper better.

ENEE101 has been a cornerstone of my time at Maryland. I would like to acknowledge the 101 team, including Prof. Julius Goldhar, Prof. Mel Gomez, Bryan Quinn and Tech Ops team for their support over the last four years. Being a part of ENEE101 has made me a better engineer, communicator, and leader and has instilled a passion for engineering education in me. The financial support from ENEE101 allowed me to complete this degree.

I would not be here without my family. I am grateful for their unwavering support.

Table of Contents

List of Figures

List of Abbreviations

LPC	Linear Predictive Coding
RFFF	Radial Forearm Free Flap
VoT	Voice Onset Time
VCV	Vowel-Consonant-Vowel

Chapter 1: Introduction

The tongue plays a pivotal role in speech production. As such, a glossectomy, the removal of all or part of the tongue has inherent repercussions in speech.

In this thesis, we investigate the case study of a single male native English speaker glossectomy patient over the course of 30 years. The patient was diagnosed with a left lateral squamous cell carcinoma (SCC) T1N0M0 tumor of the lateral oral tongue at age 42, indicating a tumor size of ≤ 2 cm with a depth of ≤ 5 mm, no lymph node involvement, and no metastasis. The tongue tip was preserved, and the tongue was reconstructed using a radial forearm freeflap. An image of the tongue with the flap is shown in Figure 1.2a. After 16 years, the cancer came back on the flap and the flap had to be removed in its entirety, in addition to the left side of the palatoglossal arch.

1.1 Background on Glossectomy

A glossectomy is used as a treatment for tongue cancer, where the tumor, along with any necessary margins and lymph nodes, are surgically removed from the oral cavity. Oral cancer tumors are staged according to the TNM system, where T represents tumor size, N represents cervical lymph node involvement, and M represents metastasis. [2] The glossectomy can be closed with a simple suture or a radial forearm freeflap (RFFF, or simply flap). A flap entails the removal of approximately 9 cm of the forearm skin and the radial artery to close the hole left

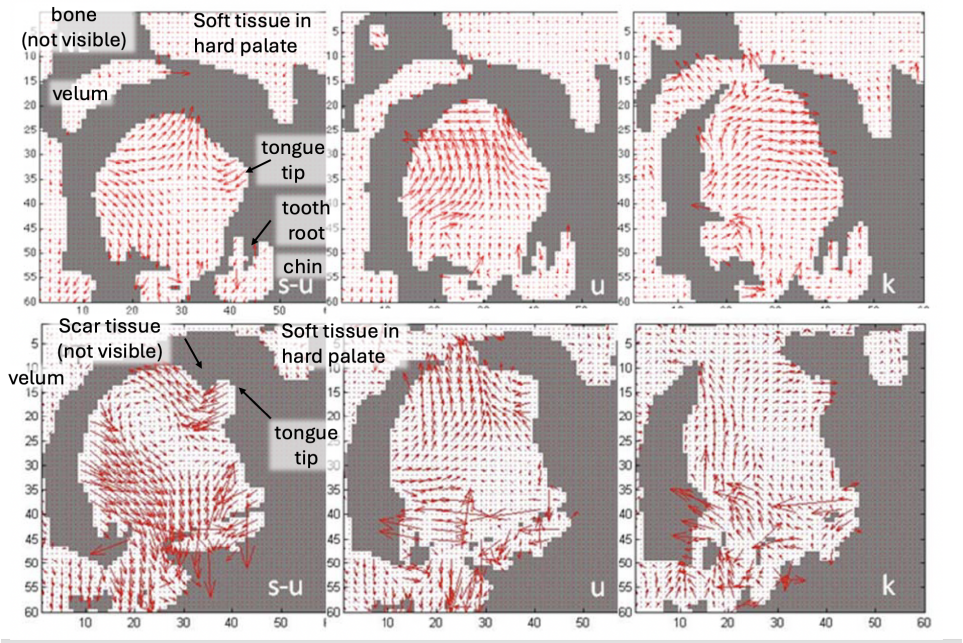


Figure 1.1: Velocity fields of tongue motions for control (top) and patient (bottom) for the utterance 'asouk' at the s-u transition (left), during the u, and at the first full frame of k. Image from [1]

by the glossectomy. [3]

1.2 Understanding Compensation

The subject of this study was previously the subject of a paper by Stone et al. in which they used tagged cine-MRI to understand the tongue movement after a glossectomy [1]. Despite missing significant portions of the tongue missing, glossectomy patients are still able to produce intelligible speech. To do this, they must compensate for not only the missing tongue, but the rigid scar tissue from the surgery. To understand the compensation, we study Figure 1.1 from the original paper, which plots the tongue's velocity field at a midsaggital view during three frames of the utterance of 'asouk' for the patient (bottom) and a control (top).

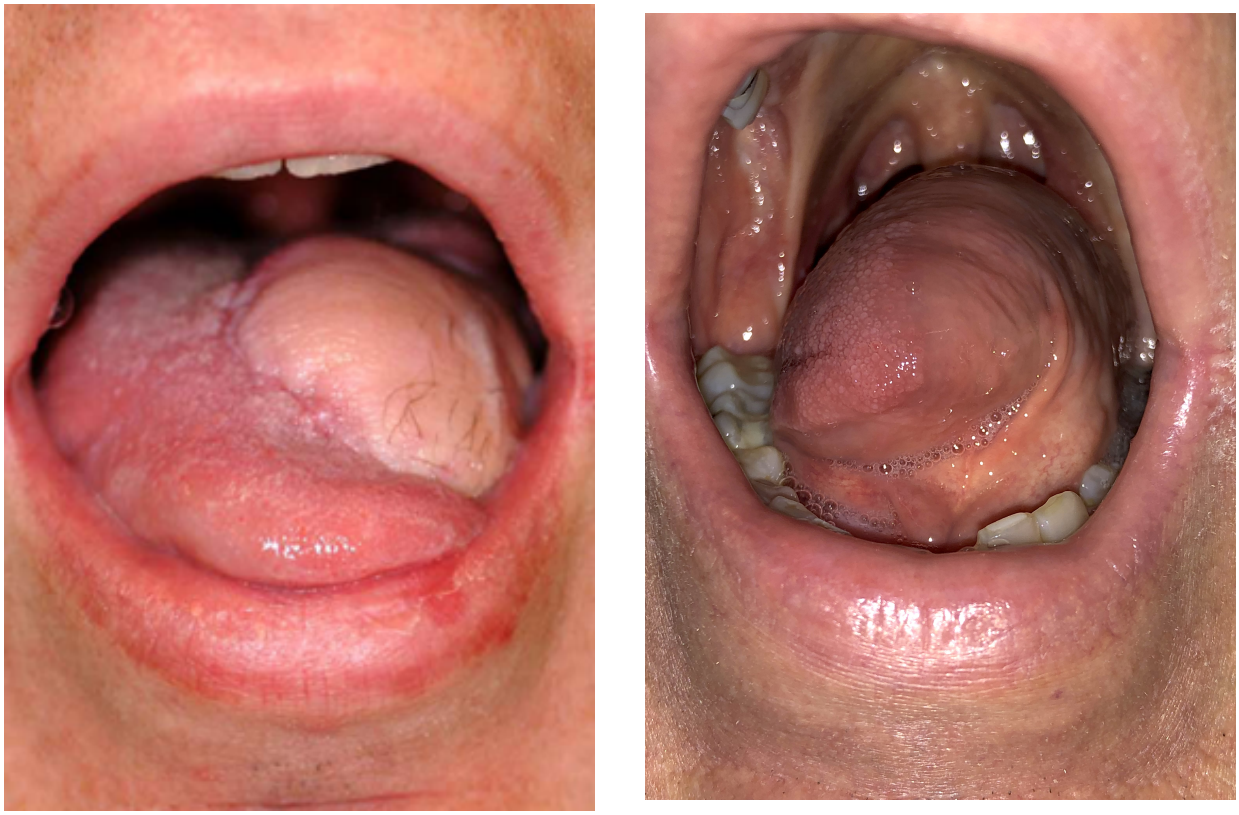
In the first panel, we see the the constriction for the /s/ being released at the front of the mouth, with the velocity arrows pointing down and back as the tongue tip retracts. The back of the tongue moves forward and up, causing the tongue to converge in the middle. The tongue body moves down and in, compressing the tongue upwards. This upwards motion increases in the second panel for the close-back place of articulation for the /u/. Additionally, there is a backwards pull from the styloglossus muscle. In the final panel, the tongue lightly taps the velum to make the constriction for /k/, a velar plosive. The velocity field shows the tongue moving forwards for inhalation. [1]

In comparison, the patient's tongue is constantly in rotational motion to compensate for the rigidity caused by the glossectomy and flap scars. In the first panel for the patient at the s-u transition, the tongue tip is pulling in like in the control, but it needs much more force to move around the rigid scar tissue, which is shown as a divot in the top of the tongue because the tissue is not visible in MRI. The base of the tongue also compensates by pulling inward and squeezing to elevate the tongue to the proper place of articulation. All of this contributes to a rotational motion around the rigid body of the scar. In the second panel, we continue to see compression and expansion at the base of the tongue to elevate it to the proper place of articulation. There is an additional piece of the tongue sticking up past the rest, which is abnormal – perhaps an artifact of the uneven tongue surface caused by the flap and scar tissue. When the tongue first makes the constriction for /k/ shown in the final panel, there is not the light tap seen in the control, but rather the tongue is smashed against the velum, indicating loss of control. However, the more forceful constriction likely does not affect the patient's speech audibly, otherwise they would compensate for it. [1]

As we can see, speaking with a glossectomy requires complex contortions of the tongue to

maintain intelligibility. We can infer that having a greater portion of the tongue missing would require at least as much compensation, if not more. It is from this perspective that we begin this study, which seeks to build upon the insights from Stone et al. and explore the tongue compensation from the the spectral perspective

1.3 Methods



(a) Patient's tongue shortly after glossectomy with radial forearm freeflap

(b) Patient's tongue after radial forearm freeflap has been removed

Figure 1.2: Images of patient's tongue with and without flap.

One of the challenges in evaluating the speech of glossectomy patients is having a good baseline to compare it with. Past studies have used other people without glossectomies as controls [1] [4] or evaluated patients immediately before and after the glossectomy [5] [6]. However, these

neither account for the individual variances within speech that may cause a patient's speech to be dissimilar from the control outside of the glossectomy, nor the fact that immediately before the surgery, the patient still has cancer, and the tumor could be altering the patient's speech. There is no way to predict someone will have oral cancer in the future, so unless baseline recordings are take of everyone, it is not possible to have an individualized baseline. Instead, this study relies on home videos of the subject taken years prior to the cancer diagnosis to provide an accurate baseline for the patient. The rest of the study is subdivided into three stages:

- *Flap*: after the initial partial glossectomy with RFFF. This includes additional biopsies and surgeries that removed small portions of the tongue to prevent further cancer growth, 2007-2018
- *Post*: after the RFFF had been removed due to additional cancerous growth on the flap, 2019-present
- *Cold*: without the RFFF, with speech occuring after the patient had been outside in freezing temperatures for a sustained period of time.

Video recordings were also used to evaluate speech in these stages. A mix of natural speech and clearly articulated speech was recorded. For clear speech, the patient repeated a variety of sentences, as well as vowel-consonant-vowel (VCV) "nonsense" words that particularly targeted difficult phonemes for glossectomy speakers. For each recording, the video was removed and the audio was converted to a .wav file. The sampling frequency was resampled to 16kHz. The audio was analyzed in WaveSurfer, and a spectral slice was taken at each relevant phoneme. For consonants, the FFT was taken using a 512-point Hamming window. For vowels, the frequency of the first two formants were recorded using LPC, also with a 512-point Hamming window,

with and order of 20-35. Although an order of 20 is appropriate when evaluating speech with a sampling frequency of 16kHz, for certain higher frequency formants, it was necessary to increase the order.

1.3.1 Database

A large volume of recordings were analyzed for this project. However, the nature of using home videos means that it is limited by what each family chooses to capture, the environment in which they record in, and the consumer-grade technology available at the time period, unlike recordings made for research purposes where the environment, technology, and speech can be controlled. For this family, many of the recordings happened during large family gatherings where many conversations were happening at once, a challenge in speech processing referred to as the cocktail party problem. At other times, there was too much environmental noise to properly hear the speaker or the recordings, the first of which were from 1995, were just not high enough quality to make good measurements. Thus, a large portion of the data had to be discarded. Furthermore, the subject is known for being a man of few words, so one of the challenges of this project was scouring through the remaining data to find speech samples to ensure each phoneme was adequately represented.

There were not many home videos taken in the period immediately following the glossectomy. Much of the analysis for this period relies on videos from the original study provided by Maureen Stone. They were supplemented by recordings from a class taught by the subject. The class recordings were created nearly a decade after the glossectomy, during which there had been additional tongue surgeries, but because a majority of the flap remained intact, it was determined

that it was still an adequate representation of that speech stage. The class was recorded ”off the cuff” without a script, so it is representative of a conversational style. A further exploration of the class recordings can be found in Section 3.3.

The flap was removed five years ago and there have no additional surgeries since, so the ”post” stage is all present day data. To represent the mix natural conversational and clearly articulated speech in the existing data, the subject was taped during a normal conversation, reading given sentences, and saying VCV nonsense words. Most of the sentences and VCVs were taken from the pre or flap recordings to help control the context of the speech.

The cold recordings are also from present day. The subject was asked to repeat the same phrases as in post, and a small conversation was recorded. The subject was also recorded reciting a Shakespeare soliloquy. For all these recordings, the subject had been outside in freezing temperatures for at least half and hour and had identified that his speech was degrading.

1.3.2 Studying the Effects of Cold Weather

To the best of my knowledge, there is no existing literature on how the cold affects glossectomy patients. However, the subject’s speech intelligibility is known to significantly deteriorate in the cold. As such, it was determined to be a worthwhile line of exploration for this thesis.

1.4 Outline of Thesis

In Chap. 2, we discuss the effect of the glossectomy and additional speech stages on consonants. We focus the discussion on the sibilant consonants /s/ and /ʃ/, which are known to be difficult for glossectomy patients because they require control of the tongue tip. We also

analyze /t/, which shares the alveolar place of articulation with /s/, and /k/, which has a velar place of articulation and thus would be more impacted by the flap than the tongue tip. We look at the center of gravity of each of these consonants across each stage of speech to understand how the overall frequency placement changes.

In Chap. 3, we investigate the effect of the glossectomy on the vowels. We start off by analyzing the average formants for each vowel and the overall vowel space that they produce for each speech stage. We then narrow in on the open vowels, which have been noted to be different for glossectomy patients in other studies. [7] We then analyze the extent to which clear articulation affects the vowel formants

Chapter 4 provides the conclusion to the thesis.

2.5 Overview

Chapter 2: Consonants

2.1 Sibilant fricatives

The sibilant fricatives, /s/ and /ʃ/ have been noted to be particularly difficult for glossectomy patients. [1] [4] [8] Sibilant fricatives have an alveolar and palatal place of articulation, respectively, which require precise and stable control of the tongue tip. After surgery, glossectomy patients are left with a significant amount of scar tissue on their tongue, which is more fibrous, and thus rigid, than normal tissue. The tongue is then less flexible in that area and must compensate with other motions to maintain proper articulation. This compensation is depicted in Figure 1.1, which compares the motion of the tongue of a subject (top) to that of the patient (bottom) as they say 'asouk'. The /s/-/u/ transition is notably more convoluted for the patient, who requires rotation around the rigid flap and scar tissue.

An additional source of impairment in the sibilant fricatives comes from extra air escaping the constrictions. In this case, less air can be forced through the constriction, creating a less turbulent fricative.

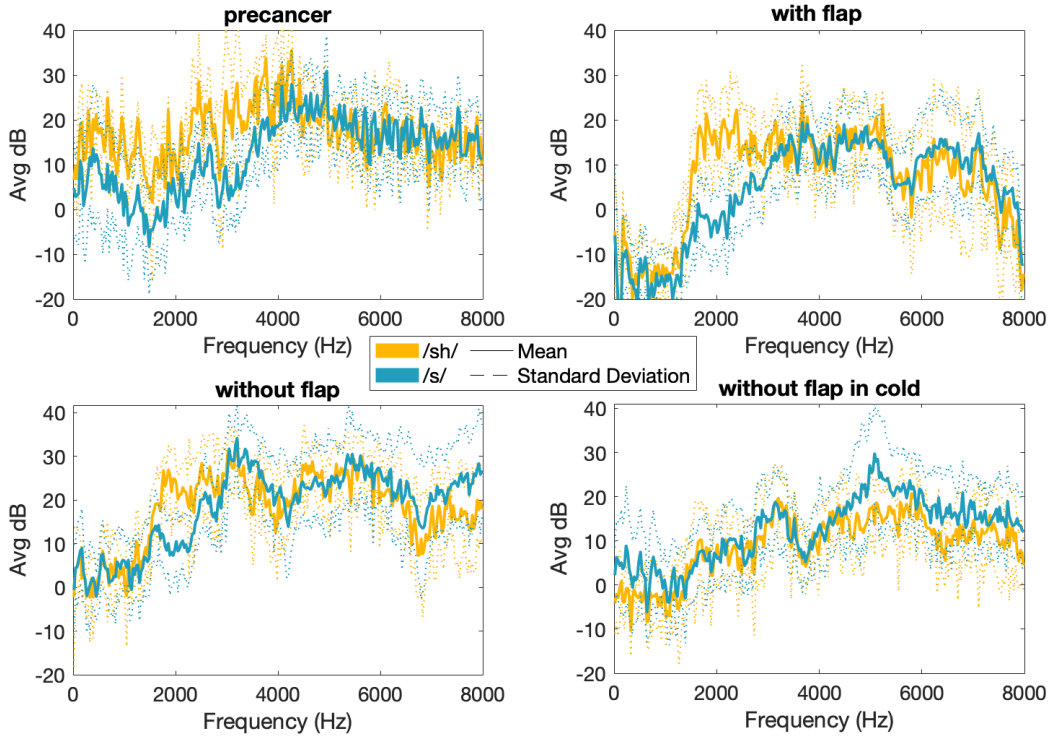


Figure 2.1: Spectrum of /s/ and /ʃ/ across all stages.

2.1.1 Results

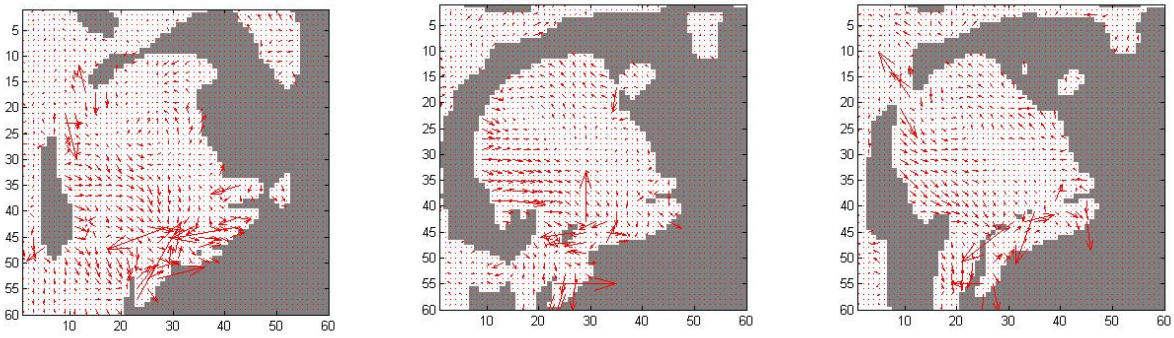
Figure 2.1 shows the spectrum of /s/ and /ʃ/ before cancer, after the glossectomy with the flap, after the glossectomy without the flap, and after the glossectomy without the flap in the cold. To control for non-uniform noise, the figure was created by finding a section of ”silence” in the original audio, taking the spectral slice, and subtracting it from the spectrum of the original audio before plottimg. Precancer, the spectral peaks of /s/ and /ʃ/ at are 4400Hz and 2500Hz, respectively, which is about what would be expected for an adult male.

In all post-operative stages, the spectral peaks decrease for both consonants. With the flap, /s/ is measured at 3700Hz and /ʃ/ is measured at 1800Hz. Without the flap, the distance between the two peaks shortens, so /s/ is at 3200Hz and /ʃ/ is at 1700Hz. In the cold, the spectra of /s/

and /ʃ/ become nearly indistinguishable, with both having peaks around 1700Hz and 3100Hz. However, /s/ additionally has a small peak around 5700Hz. In the post-operative stages, we continue to see the falloff in higher frequencies; however, this is amplified in the cold, where the shape of the spectral slice for both /s/ and /ʃ/ changes such that there is barely a distinct peak like in the other three stages.

The decrease in frequency is representative of the changing sound of the consonant. This is particularly true for /s/: as it shifts lower, it converges onto the typical range for /ʃ/, thus decreasing the intelligibility of the /s/. However, we see an increased separation in frequency between /s/ and /ʃ/, indicating an attempt to differentiate between them.

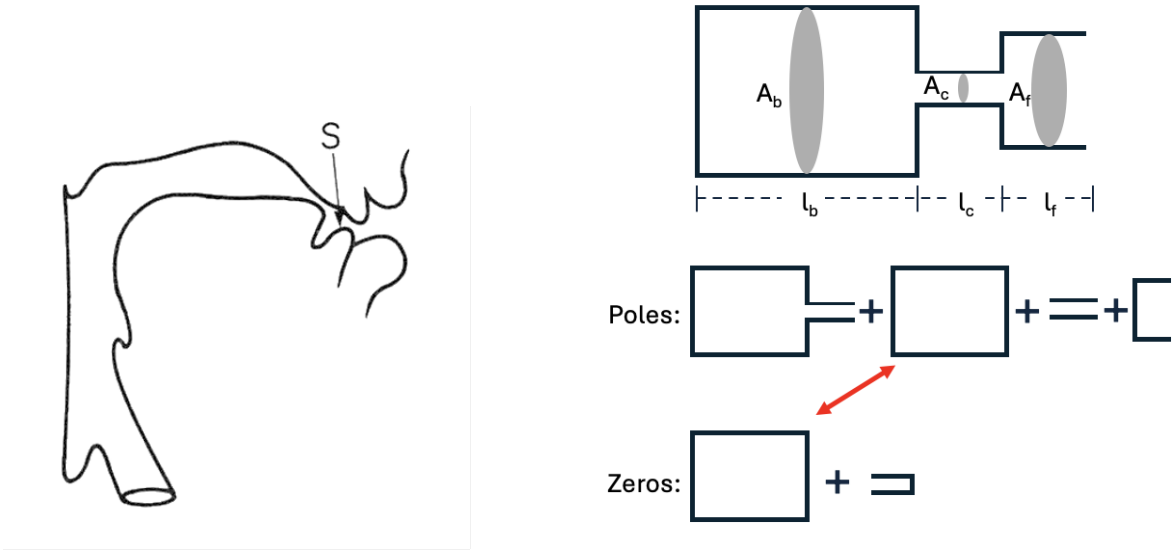
2.1.2 Vocal Tract Anatomy and Modeling



(a) Right sagittal cross-section. (b) Midsagittal cross-section (c) Left sagittal cross section

Figure 2.2: Velocity fields during /s/ constriction for 'asouk' [1]

Figure 2.2 shows the right, mid, and left sagittal views of the tongue's velocity field during the constriction for /s/ while saying 'asouk'. Normally, this constriction occurs at the alveolar ridge, but here, we see a constriction closer to the velum. This constriction is only seen on the right side, where the tongue is still intact. Because the flap is impervious to the MRI, we cannot



(a) Vocal tract during /s/ constriction [9]

(b) Consonant simple tube model

Figure 2.3: Vocal tract model for /s/

see the the tongue tip in the midsagittal or left sagittal views, but the remaining tongue does not appear to have the same constriction as the right side. Both sides have an additional constriction further back, which is unusual for an /s/. A too posterior constriction could cause a decresase in formant frequency.

A typical constriction for /s/ can be seen in Figure 2.3a. Here, we see the tongue tip making the constriction at the alveolar ridge. The vocal tract can be represented by the simple tube model, where the back, constriction, and front of the vocal tract are represented by a set of interconnected tubes. These tubes model the resonance patterns of the consonant. We assume that the constriction is small enough such that $A_b \gg A_c$ and $A_f \gg A_c$ so that the tubes can be decoupled and treated independently. For the poles, we have resonances coming from the back cavity that acts as a half-wavelength resonator, where both ends – the glottis and the constriction – are treated as being closed off. At the constriction, we have a tube with both ends open, also a half wavelength resonator. The front cavity is treated as closed at the constriction and open

at the lips, acting as a quarter wavelength resonator. Additionally, at low frequencies, the back cavity and the constriction in front of it act as a Helmholtz resonator resulting in the first formant. Mathematically, this comes out to:

$$f_b = \frac{c}{2l_b}n$$

$$f_c = \frac{c}{2l_c}n$$

$$f_f = \frac{c}{4l_f}(2n+1)$$

$$f_H = \frac{c}{2\pi}\sqrt{\frac{A_c}{A_b l_b l_c}}$$

For the zeros, we get the anti-resonances produced by the back cavity and the constriction with a closure at the right end, so no volume velocity flows into the front cavity:

$$f_b = \frac{c}{2l_b}n$$

$$f_c = \frac{c}{4l_c}(2n+1)$$

2.1.3 Pole-Zero Cancellation

In theory, the pole and zero from the back cavity should cancel each other out. However, as with any idealized model, that is not always the case in reality. Using typical lengths and areas for the /s/ constriction, we have A_b = 5cm^2, A_c = 0.5cm^2, A_f = 4cm^2, l_b = 11cm, l_b = 3cm, and l_b

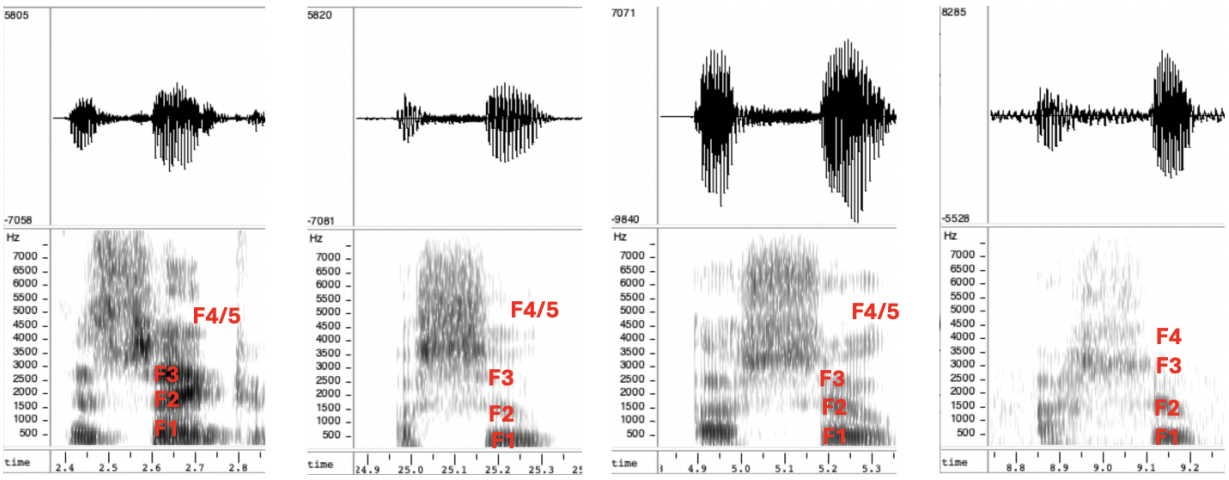
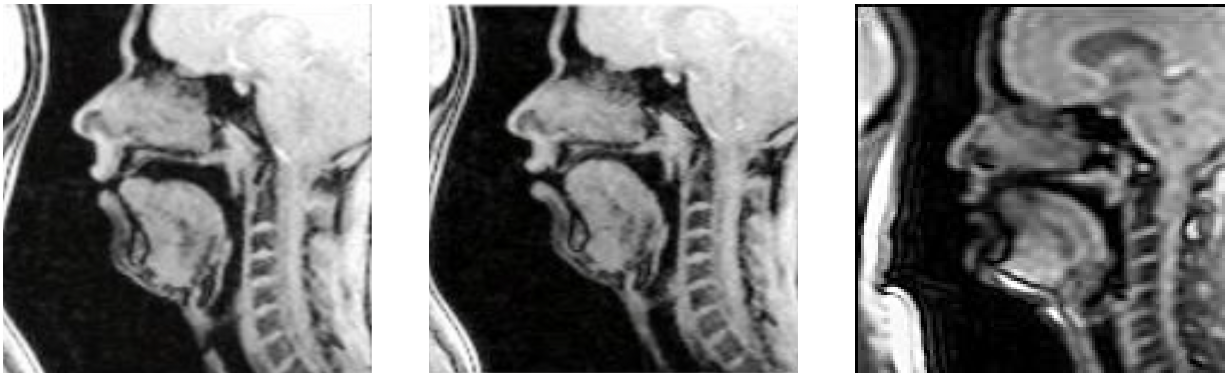


Figure 2.4: Waveforms and spectrograms for /s/ consonant during the word ”the sake” precancer (a) and ”asouk” with the flap (b), after the flap was removed (c), without the flap in the cold (d)

= 2.5cm. This creates the first five poles at 307Hz, 1590Hz, 3182Hz, 3500Hz, 5833Hz and zeros at 0Hz, 1590Hz, 2917Hz, and 3182 Hz. The first three poles and zeros should cancel out, making F4 the first visible formant at 3500Hz. For the present study, the unusual anatomy of the oral cavity could lead to lack of cancellation between the poles and zeros, leading to lower formant frequencies as we see in Figure 2.4. This begins in the precancer stage, where F3, and some of F2 are visible during the frication. After the glossectomy, both with and without the flap, F2 and F3 get stronger, indicating less pole-zero cancellation. In the cold, we additionally see less energy in high frequencies, so F3 is the most prominent formant.

The weakening of formant cancellation was observed in [10], where a male patient with a partial glossectomy and radiation treatment following a T2 tumor consistently had a low main resonance frequency. In his case, all /s/ were perceived as /ʃ/ and the main resonances were significantly lower than controls. However, the constriction location was not too posterior. Instead, the subject had too large of a constriction ($1cm^2$ as compared to the typical $0.2cm^2$), and the cavities could not be decoupled, thus the poles and zeros do not cancel out [10].



(a) Control /s/ [4] (b) Control /ʃ/ [4] (c) Patient /s/ [1]

Figure 2.5: Midsagittal MRI of sibilant fricatives

In comparing the midsaggital MRI of the patient's /s/ as compared to a control saying /s/ and /ʃ/, we see that the shape of the tongue more closely resembles the /s/ shape, even it the constriction does not appear quite as forward as the control. However the control appears to be able to more precisely point the tongue at the constriction, where as the patient's tongue tip is thick, similar to how the constriction was imprecise during the /k/ for asouk (Figure 1.1. This indicates that as in [10], the constriction is likely not complete so that the cavities cannot be decoupled and thus the resonances are not weakened by antiresonances, leading to a low main resonance. The MRI data was from when the patient had the flap, but given the similarities in spectrum between the flap and no flap data, the same phenomenon is likely occurring in both cases, despite the changes in anatomy.

Furthermore, the patient was noted to have a lateral lisp [1]. A lateral lisp occurs when excess air is escaping through the sides of the mouth as opposed to the constriction. Lateral lispers have been shown to have increased tongue pressure in the center of their mouth, whereas non-lispers have high pressure on the sides of their mouth but low pressure in the center where the air is supposed to escape. [11] This further supports the claim that there is may be an incomplete constriction for the /s/. With more of the tongue removed after the glossectomy, the ability for air

to escape out the side of the mouth for a lateral lisp likely increases, which may account for the further decrease in spectrum after the glossectomy. The loss of high frequency energy in the cold indicates leakage from an improper constriction, which could be related to the lateral lisp.

2.1.4 Cold Temperature

The data from cold weather presents the largest change in spectrum as compared to the precancer state. There is no significant main resonance in either the /s/ or /ʃ/ as there is in all the other stages, and the spectrum for both consonants is nearly the same.

There is little existing research on how the cold effects speech, either in glossectomy patients or normal people. Instead, we looked to research on how language developed, with the underlying idea that humans are prone take the path of least resistance, and thus, language features that are difficult in specific environmental conditions will not occur in languages that develop in regions with those environmental conditions. The study showed that tonal languages are most likely to be found in areas with warm, humid climates, and connected this to findings that cold air, which is also more likely to be dry, provokes muscle tension dysphonia, making it difficult to create the precise sounds needed in tonal languages [12]. If we assume the cold, dry air affects the tongue in a similar way, it is then reasonable to conclude that cold weather also causes tension in the tongue making its movement imprecise. For a glossectomy speaker, this effect could be amplified because it is already difficult for them to make the proper constriction. The lack of any resonances in the cold spectrum makes it start to resemble noise, which could be explained if the imprecise constriction increases lateral emission.

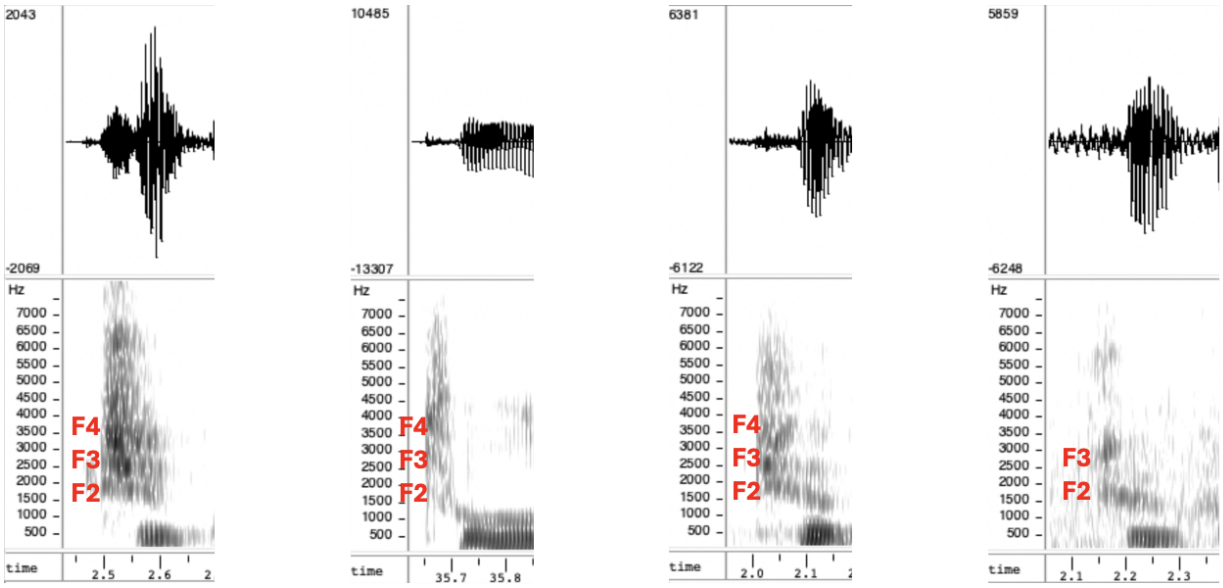


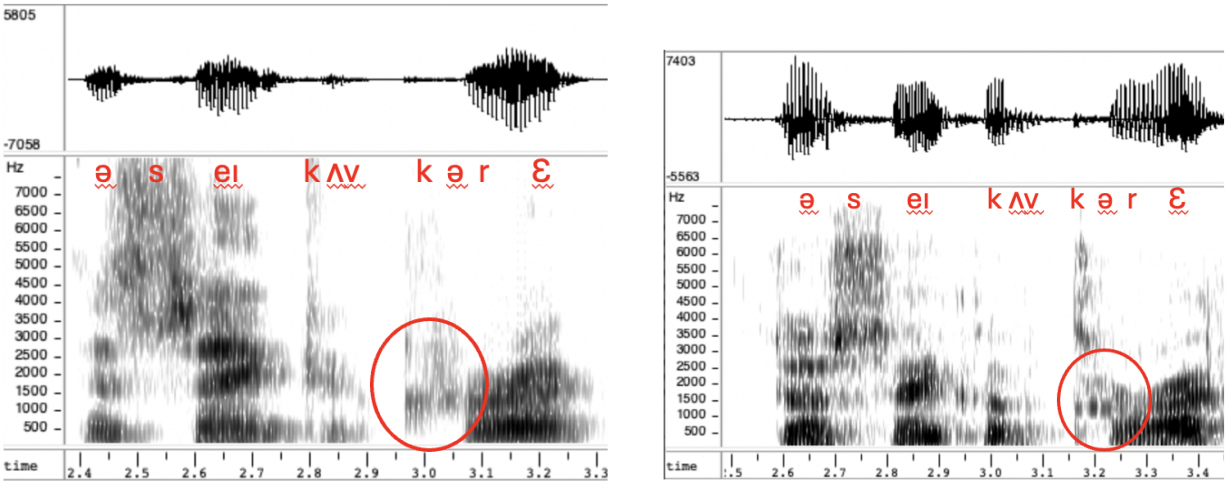
Figure 2.6: Waveforms and spectrograms of /t/ in ”two” precancer (a), and /t/ in ”tool” with the flap (b), after the flap was removed (c), and in cold weather(d)

2.2 Other Consonants

2.2.1 /t/

/t/ is an alveolar stop, which means it has the same place of articulation as /s/, but instead of the turbulent noise required with /s/, the /t/ is produced with a sharp burst of energy.

/t/ was analyzed with a succeeding /u/ due to data availability; however, in all cases there is co-articulation which brings the /t/ spectrum down. Like in the case of /s/, we expect the first formant should be F4 around 3500Hz, but instead F2 and F3 are present in all cases. Previous studies have observed that patients who have undergone a glossectomy with only a primary closure fare better in intelligibility than those with reconstruction, specifically for blade and mid portion glossal sounds. [13] While this was not the case for the /s/, the inflexibility of the flap as compared to the the closure after its removal may limit the ability to create the short jab for a /t/ while allowing the slower motion of the /s/. This aligns with notes from the initial study that



(a) Precancer

(b) After flap was removed

Figure 2.7: Spectrogram of the partial phrase ”sake of correct(ness)”.

tongue movement with the flap was slow. [1] In the cold, we again notice the lack of strength in higher frequency signals.

2.2.2 /k/

In analyzing the patient’s speech, specific qualities of speech had to be understood and accounted for.

Voiceless consonants are created when air is pushed through a constriction the oral cavity from a super glottal source. Typically, back cavity resonances cancel and we do not see any low-frequency formants. However, in Figure 2.7 we see formants continue throughout the consonant for both /s/ and /k/, precancer and after the flap was removed, albeit with low amplitude. This is indicates the constriction is not tight and allows leakage.

Additionally, we see the /k/ being produced two different ways. The first /k/ in the precancer spectrogram occurs at about 2.8s. It occurs at the end of a word in a falling stress pattern. Preceding it, we see the second and third formants of the vowel create the velar pinch as expected.

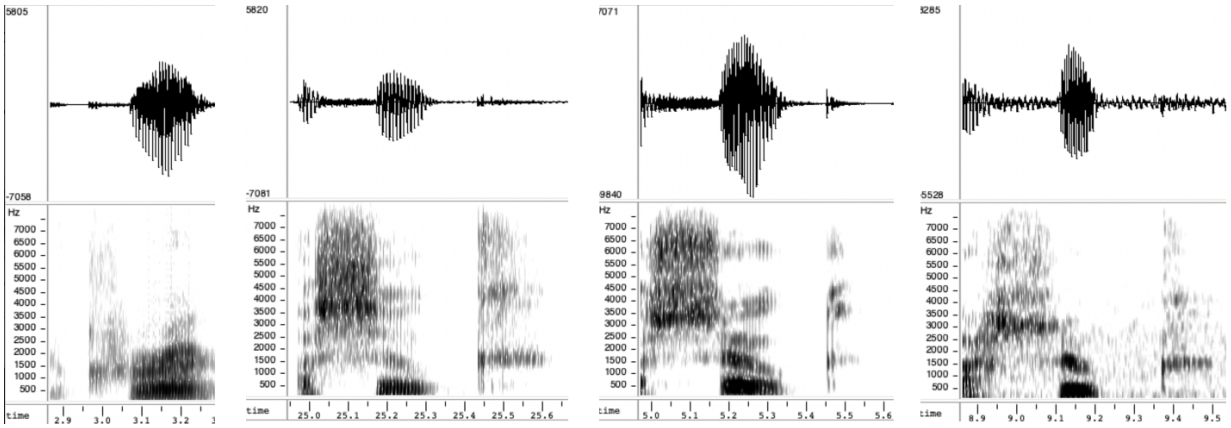


Figure 2.8: Waveforms and spectrograms of /k/ in ”correctness” precancer (a) and in ”asouk” with flap (b) after the flap was removed (c), and in cold weather(d)

However, there is no voice onset time (VoT), and instead, voicing continues right until the burst. This may mean the /k/ is created more like the voiced stop /g/. The defining feature of a canonical stop is airflow being completely blocked at the place of articulation, then released suddenly. While we see the burst releasing the /k/, the blocked airflow is missing. We see a similar, but less distinct, occurrence post glossectomy. The second /k/, starting the word ”correctness”, has a half second voice onset time, providing an appropriate stop, although the burst is less prominent. In both cases, it is followed by aspiration before leading into the rest of the word. We see F2 and F3, circled in red, falling into the succeeding /r/. Given that they comes at the beginning of a word, these /k/s are expected to be much stronger and closer to a canonical depiction.

These atypical features that naturally occur in the patient’s speech are important to identify so that they are not mischaracterized as effects of the glossectomy. Like /t/, /k/ is a stop, but it has a velar place of articulation. This means that it is mostly behind the flap and thus should be less affected. We saw in Section 1.2 that the constriction occurred in the proper location, but was imprecise and forceful, as compared to the light hit expected of stops. However, the sounds was noted to not be affected. [1]

In Figure 2.8, we analyze the /k/ starting correctness precancer as compared to the one ending ”asouk” in all other stages. While a consonant at the end of the word is typically a less clear example, in this case, the /k/s in asouk were all clearly pronounced. They all appear mostly as expected, regardless of speech stage. In each case, we see formants around 1700Hz and 4500Hz. While a /k/ would normally have the second visible formant (F5) around 5250Hz, this is dependent on the context and individual oral cavity measurements, so 4500Hz is still reasonable. We also see appropriate VoT, burst, and aspiration. We again see loss of high frequency spectrum in the cold.

Given the similarities between all the spectra, we see that the glossectomy, as well as additional complications – loss of the flap and cold temperature – play a much stronger role for sounds at the front of the mouth.

Chapter 3: Vowels

3.1 Overview

The body of research on vowels in glossectomy patients is comparatively smaller than that on consonants. This is perhaps because they are more easily compensated for, and thus, more intelligible. [14] However, they are still and important factor in understanding the full speech production of a glossectomy patient.

3.2 Vowel Space

First, we look at the average F1 and F2 formants for each of the vowels. Some vowels (ʊ, ɔ) were excluded due to insufficient data. Vowels, especially, can be shaped by the context in which they are spoken, so it was crucial to ensure there was enough data that the vowels were being accurately represented. The cold /a/ has data all from the same context, but as a corner vowel it has an important role in defining the vowel space, so instead of removing it from the dataset, it will be analyzed with the context in mind. The remaining vowels can be found on Figure 3.1. They are grouped by vowel (color) and stage of speech (shape). Looking at the color groupings, we see that overall groupings are reasonably close together with each vowel grouping mostly occupying its own space on the chart, even when an average formant may change by a few

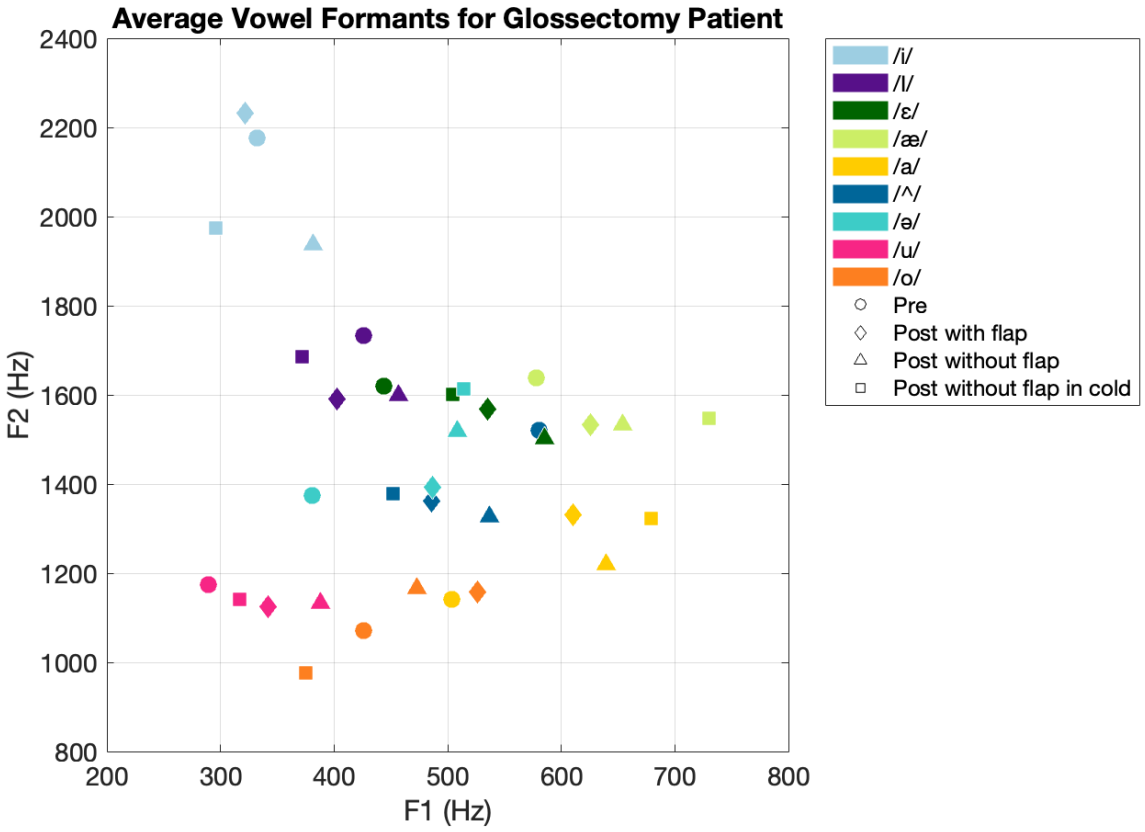


Figure 3.1: Average Formants of Vowels Across Stages of Speech

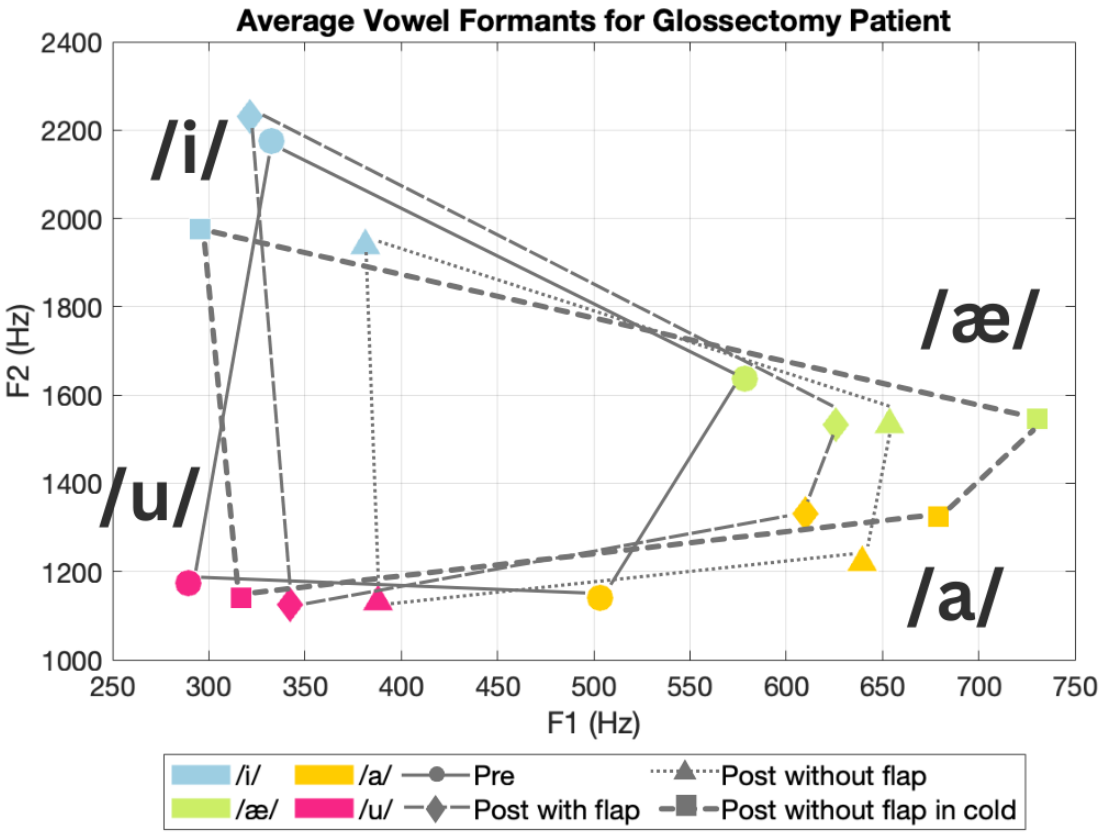


Figure 3.2: Change in Vowel Space for Glossectomy Patient

hundred hertz. The lack of major overlap between the vowel formants provides support for the claim from [14] that vowels are more intelligible: if there is less overlap, vowels are less likely to mistaken for one another, even if the formants have moved.

Figure 3.2 shows the vowel trapezoid – the shaped formed by connecting the four corner vowels, /i/, /u/, /æ/, and /a/ – for the patient across all four speech stages. As we move from precancer to cold, we see that the vowel space gets shorter and wider. F1 remains similar for the close vowels, /i/ and /u/, but spreads for the open vowels, /æ/ and /a/.

3.2.1 Open Vowels

Most existing literature does not observe changes in F1 for either open or closed vowels. [8] [15] However, a similar phenomenon was noted in a study of patients with larger tumors that required greater portions of the tongue to be removed, as well as the genioglossus muscle in the floor of the mouth. These patients' /æ/ F1 increased by over 100Hz, which the authors associated with the lower tongue height they measured post-surgery [16]. While the patient in the current study did not have as large of a tumor, he did have significant portions of the floor of his mouth removed during the initial glossectomy, and thus is likely following the trend of a lower tongue height, causing the lower F1 values on the open vowels.

3.2.2 Special Case of /i/

We notice that /i/, specifically, has a larger spread in F2. While the F2 with the flap remains similar to the precancer value around 2200Hz, the post flap and cold values drop significantly to 1950Hz. Even though the average for the flap remains similar, we see the range increases significantly, and the lower quartile of flap data overlaps with the upper quartile of the post and cold data. The full spread of the F2 values for /i/ can be seen in Figure 3.3. The /i/ vowel has the most forward place of articulation, requiring the tongue to be placed at the very front of the mouth. The glossectomy limits the tongue motion in that way, thus preventing the the F2 value from getting adequately high. The lowering of F2 values for /i/ has been noted in several studies, where they directly linked it to the inability to move the tongue forward enough [7] [15]. In [15], it was specifically noted that the decreased F2 corresponded to decreased intelligibility.

During the initial study from which this thesis is based on, the patient's tongue movement

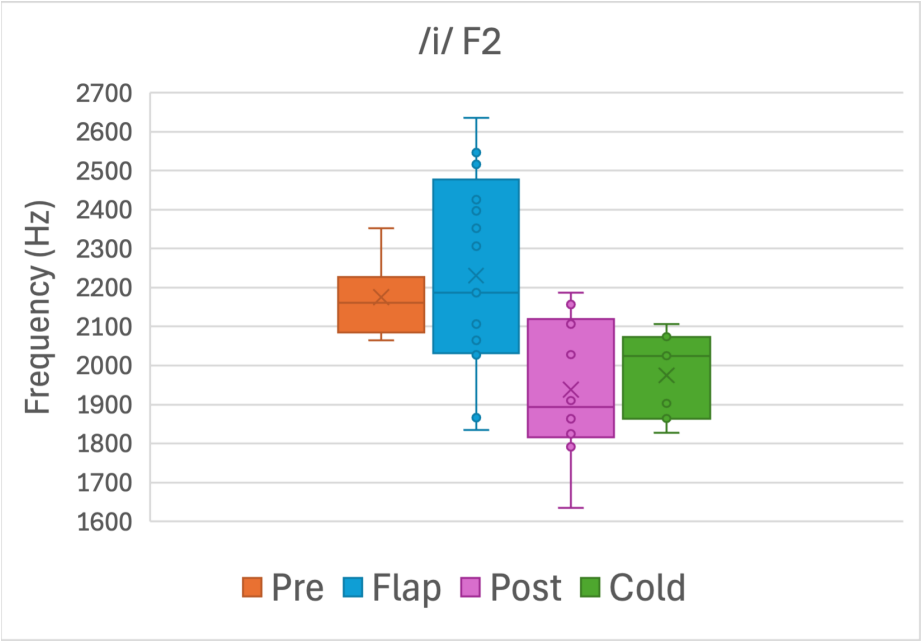


Figure 3.3: Boxplot of the F2 values for /i/ across speech stages

was slow, and he could not quite reach when asked to touch his lips. [1] At this point, the F2 value was still equal to the precancer value. It can then be assumed that with the additional loss of tongue tissue, decreases the ability to reach the far-forward place of articulation for /i/, thus causing the F2 formants to drop.

3.3 Articulation

In Figure 3.3, we saw a large range in F2 values for /i/ said with the flap, so the each speech stage was further broken down by method of articulation:

- Flap:
  - *Natural:* Natural conversational speech.
  - *Teaching:* Recordings of a lecture taught by the patient, similar to natural speech.
  - *Read Sentences:* Recordings of the patient reading a set of sentences

- *Clearly Articulated:* Recordings of the patient repeating VCV "nonsense" words with purposefully clear articulation

- Post:
  - *Natural:* Natural conversational speech.
  - *Read Sentences:* Recordings of the patient reading a set of sentences
  - *Clearly Articulated:* Recordings of the patient repeating VCV "nonsense" words with purposefully clear articulation

- Cold:
  - *Natural:* Natural conversational speech.
  - *Recited:* Recordings of the patient reciting a Shakespeare soliloquy. Although there is a (long ago memorized) script, the speech quality is similar to spontaneous speech
  - *Read Sentences:* Recordings of the patient reading a set of sentences
  - *Clearly Articulated:* Recordings of the patient repeating VCV "nonsense" words with purposefully clear articulation

These are plotted in Figure 3.4, where we see the more spontaneous style speech, especially for the flap and post stages, with significantly lower F2 values. In Figure 3.5, which shows the same subdivisions for /u/, we see natural speech is no longer the lowest. Instead, we see the flap teaching category has a exceedingly high F2. Here, again, the data changed can be accounted for by the context in which the vowel was said. Looking at the F2/F1 plot for /u/ in Figure 3.6, the data is clearly grouped by speech type moreso than articulation mode. The trends seen for the

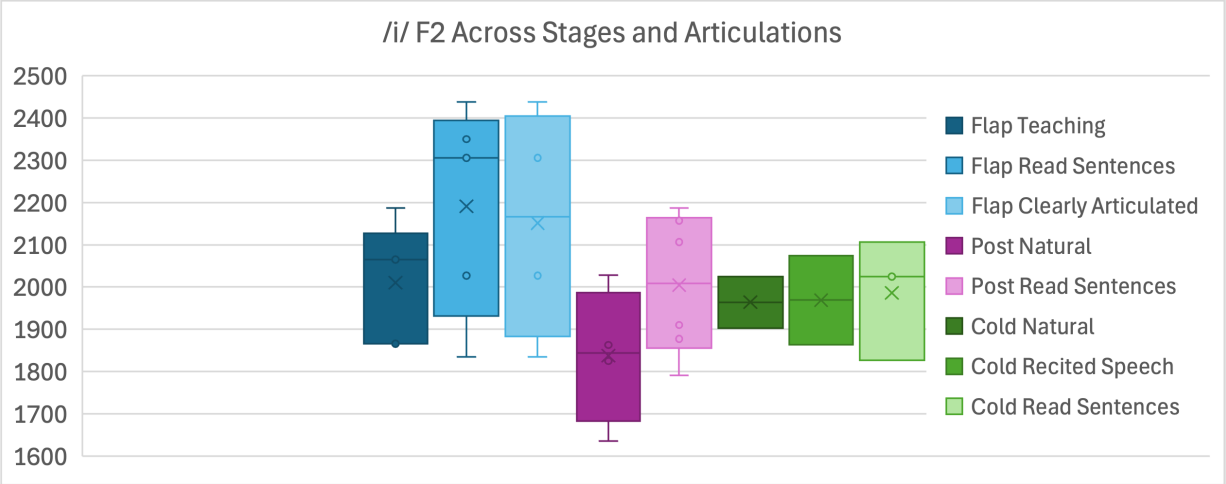


Figure 3.4: /i/ F2 values across speech stages and articulation types

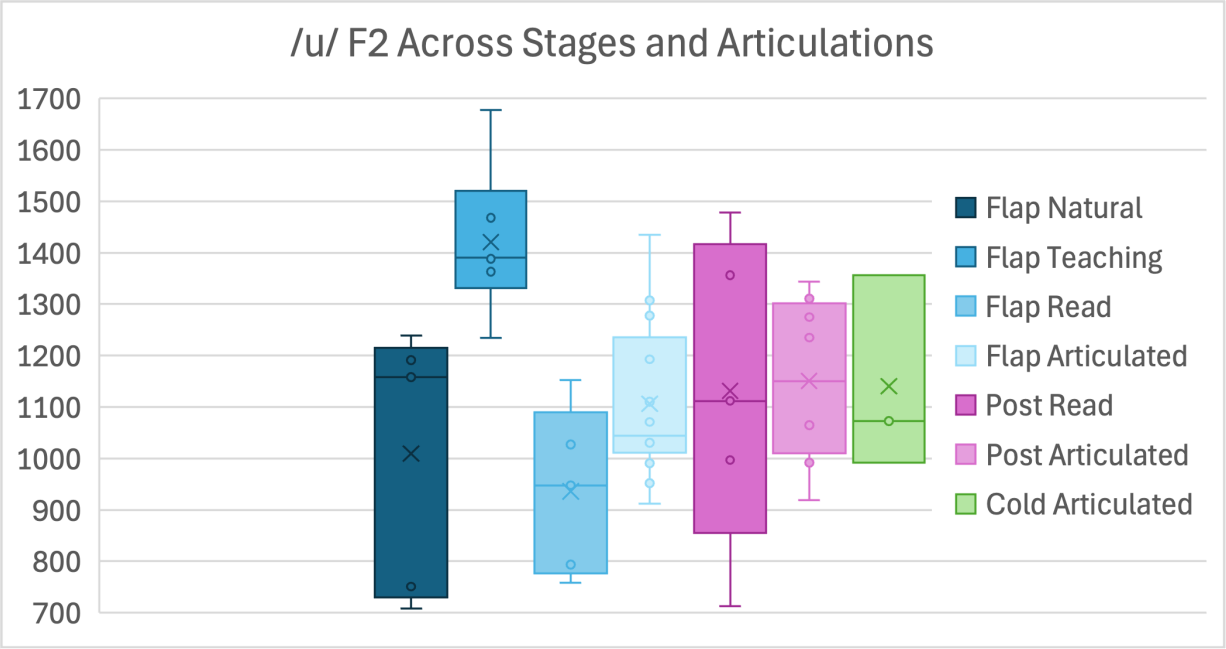


Figure 3.5: /u/ F2 values across speech stages and articulation types

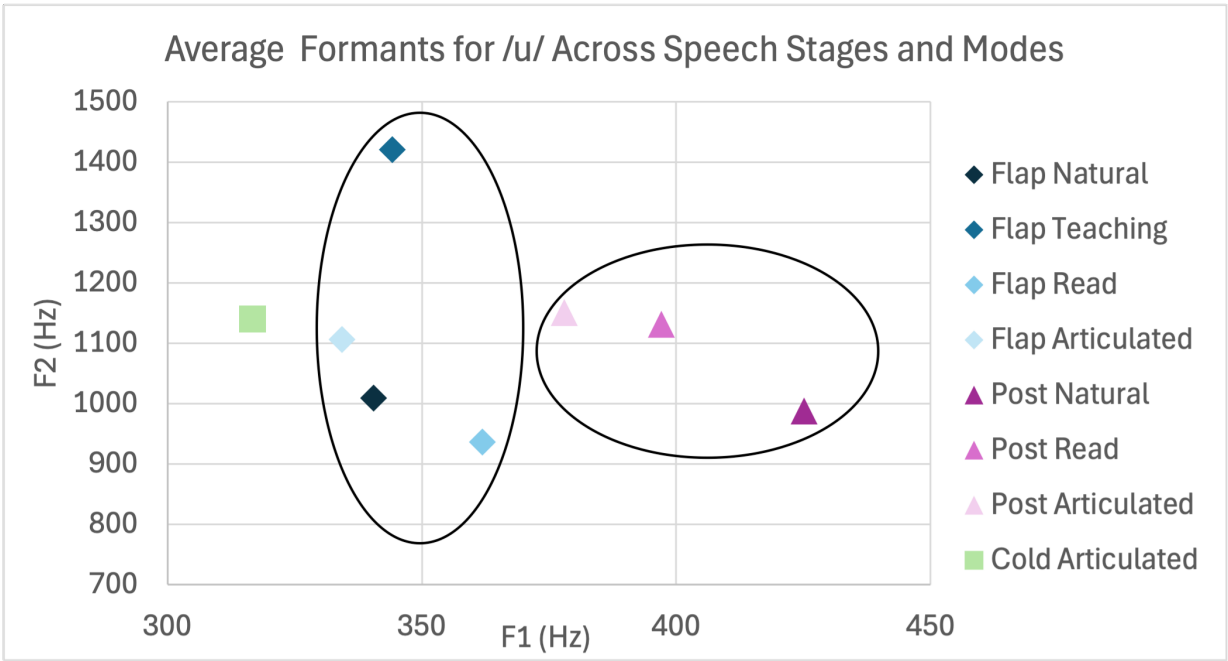


Figure 3.6: /Average formant values for /u/ across speech stages and articulation types

other vowels also fall along the lines of speech type, not articulation mode. The case of /i/ then again becomes an outlier, but given that it is known to be difficult for glossectomy patients to reach, it logically follows that the target F2 values may only have been reached when the patient was specifically trying to articulate extra clearly and expressly stretches his tongue forward, but when speaking naturally, the F2 only goes as high as needed to become intelligible. This doesn't happen for the other vowels because they do not have the same forward place of articulation.

Chapter 4: Conclusion

4.1 Summary

This study provides a longitudinal spectral analysis of speech following partial glossectomy, offering a perspective into a glossectomy patient's long-term articulatory adaptation across multiple surgical and environmental conditions. This thesis aims to provide a complete overview of the glossectomy patient's speech production, covering both consonants and vowels, and specifically investigating areas where the patients speech degrades: saying sibilant fricatives and speaking in cold temperatures. Furthermore, the results are individualized, taking into account the patient's own speech before surgery, not controls who only provide an estimate of what the patient's speech may have been.

We first investigated sibilant fricatives and observed the how both /s/ and /ʃ/ decrease in frequency after surgery. Despite the decrease in frequency, the spectra /s/ and /ʃ/ remain distinct enough to audibly distinguish. We show how this can be caused by too posterior of a constriction and too large of a constriction that causes the back cavity resonances not to cancel out, both of which are caused by the rigidity of the scar tissue which leaves the tongue body inflexible and imprecise. In the cold, the spectral distinction disappears, and we posit that this is due to the cold making the tongue muscle tense, leading to further imprecision in the constriction. We see that the lowered tongue mobility affects the spectra in /t/, which requires a fast burst, and /k/, where

the constriction is shown to be imprecise, but both to a lesser extent than the sibilant fricatives.

For vowels, that F1 increased for open vowels after the glossectomy, due to lower tongue height from removing part of the floor of the mouth. The F2 for /i/ is significantly lower because it requires the most anterior tongue position, making it hard to reach. The /i/ F2 values became higher with clearer articulation, but clearer articulation did not affect other vowels.

Despite all of the atypical speech qualities presented, the subject remains intelligible, demonstrating the immense compensation of the tongue. Overall, these findings highlight the glossectomy's disproportionate impact on speech sounds that depend on fine tongue-tip control and suggest that preserving anterior tongue mobility should remain a key consideration in surgical planning.

4.2 Future Work

Much of this research relied on imaging from [1]. However, these images only show the oral cavity during the flap stage. Updated imaging would be highly beneficial in understanding how the tongue has further adapted to the loss of tissue after the flap was removed and corroborate some of the claims made in this thesis.

Additionally, as this study focuses on a single case, the findings may not entirely be applicable to the general population. While the longitudinal data provides valuable insight into this subject's individual speech before and after a glossectomy, the ability to generalize these results to a broader population is inherently limited. Further research involving a larger and more diverse group of patients is necessary to validate the trends observed here. Given the rise of technology and ubiquity of high-quality recording devices in daily in modern daily life, such a

study is becoming increasingly possible.

Finally, more research is needed into speech changes in cold temperatures. While this study hypothesizes potential causes for the significant change in spectra, a larger study including controls that investigates the spectrum alongside physical phenomenon.

Bibliography

[1] Maureen L. Stone, Xiaofeng Liu, Hideo Shinagawa, Emi Z. Murano, Rao P. Gullapalli, Jiachen Zhuo, and Jerry L Prince. Speech patterns in a muscular hydrostat : normal and glossectomy tongue movement. 2008.

[2] Jo-Lawrence M. Bigcas and Oluwafunmilola T. Okuyemi. Glossectomy, 2024.

[3] David Rengifo and Timothy T. Lian. Radial forearm tissue transfer, 2023.

[4] Xinhui Zhou, Jonghye Woo, Maureen Stone, and Carol Espy-Wilson. A cine mri-based study of sibilant fricatives production in post-glossectomy speakers. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7780–7784, 2013.

[5] Christophe Savariaux, Pascal Perrier, Daniel Pape, and Jacques Lebeau. Speech production after glossectomy and reconstructive lingual surgery: a longitudinal study. In *Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2001)*, pages 163–168, 2001.

[6] Gomathi Saravanan, Venkatesan Ranganathan, Anitha Gandhi, and V Jaya. Speech outcome in oral cancer patients - pre- and post-operative evaluation: A cross-sectional study. In *Indian journal of palliative care*, 2016.

[7] Jun Takatsu, Nobuhiro Hanai, Hidenori Suzuki, Masahiro Yoshida, Yasuhiro Tanaka, Seiya Tanaka, Yasuhisa Hasegawa, and Masahiko Yamamoto. Phonologic and acoustic analysis of speech following glossectomy and the effect of rehabilitation on speech outcomes. *Journal of Oral and Maxillofacial Surgery*, 75(7):1530–1541, 2017.

[8] Xinhui Zhou, Maureen Stone, and Carol Y. Espy-Wilson. A comparative acoustic study on speech of glossectomy patients and normal subjects. In *Interspeech 2011*, pages 517–520, 2011.

[9] Ken N. Stevens. *Acoustic Phoenetics*. The MIT Press, 2000.

[10] Christine H. Shadle, Maureen Stone, and Wei rong Chen. The acoustic consequences for sibilants produced by glossectomiescompared to healthy controls. In *176th meeting ASA*, Victoria, B.C., Canada,, November 2018.

[11] Robert E. McGlone and William R. Proffit. Patterns of tongue contact in normal and lisping speakers. *Journal of Speech and Hearing Research*, 16(3):456–473, 1973.

[12] Caleb Everett, Damián E Blasi, and Seán G Roberts. Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 112, pages 1322–7, 2015.

[13] Chen Chuanjun, Zhang Zhiyuan, Gao Shaopu, Jiang Xinquan, and Zhang Zhihong. Speech after partial glossectomy: A comparison between reconstruction and nonreconstruction patients. *Journal of Oral and Maxillofacial Surgery*, 60(4):404–407, 2002.

[14] S. Bhattacharya, K. Thankappan, S.V. Sukumaran, M.Mayadevi, D. Balasubramanian, and S. Uyer. Volume and location of the defect as predictors of speech outcome after glossectomy: correlation with a classification. In *International Journal of Oral adn Maxillofacial Surgery*, volume 50, pages 1533–1539, 2021.

[15] Tara L. Whitehill, Valter Ciocca, Judy C-T. Chan, and Nabil Samman and. Acoustic analysis of vowels following glossectomy. *Clinical Linguistics & Phonetics*, 20(2-3):135–140, 2006. PMID: 16428229.

[16] Sandra L. Hamlet, Robin L. Patterson, and Susan M. Fleming. A longitudinal study of vowel production in partial glossectomy patients. *Journal of Phonetics*, 20(2):209–224, 1992.