**MODELING ASPECTS OF VOCAL FOLD OSCILLATIONS WITH
VALIDATION TO CLINICAL DATA**

**A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF**

MASTER OF SCIENCE

IN

MECHANICAL ENGINEERING

AUGUST 2006

**By
Sarah A. Bentil**

Thesis Committee:

**John S. Allen, III, Chairperson
Yuling Yan
Todd Reed**

We certify that we have read this thesis and that, in our opinion, it is satisfactory in scope and quality as a thesis for the degree of Master of Science in Mechanical Engineering.

10 002605517

UNIVERSITY OF HAWAII

HAWN

Q111

.H3

no. 4077

THESIS COMMITTEE

Chairperson

DEDICATION

**In loving memory of my grandfather,
Paul Bentil**

ACKNOWLEDGEMENTS

I would like to thank my current advisor and committee chairman, Dr. John Allen, III, for his generous guidance, patience, and constant encouragement. His invaluable comments and suggestions during the last research phase have contributed to the contents that I present in this thesis. He was a great mentor to me and without his help this thesis would not have been completed.

I would like to thank Dr. Yuling Yan, who was my advisor from August 2004 – December 2005, for introducing me to this exciting research area. I would like to thank her for her supervision, insight and knowledge of voice pathophysiology, laryngeal imaging, and diagnostic and analytic techniques that she introduced to me. I would also like to thank her for the clinical data she shared with me and for the financial support that I received from her NSF grant (#BS 0402439) during her tenure at the University of Hawai‘i, which made it possible for me to accomplish this thesis work.

I would also like to thank Dr. Todd Reed for his time and interest in serving on my thesis committee. His exceptional experience, insights, and knowledge in the field of signal processing and analysis greatly influenced the results of my research work. I am truly thankful and honored to have had such exceptional mentors on my thesis committee.

I would like to thank my boyfriend William Jackson for his patience, emotional support, and contributions to this thesis. I appreciate all the time and effort that he generously gave to help me to finish my thesis.

Finally, I would like to thank my father, Dr. Daniel Bentil, my mother, Mrs. Beatrice Bentil, my sister, Paulina Bentil, and my brother, James Bentil, for their love, support, patience, and encouragement. I would especially like to recognize my father for being a wonderful mentor and contributing towards my scientific goals. My gratitude to him is beyond description.

ABSTRACT

The two mass model describing vocal fold oscillations is used to study voice abnormalities such as Diplophonia, Papillomas, and Parkinson's Diseases. This model, which utilizes masses, springs, and dampers, was able to capture some of the behaviors associated with clinical data describing both normal and abnormal vocal fold dynamics. The clinical data was obtained through standard techniques, acoustic and endoscopic high-speed laryngeal imaging techniques, of the resulting acoustic signals and glottal area waveforms.

A comparison of the two mass mechanical model predictions and clinical data was done using signal processing techniques: Fast Fourier Transforms, Wavelet Transforms, and Spectrograms. Results from the analyses provided a comprehensive representation of the quantitative characteristics of the vocal fold dynamics. The methods and techniques presented in the thesis might potentially be used by clinicians and other specialists to facilitate, more accurate, diagnosis and treatment of vocal fold abnormalities, such as found in patients with Diplophonia, Papillomas, and Parkinson's Diseases.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT.....	vi
LIST OF FIGURES	viii
CHAPTER 1 INTRODUCTION	1
1.1 Overview	1
CHAPTER 2 VOCAL FOLD PHYSIOLOGY	5
2.1 Vocal Fold Structure and Function.....	5
2.2 Vocal Fold Disorders	12
CHAPTER 3 VOCAL FOLD MODELLING	15
3.1 Review of Vocal Fold Models	15
3.2 Two Mass Model.....	19
3.2.1 The Steinecke et. al. Model	20
3.2.2 Two Mass Model - Normal Case.....	22
3.3 Extension of Two Mass Model - Parkinson Case.....	40
CHAPTER 4 ACQUISITION OF CLINICAL DATA	44
4.1 Introduction to the Acquisition of Voice Samples	44
4.1.1 Glottal Area Waveform (GAW)	46
4.1.2 Acoustic Waveform (AW).....	48
4.1.3 Other Acquisition Methods.....	49
CHAPTER 5 RESULTS AND DISCUSSION.....	51
5.1 Signal Processing Methods	51
5.1.1 Fast Fourier Transform (FFT)	51
5.1.2 Short Time Fourier Transform (Spectrogram).....	53
5.1.3 Wavelets.....	58
5.2 Spectral Analysis of Normal and Pathological Vocal Folds	61
5.2.1 Waveforms.....	62
5.2.2 FFT	72
5.2.3 Spectrogram	82
5.2.4 Wavelets.....	87
5.2.5 Model Predications - Diplophonia	98
5.2.6 Model Behavior Using Spectral Analysis	100
CHAPTER 6 CONCLUSION AND FUTURE DIRECTIONS	105
6.1 Conclusion.....	105
6.2 Future Work	107
REFERENCES.....	108

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Location of the larynx in relation to the head and neck region.....	5
2. (a) View of the larynx from above. (b) View of the larynx from behind	6
3. Cross section schematic of the right vocal fold featuring the tissue layers	9
4. Three different schemes for labeling the structures of the vocal folds.....	10
5. Diagram of an idealized cycle of the vocal fold vibration.....	11
6. One Mass Model	16
7. Two Mass Model.....	19
8. Simulation of the upper and lower mass displacements.....	24
9. Reproduction of the upper and lower mass displacements.....	24
10. Simulation of the glottal volume flow velocity.....	25
11. Reproduction of the upper glottal volume flow velocity.....	25
12. SHO – (i) Without Dissipation and (ii) Without Forcing (Displacement).....	28
13. Two Mass System – (i) Without Dissipation and (ii) Without Forcing (Displacement).....	29
14. SHO – (i) Without Dissipation and (ii) Without Forcing (Energy).	30
15. Two Mass System – (i) Without Dissipation and (ii) Without Forcing (Energy).....	30
16. SHO – (i) Without Dissipation and (ii) Forcing (Displacement).....	31
17. Two Mass System – (i) Without Dissipation and (ii) Forcing (Displacement).....	32
18. SHO – (i) Without Dissipation and (ii) Forcing (Energy).	33
19. Two Mass System – (i) Without Dissipation and (ii) Forcing (Energy)	33
20. SHO – (i) Dissipation and (ii) Without Forcing (Displacement).....	34
21. SHO – (i) Dissipation and (ii) Without Forcing (Phase Plot).....	35
22. Two Mass System – (i) Dissipation and (ii) Without Forcing (Displacement).....	35
23. SHO – (i) Dissipation and (ii) Without Forcing (Energy).....	36
24. Two Mass System – (i) Dissipation and (ii) Without Forcing (Energy)	37
25. SHO – (i) Dissipation and (ii) Forcing (Displacement).....	38
26. Two Mass System – (i) Dissipation and (ii) Forcing (Displacement).....	38
27. SHO – (i) Dissipation and (ii) Forcing (Energy).....	39
28. Two Mass System – (i) Dissipation and (ii) Forcing (Energy).....	40
29. Simulation for the lower mass displacements for a Parkinson's disease patient.....	42
30. Reproduction of the lower mass displacements for a Parkinson disease using Zhang et. al.'s model	43
31. Age-related changes in speaking fundamental frequency for males and females.....	45
32. Vibrating vocal folds for 24 frames (0.012 sec).....	46
33. Example of GAW for a normal 36-year-old female patient.	47

34.	Example of Acoustic Waveforms (AW) for a normal 36-year-old female patient	49
35.	FFT example using a sine curve.....	52
36.	Waveform using a sine curve.....	53
37.	Sample of tiled scheme for the windowed Fourier Transform (Spectrogram)	56
38.	Spectrogram example using a sine curve.....	57
39.	Sample of the logarithmic tiling scheme for the Wavelet Coefficient	59
40.	Wavelet example using a sine curve - db4.....	60
41.	Wavelet example using a sine curve - db9.....	61
42.	Normal Patient - Acoustic Waveform	63
43.	Normal Patient - GAW	63
44.	Normal Voice: Model Data vs. Clinical Data - GAW.....	64
45.	Parkinson Patient - GAW.....	65
46.	Parkinson Voice: Model Data vs. Clinical Data - GAW	66
47.	Papillomas Patient - Acoustic Waveform.....	67
48.	Papillomas Patient - GAW	67
49.	Papillomas Voice: Model Data vs. Clinical Data - GAW	68
50.	Diplophonic Patient - Acoustic Waveform (Pre-treatment).	69
51.	Diplophonic Patient - Acoustic Waveform (Post-treatment).....	70
52.	Diplophonic Patient - GAW (Pre-treatment).	70
53.	Diplophonic Patient: Normal vs. Diplophonic Patient (Post-treatment) - GAW.....	71
54.	Diplophonic Voice: Model Data vs. Clinical Data - GAW (Pre-treatment).....	72
55.	Normal Patient - Acoustic Waveform (FFT).	73
56.	Normal Patient - GAW (FFT).....	73
57.	Normal Voice (Model) – GAW (FFT)	74
58.	Parkinson Patient - GAW (FFT)	75
59.	Parkinson Voice (Model) - GAW (FFT)	76
60.	Papillomas Patient - Acoustic Waveform (FFT).....	76
61.	Papillomas Patient (40 years old) - GAW (FFT)	77
62.	Papillomas Voice (Model) - GAW (FFT).....	78
63.	Diplophonic Patient - Acoustic Waveform (FFT: Pre-treatment).....	78
64.	Diplophonic Patient - Acoustic Waveform (FFT: Post-treatment)	79
65.	Diplophonic Patient - GAW (FFT: Pre-treatment).....	80
66.	Diplophonic Patient - GAW (FFT: Post-treatment)	80
67.	Diplophonic Voice (Model) - GAW (FFT: Pre-treatment).....	81
68.	Normal Patient - Acoustic Waveform (Spectrogram)	83
69.	Normal Patient - GAW (Spectrogram).....	84
70.	Papillomas Patient - Acoustic Waveform.(Spectrogram).....	85

71.	Diplophonic Patient - Acoustic Waveform (Spectrogram: Pre-treatment).....	86
72.	Diplophonic Patient - Acoustic Waveform (Spectrogram: Post-treatment).....	86
73.	Normal Patient - Acoustic Waveform (Wavelet).....	88
74.	Normal Patient - GAW (Wavelet).....	89
75.	Normal Voice (Model) - GAW (Wavelet).....	90
76.	Parkinson Patient - GAW (Wavelet).....	90
77.	Parkinson Voice (Model) - GAW (Wavelet).....	91
78.	Papillomas Patient - Acoustic Waveform (Wavelet).....	92
79.	Papillomas Patient - GAW (Wavelet).....	93
80.	Papillomas Voice (Model) - GAW (Wavelet)	94
81.	Diplophonic Patient - Acoustic Waveform (Wavelet: Pre-treatment).....	95
82.	Diplophonic Patient - Acoustic Waveform (Wavelet: Post-treatment).....	96
83.	Diplophonic Patient - GAW (Wavelet: Pre-treatment)	96
84.	Diplophonic Patient - GAW (Wavelet: Post-treatment).....	97
85.	Diplophonic Voice (Model) - GAW (Wavelet: Pre-treatment)	98
86.	Diplophonic Voice (Model) - GAW (FFT: Increasing nonlinear parameters).....	99
87.	Diplophonic Voice (Model) - GAW (Wavelet: Increasing nonlinear parameters).....	100

CHAPTER 1

INTRODUCTION

1.1 Overview

The human larynx, also known as the voice box, is an organ in the body that serves two vital purposes. First, it is a component of the respiratory system, which prevents air from escaping from the lungs by acting as a valve. In a similar fashion, it also prevents foreign substances from entering the lungs. The secondary role of the larynx is to act as a device used by humans for communication purposes. As a result, humans are able to communicate through voicing and/or vocalization. “Voicing” is a term that is used to describe all sounds related to speech, including words, and “Vocalization” occurs when sounds, which do not contain any words, are produced. In general, the process of voicing or vocalization is called phonation¹.

Being plagued with an abnormality that can affect your ability to communicate with the larynx can have devastating physical and social implications, especially to an individual who relies on this organ for their profession such as an opera singer or a teacher. The diagnosis of these disorders that plague the laryngeal region can be challenging since some disorders share similar traits and characteristics². In addition, the interpretation of the disorder can be subjective, especially if the clinician is interpreting the results solely on how the patient’s voice sounds and the images or video clips of the vocal folds, which are soft tissue structures contained within the larynx³.

We attempt to help quantify the subjective nature present in the diagnosis of disorders by seeking a better understanding of the characteristics of specific pathologies in this thesis. We will compare data describing the patient's voice obtained clinically with that obtained through simulation. An investigation through the use of simulation, specifically how the parameters describe characteristics that are observed in the patient data is explored. Our results may assist with current diagnosis methods and provide clinicians with more comprehensive understanding, which may lead to the optimal treatment plan for the patient.

Chapter 2 will examine the basic anatomy and physiology of the larynx, focusing on the vocal folds. Muscles and cartilages, which work concurrently with the vocal folds resulting in sound production, are explained. An understanding of the biomechanics is necessary to characterize vocal fold abnormalities, which leads to accurate diagnosis. Examples of vocal fold abnormalities along with their pathophysiological explanation are also presented in this chapter.

There have been numerous studies focused on the development of diagnostic techniques that can be used by clinicians in the detection and characterization of vocal fold abnormalities⁴. These studies include both empirical and theoretical research. Chapter 3 outlines these methods emphasizing the theoretical work that utilizes both simple and complex models that describe the oscillatory behavior of the vocal folds. These models can be described as discrete elements, which include a combination of masses, springs, and dampers. In reality, the continuum space describes the continuous shape of the vocal folds and influences how this shape affects the oscillatory behavior⁵. Since the focus of this thesis is on the development

of diagnostic tools, the use of more complex models at either the discrete or the continuum level is not necessary to achieve this goal⁵. Simpler discrete models provide adequate information to capture the general characteristics present in various abnormalities. In order to develop diagnostic tools, we use differential equations which describe a two mass model, a combination of two masses, dampers, and springs, for the vocal fold vibratory behavior associated with sound production.

Current diagnostic techniques that are used by clinicians may be subjective³. These include using perceptual ratings, and stroboscopy, to capture video images which are interpreted during diagnosis³. We outline these techniques in Chapter 4. In Chapter 5, we highlight some objective diagnostic tools, which might be used by clinicians in the future. These tools include signal processing techniques, specifically time-frequency methods such as the fast Fourier transform (FFT), short time Fourier transform (STFT), and Wavelet transforms (WT). The FFT and STFT provide a means of transforming the signal with the oscillatory behavior describing the vocal folds into the frequency domain⁶. The transform uses sinusoidals as the basis function for transforming the time signal into the frequency domain. Characteristics of the pathology, which may otherwise be missed in the time domain, can be seen in the frequency domain. Wavelets transform the signal into a pseudo-frequency domain. The basis function is no longer composed of sinusoidals but rather wavelets, which are functions that can better extract the transients present in the signal⁶. The methods proposed in Chapter 5 are a means of interpreting and understanding some of the characteristics of the data acquired by clinicians for diagnosis. The time-frequency methods are applied to the two mass model to gain insight into the specific

patient data. In addition, characteristics of the vocal fold can be extracted from both the patient data and models. These results might aid the clinician in the diagnoses of the vocal fold abnormalities. The pathophysiology of the vocal folds can be described by using the model along with the results acquired from the time-frequency methods. Chapter 6 summarizes and provides a conclusion of the aforementioned chapters. Future work and directions are also discussed.

CHAPTER 2

VOCAL FOLD PHYSIOLOGY

2.1 Vocal Fold Structure and Function

The larynx is located in the neck in front of the spinal cord and is surrounded by a large number of blood vessels, nerves, glands, and muscles of the human body (Figure 1)¹. The oscillatory and symmetrical natures of the vocal folds, which are contained within the larynx, produce an air space between the two vocal folds referred to as the glottis. This airspace occurs when the vocal folds are adducted. The subglottal area is the area below the glottis and the supraglottal area is the area above the glottis.

When the vocal folds oscillate in response to changes of pressure between the supraglottal pressure and the subglottal pressure, the result is the onset of sound from the larynx. These changes in pressure are the result of the pressure exerted on the lungs from the diaphragm and abdominal muscles.

The biological structures that are necessary for the vocal folds to oscillate include the laryngeal cartilages, intrinsic muscles (which are muscles that are located inside the larynx and connect to the cartilage), and extrinsic muscles (which connect the larynx to other external structures

such as the hyoid bone). A comprehension of these structures, which directly influence the vibratory motion of the vocal folds, is important for the characterization of vocal fold pathologies.

The following figure shows the locations of these cartilages and muscles within the larynx.

We describe the muscles and cartilages that are important in the oscillatory behavior of the vocal folds below. These muscles are symmetrically arranged on the left and right sides of the larynx⁷. Unless otherwise noted, the muscles described are intrinsic and hence are components of the larynx that control phonation.

1. **Cricoid cartilage:** This cartilage lies directly below the thyroid cartilage and forms a solid ring surrounding the laryngeal airway¹.
2. **Thyroid cartilage:** This cartilage fits over the cricoid cartilage and it comes to a point in the front. This point is known as the thyroid notch, or Adam's apple. The front ends of the vocal folds are attached to the inside of the thyroid notch⁷.
3. **Arytenoid cartilage:** This cartilage allows for the attachment of the vocal fold ligament¹. They sit on top of the back portion of the cricoid cartilage and hold the back ends of the vocal folds⁷.
4. **Posterior cricoarytenoid muscle:** This muscle is the primary abductor for the vocal folds¹. They open the glottis by pulling the back ends of the arytenoid cartilages together. This pulls the front ends (where the vocal folds attach) apart, therefore pulling the vocal folds apart⁷.
Abduction occurs during respiration.
5. **Lateral cricoarytenoid muscle:** This muscle acts as an adductor by allowing for the closure of the glottis to be possible by pulling the back end of the arytenoid cartilages apart. This action pulls the front ends together, making the vocal folds as a whole come together⁷. The process of adduction occurs during phonation.
6. **Thyroarytenoid muscle:** This muscle forms the body of the vocal folds. They shorten the vocal folds by pulling the arytenoid cartilage toward the thyroid cartilage. In other words, the back ends of the vocal folds are pulled towards the front end of the vocal folds. This motion

shortens the vocal folds causing them to vibrate more slowly, thus lowering the pitch. The thyroarytenoid muscles also create the force necessary to strengthen glottic closure. This is accomplished by keeping the vocal folds together and resisting the air stream from the lungs⁷.

7. Cricothyroid muscle: This muscle lengthens the vocal folds by pulling the thyroid cartilage down on its hinge. This motion increases the distance between the arytenoids and the thyroid notch. As a result, the vocal folds are lengthened and vibrate faster, thus raising the pitch⁸.
8. Interarytenoid muscles: These muscles are composed of the transverse arytenoid muscles and the oblique arytenoid muscles. These muscles bring the two arytenoid cartilages together to provide medial compression of the vocal folds. The act of compressing the vocal folds to resist the air pressure from the lungs is due to these muscles⁷.
9. Hyoid bone: This bone acts as an anchor for tissue connections and protects some of the soft tissues in the upper larynx and lower pharynx against injury that might be due from blows to the neck¹.
10. Sternohyoid muscle (extrinsic): This muscle connects the sternum and the hyoid bone. Movement of this muscle results in the lowering of the larynx¹.
11. Thyrohyoid muscle (extrinsic): This muscle connects the thyroid cartilage to the hyoid bone and results in the elevation of the larynx¹.

Modeling the oscillatory behavior of the vocal folds, which requires knowledge of the muscles and cartilages, is quite complicated due to the multiple structures present and their different corresponding functions. To simplify the modeling process, all cartilages and all muscles, with the exception of the thyroarytenoid muscle, may be neglected in a model¹. Although these neglected muscles contribute to the overall physiology of the vocal folds, they are not necessary to explain and model the general oscillatory behavior of the folds, which can be accomplished by describing only the behavior of the thyroarytenoid muscle¹.

The thyroarytenoid muscle and four other layers of tissues contiguous to this muscle collectively describe the vocal fold structure. Figure 3 shows a schematic of these five layers of the vocal fold tissue necessary to model the function and structure of this organ. Each tissue layer has different properties and structures.

The epithelium layer, which is the outermost layer, is a thin skin that acts as a covering or lining. It is layered and scale-like and is between 0.05 to 0.10 mm in thickness¹. The superficial layer contains loosely organized elastin fibers that are surrounded by interstitial fluid.

The elastin fibers contribute to the elastic properties of the vocal folds and the interstitial fluid provides a means of transporting nutrients and waste. The superficial layer is approximately 0.5 mm in thickness¹. The intermediate layer is similar to the superficial layer in the sense that both contain elastin fibers. However, they are more

uniformly oriented to allow for more defined elastic behavior¹. The intermediate layer also contains some collagen fibers that provide strength and flexibility to the tissue in addition to limiting elongation¹. The deep layer contains mostly collagen fibers. The thickness of the intermediate and deep layer combined is 1 to 2 mm¹. The thickness of the thyroarytenoid muscle is 7 to 8 mm¹. Variation of these tissue properties, which cause a change of structure or size, for example, may result in vocal fold oscillations that deviate from the norm and give rise to vocal fold abnormalities³.

Different labeling schemes are used to group the five different vocal fold soft tissue layers¹. These labeling schemes can be used to explain the mechanics of phonation in detail or in general. The schemes are summarized in Figure 4. The two-layer scheme will be used when modeling the vocal folds using the two mass model as is described further in Chapter 3.

The structure of the vocal folds described by the two-layer scheme is also referred to as the cover-body concept⁹. The cover layer is composed of mucosal tissue that serves as a sheath around the body-layer. This cover tissue allows a

surface wave (mucosal wave) to be generated on the vocal fold surface. This mucosal wave propagates longitudinally along the axis of airflow during phonation⁹. The body tissue (inertial acoustic loading) is used to describe the lateral displacement of the vocal folds. The use of the cover-body layer description is convenient and frequently used due to its simplicity in the study and modeling of vocal folds' vibratory characteristics⁹.

Figure 5 shows an idealized cycle of the vocal fold vibration where the mucosal wave behavior occurs.

Figure 5a starts the cycle with both left and right vocal folds in contact and closed. At this stage, the airway is closed and phonation is occurring. Figure 5b and Figure 5c indicate a lateral movement of the cover surface that continues until the left and right sides separate, thus causing an opening in the airway. Figure 5d displays the action of the vocal folds once the maximum lateral displacement has been achieved.

Figure 5e illustrates the movement of the vocal folds towards the center starting from the lower portion of the vocal folds and then followed by the upper portion. Figure 5f shows the collision of the left and right sides of the lower portions of the vocal folds. This results in the closure of the airway once again. Figure 5g continues the closure process as it migrates from the bottom of the vocal folds and continues upwards. Figure 5h⁹ shows the complete closure of the glottis. The entire oscillatory process repeats itself cyclically at the fundamental frequency of vibration (F_0) as depicted in Figure 5. It should be noted that the fundamental frequency depends upon the length of the vocal folds, tension, mass, along with the subglottic pressure⁸.

2.2 Vocal Fold Disorders

The vocal folds might be considered a biomechanical device in which disturbances or loss of normal function causes disorders³. For example, disorders due to disturbances can be created by changes in the material properties of the vocal folds thus directly affecting the response. These variances could cause a loss of the elasticity properties of the intermediate layer, which would affect the oscillatory pattern of the vocal folds and as a result, the quality of sound that is produced might change.

Viral or bacterial infections, stress, and pollutants can affect the material properties of vocal folds resulting in abnormal sounding voice⁸. These changes can be classified as either organic disorders or functional (behavioral) disorders¹. Organic disorders result from detectable physiological or structural change in the organs, for

example lesions caused by irritants or infections. Functional disorders are idiopathic in nature and can be the result of the individual's improper use of the voice or emotional stressors^{1,8}.

These two general categories of disorders can be decomposed into five more descriptive groups as listed below¹.

1. Infectious and inflammatory conditions-the flu or allergies causes hoarseness (dysphonia) and/or lowers the pitch of the voice.
2. Vocal misuse/abuse (chronic)-constant screaming causes swelling and excessive muscle tension of the vocal folds.
3. Benign and malignant growths-the growth of cysts or papillomatosis can affect the quality of voice.
4. Neuromuscular diseases-vocal tremor or Parkinson's disease can produce a strained or breathy voice.
5. Psychogenic conditions-patients suffering from a psychological trauma can experience a loss of voice or abnormal voice.

The classification of vocal fold disorders can be quite complex since some organic disorders may display symptoms of functional disorders and vice versa¹. We utilize signal-processing techniques to isolate the characteristics present in selected vocal fold disorders related to the two general categories of disorders. Vibration analysis techniques are used on selected voice disorders as a means to explain the pathophysiology relationship. The voice disorders examined in the analysis include that of patients with normal voices as well as patients suffering from Parkinson's disease, Papillomas of the vocal folds, and Diplophonia. Specific cases along with

associated signal processing and vibration analysis techniques are discussed in more detail in Chapter 5. We briefly describe the disorders in this chapter.

Diplophonia is a voice disorder caused by the differential tension of the vocal folds^{10,11}. This produces multiple tones or frequencies simultaneously. The pre-treatment and post-treatment data of a female patient that suffered from Diplophonia is analyzed. Post-treatment of this disorder, in some cases, involves behavioral modification and surgery to remove the lesions arising from the differential tension¹².

During voice production, a fluctuation in the voice fundamental frequency and intensity is a phenomenon necessary for sound production. However, abnormal fluctuations of the fundamental frequency due to involuntary muscular contraction of the vocal folds between the frequency ranges of 4 – 8 Hz produces an event referred as a vocal tremor, which affects the quality of sound produced¹³. Neurological disorders such as Parkinson's disease often exhibit vocal tremor as one of its symptoms and these are examined in later chapters¹⁴. Typically, a frequency band from 4-7 Hz characterizes the vocal tremors that often occur in Parkinson's patients¹⁵.

Papillomas refer to warts on the vocal folds caused by viral infections¹⁶. These warts cause recurrent lesions such that they tend to grow back even if they are completely excised surgically⁸. The warts, which at times are compounded and thus overlap each other, may grow to a size on the vocal folds that the airway becomes completely blocked⁸. In addition, the sound produced by a patient with Papillomas is hoarse as the degree of hoarseness is proportional to the size of the wart¹⁶.

CHAPTER 3

VOCAL FOLD MODELLING

3.1 Review of Vocal Fold Models

Modeling of vocal fold vibrations is useful for the study of both speech synthesis systems and specific aspects of the mechanics of vocal fold vibration¹⁷. To study the mechanics of vocal fold vibrations from a physiological or clinical perspective, mathematical models may be developed to simulate different conditions in order to understand and investigate the role of various abnormalities, of the vocal folds.

In the past two decades, there have been a number of models developed for simulating the dynamics of normal vocal folds as observed from vowel speech data¹⁸. Some of these models were developed to explain the process involved during the initiation of vocal fold vibration (vocal onset), and the period (in time) when the vibratory behavior is sustained¹⁸.

Prior to the development of models that examines the sustained region of phonation, the driving force, which sustains oscillation, needed to be described¹⁹. In 1958, Van den Berg performed a theoretical study on the relationship between the vocal fold tissue elasticity and aerodynamic forces produced by the Bernoulli Effect as it related to oscillation of the vocal folds²⁰. He found that the high velocity of air through the narrow opening of the glottis produced a backpressure that caused the vocal folds to close. He conjectured that the build up of pressure on the closed vocal folds eventually forces the closed vocal folds to open and induce sound waves⁹. This

release of air through the opened vocal folds results in a drop in pressure below the vocal folds, and consequently causes closure. This process is referred to as the Bernoulli Effect⁸.

However, later research showed that in addition to the Bernoulli Effect, the asymmetry characteristics of the vocal folds caused by the mucosal wave and inertial acoustic loading, provided by the body layer, created a time delay between the upper and lower portions of the vocal folds. The build up and release of supraglottal pressure then causes the glottal area to open and close⁹. Therefore, to better understand the mechanisms for vocal fold oscillations, it was necessary to include (i) asymmetry, (ii) subglottal pressure, (iii) viscoelastic properties of the vocal folds, and (iv) Bernoulli Effect¹⁹.

In 1968, Flanagan and Landgraf proposed a one mass model, subsequently

referred to as the FL68 model in this thesis. The FL68 is a single mass-spring oscillator that is driven by airflow from the lungs and adequately describes the body of the vocal folds consisting of the deep layer and thyroarytenoid muscle²².

Figure 6 illustrates this

model which describes the vocal fold behavior using the knowledge of the time delay

characteristic due to the mucosal wave and the Bernoulli Effect²². Since, the model has a single degree of freedom, it is not able to adequately describe the mucosal wave observed in the cover layer (i.e., epithelium, superficial layer, and intermediate layer)²². Indeed, accounting for the three tissue layers is necessary to describe flow-induced oscillation¹⁷. Thus, the displacement, which arises from the orientation of the elastin fibers cannot be used to describe sufficiently vocal fold abnormalities that are found in the cover layer.

In 1972, Ishizaka and Flanagan developed a two mass model (subsequently referred to as the IF72 model in this thesis) for vocal fold vibrations. The model has its underpinnings on an integrated knowledge of acoustics and laryngeal biomechanics²³. It was inspired by the theoretical work of Ishizaka and Matsudaira, which concluded that the natural vibrational pattern of the vocal folds require out-of-phase motions of the upper and lower surface of the folds²⁴. The IF72 two mass model is able to describe this upper and lower surface of the folds using two separate masses that were coupled by a spring. The model is able to reproduce oscillations with or without the presence of the vocal tract (the area between the glottis and the mouth), which reduces the complexity of the generated waveforms. Inclusion of the vocal tract includes frequency ranges that are not characteristic of the vocal fold regions and artifacts due to the effects of turbulence in the constricted regions of the vocal tract during phonation¹. In addition, the model contains sufficient degrees of freedom to describe the cover-body tissues necessary to simulate motion of the vocal folds in both the longitudinal and lateral directions.

Initially, the FL68 and IF72 model were developed to serve as a voice source for a speech synthesis system^{5,17}. It was later reported by Ishizaka and Isshiki that the IF72 model was valuable in the study of various pathologies of the vocal folds²⁵. The sound generation capabilities of such a model when coupled with a digital simulation of a transmission line analog of the vocal tract, generated sounds that were comparable to those of pathological voices from patients²⁶.

Researchers have developed more complex models, which attempt to simulate, more accurately, naturally sounding artificial voices and diverse vocal pathologies⁵. For example in 1973, Titze developed a 16-mass model, subsequently called T73, which realistically described the cover-body tissue properties when compared with the IF72 model¹⁷. The major objectives of the T73 model were to strengthen the correspondence of the parameters with the physiology to increase the naturalness of the sounds produced, and to allow phonation in at least two distinct registers¹⁸. The aforementioned models utilize discrete masses. Therefore, they approximate the sounds and vibratory patterns generated by the vocal folds. Titze and Strong developed the continuum model in 1975, subsequently called TS75¹⁸. The TS75 model was used to study the effects of laryngeal muscle contraction and other configurations of acoustic phonation¹⁸. Because of its complexity, the model has had limited use.

The most widely used model in simulation and synthesis studies of vocal fold vibration, which allows for a description of both the cover and body is the two-mass model⁵. This model has the advantage of being simple and tractable. Supporting studies have shown that, using this model, normal phonation can be well represented

by only two eigenmodes, and disordered voices, which are obtained through the variation of parameters in the model that have physiological correlates, by the three strongest modes²⁷. Thus, modes describing both normal and pathologic voices that can be captured with models that are more complex, can also be obtained in a sufficient manner using the two mass model.

3.2 Two Mass Model

The IF72 model consists of coupled masses, springs, and nonlinear dampers that are driven by the lung pressure²⁸. Steinecke et. al. (1995) proposed a new model

based upon IF72 as illustrated by Figure 7. The modified spring-dashpot model has masses that were connected to a rigid boundary that produced only lateral displacements²⁷. The two masses were directly

coupled to each other through another spring that accounted for shear forces. To simulate the effect of body layer, the lower mass was made thicker and more massive than the upper mass¹. These assumptions reduced the original IF72 model from ten coupled differential equations to two.

3.2.1 The Steinecke et. al. Model

The Steinecke et. al. (1995) model consists of a set of nonlinear second order differential equations²⁷.

$$m_1 \ddot{x}_1 + r_1 \dot{x}_1 + k_1 x_1 + \Theta(-a_1) x_1 \left(\frac{a_1}{2l} \right) + k_c (x_1 - x_2) = F_1 \quad (3.2.1a)$$

$$m_2 \ddot{x}_2 + r_2 \dot{x}_2 + k_2 x_2 + \Theta(-a_2) x_2 \left(\frac{a_2}{2l} \right) + k_c (x_2 - x_1) = F_2 \quad (3.2.1b)$$

The forces F_i acting on the masses is the driving force due to the pressure P_i acting on the corresponding part of the glottis, where $F_i = l d_i P_i$ ¹⁹. This pressure is due to the laminar airflow described by the following form of the Bernoulli equation,

$$P_s = P_1 + \frac{\rho}{2} \left(\frac{U}{a_1} \right)^2 = \frac{\rho}{2} \left(\frac{U}{a_{\min}} \right)^2 \quad (3.2.2)$$

where: $a_{\min}$ describes the smallest gap in the vertical direction, which the air steadily flows through, and is given by the following equation.

$$a_{\min} = \begin{cases} a_1, & \text{if } 0 < x_1 < x_2 \\ a_2, & \text{if } 0 < x_2 \leq x_1 \\ 0, & \text{otherwise} \end{cases} \quad (3.2.3)$$

The variables and parameters depicted in Figure 7 and Equation 3.2.1 are defined as follows:

$i = 1, 2$	1- lower mass, 2- upper mass	m_i	masses
$\dot{x}_i$	displacement of the masses	r_i	damping constant
x_{0i}	distance of the vocal folds from midline in the rest position	k_i	stiffness coefficient
		k_c	coupling stiffness between m_1 and m_2

c_i	additional stiffness		sides of masses m_i
	coefficient during	a_i	glottal area formed by the
	vocal fold collision		rest position and
l	length of the glottis		displacement of mass m_i
d_i	thickness of masses	a_{0i}	glottal rest area formed
	m_i		by the rest position of
P_i	pressure inside the		mass m_i
	glottis acting on	U	volume flow velocity of
	masses m_i		air
F_i	forces acting on the	ρ	density of air
	masses m_i	P_s	subglottal pressure
$\Theta(x_i)$	restoring force present		
	only during the		
	collision of both		

After air passes through the smallest gap, a jet builds up and the static pressure equals that of the supraglottal pressure (F_2), which may be set to zero¹⁹.

The following equation describes the pressure acting on the masses m_i .

$$P_1 = P_s \left[1 - \Theta(a_{\min}) \left(\frac{a_{\min}}{a_i} \right)^2 \right] \Theta(a_i) \quad (3.2.4a)$$

$$P_2 = 0 \quad (3.2.4b)$$

The following $\Theta(x_i)$ expression describes the restoring force due to the collision of the masses m_i .

$$\Theta(x) = \begin{cases} \tanh\left(\frac{50x}{x_0}\right), & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (3.2.5)$$

The tanh function approximates the nonlinear restoring force during collision. The resulting volume flow rate can be expressed by the following expression, using the given assumption that the supraglottal pressure is zero.

$$U = \sqrt{\frac{2P_s}{\rho}} a_{\min} \Theta(a_{\min}) \quad (3.2.6)$$

When the expression yields $U_{\min} > 0$, incomplete glottal closure occurred, thus producing a leakage of air and a breathy sounding voice³⁰.

3.2.2 Two Mass Model - Normal Case

Simulations of the vocal fold vibration were produced using the two mass model by Steinecke et. al. (1995). The MATLAB® software tool (Natick, MA) was then used to solve this model numerically using the 4th order Runge-Kutta method. Prior to developing the program in MATLAB®, the system of equations was converted into four first order differential equations (Equation 3.2.7).

$$\dot{x}_1 = v_1 \quad (3.2.7a)$$

$$\left[P_1 l d_1 - r_1 v_1 - k_1 x_1 - \Theta(-a_1) c_1 \left(\frac{a_1}{2l} \right) - k_c (x_1 - x_2) \right] \quad (3.2.7b)$$

$$\dot{x}_2 = v_2 \quad (3.2.7c)$$

$$\dot{v}_2 = \frac{1}{m_2} \left[-r_2 v_2 - k_2 x_2 - \Theta(-a_2) c_2 \left(\frac{a_2}{2l} \right) - k_c (x_2 - x_1) \right] \quad (3.2.7d)$$

The following are the constants used to model the behavior of normal vocal folds²⁷. All units are presented in centimeters, grams, and milliseconds and their corresponding combinations unless otherwise noted²⁷.

$$\begin{aligned}
m_1 &= 0.125, & m_2 &= 0.025 \\
r_1 &= 0.02, & r_2 &= 0.02 \\
k_1 &= 0.08, & k_2 &= 0.008 \\
c_1 &= 3k_1, & c_2 &= 3k_2 \\
d_1 &= 0.25, & d_2 &= 0.05 \\
a_{01} &= 0.05, & a_{02} &= 0.05 \\
k_c &= 0.025, \\
P_s &= 0.008 \text{ } (\approx \text{ cm } H_2O) \\
\rho &= 0.00113
\end{aligned}$$

Figures 8-11 show the results of the mass displacements along with the glottal volume flow velocity and the waveforms using Equation 3.2.1. This allow comparison with previously published results.

Figure 8 and Figure 9 illustrates the oscillatory motion of the masses with respect to time. An initial disturbance causes the masses to oscillate, and as time progresses, the oscillation of both mass 1 and mass 2 reach steady states. The sinusoidal motion reflects the opening and closing of the glottal areas. The peaks represents maximum glottal opening while the troughs represent maximum glottal closure. Correspondingly, Figure 10 and Figure 11 demonstrate the behavior of the glottal volume flow velocity as the folds are oscillating. When comparing Figure 8 with Figure 10, the time span for the maximum glottal opening similarly describes maximum flow velocity. The maximum flow velocity, which is expressed by the peaks in Figure 10, occurs during maximum glottal opening.

Figure 8. Simulation of the upper and lower mass displacements for a normal patient with a phase delay present²⁷.

Figure 9. Reproduction of the upper and lower mass displacements for a normal patient with a phase delay present using Steinecke's model. The lower mass displacement (x_1) is in the blue color while the upper mass displacement (x_2) is illustrated by the dashed magenta color.

Figure 10. Simulation of the glottal volume flow velocity for a normal patient with a phase delay present using Steinecke's model²⁷.

Figure 11. Reproduction of the upper glottal volume flow velocity for a normal patient with a phase delay present.

Initial conditions were determined to reproduce the results depicted in Figure 9 and Figure 11. A comparison of Figure 8 with Figure 9 (mass displacements) and Figure 10 with Figure 11 (glottal volume flow velocity) illustrate comparable waveforms in amplitudes and shape though the exact initial conditions were estimated.

To verify further that the model was implemented correctly, an energy balance equation was derived and solved numerically. This allows for a test of the validity of the model, independent of the initial conditions used, and thus provides a rigorous global check.

To perform the energy conservation test, the equations of motion described by Equation 3.2.1 were multiplied by a factor of $\dot{x}_i = v_i$, which then converted the equations into respective energy, dissipation and forcing terms. The energy terms describe the elastic potential energy $\left(\frac{1}{2}kx^2\right)$ and kinetic energy $\left(\frac{1}{2}mv^2\right)$ of the system. These balance the driving force from the airflow, the restoring force present during collision, and the effects of dissipation.

Mathematically, this is derived as follows.

$$\dot{x}_1 \left[m_1 \ddot{x}_1 + r_1 \dot{x}_1 + k_1 x_1 + \Theta(-a_1) x_1 \left(\frac{a_1}{2l} \right) + k_c (x_1 - x_2) = P_1 l d_1 \right] \quad (3.2.8a)$$

$$\dot{x}_2 \left[m_2 \ddot{x}_2 + r_2 \dot{x}_2 + k_2 x_2 + \Theta(-a_2) x_2 \left(\frac{a_2}{2l} \right) + k_c (x_2 - x_1) = 0 \right] \quad (3.2.8b)$$

A rearrangement of Equation 3.2.8 results in the following sets of equations.

$$\frac{d}{dt} \left[\frac{1}{2} m_1 \dot{x}_1^2 + \frac{1}{2} k_1 x_1^2 + \frac{1}{2} k_c x_1^2 - k_c x_2 x_1 \right] = P_1 l d_1 \dot{x}_1 - r_1 \dot{x}_1^2 - \Theta(-a_1) x_1 \left(\frac{a_1}{2l} \right) \dot{x}_1 \quad (3.2.9a)$$

$$\frac{d}{dt} \left[\frac{1}{2} m_2 \dot{x}_2^2 + \frac{1}{2} k_2 x_2^2 + \frac{1}{2} k_c x_2^2 - k_c x_1 x_2 \right] = -r_2 \dot{x}_2^2 - \Theta(-a_2) x_2 \left(\frac{a_2}{2l} \right) \dot{x}_2 \quad (3.2.9b)$$

An energy balance follows as the time rate of change of kinetic and potential energy equals the sum of dissipation and forcing effects, as is described by Equation 3.2.9.

Integrating both sides of Equation 3.2.9 provides a format more suitable for solving numerically using MATLAB®.

$$\frac{1}{2}m_1\dot{x}_1^2 + \frac{1}{2}k_1x_1^2 + \frac{1}{2}k_cx_1^2 - k_cx_2x_1 = \int P_1ld_1\dot{x}_1 - r_1\dot{x}_1^2 - \Theta(-a_1)c_1\left(\frac{a_1}{2l}\right)\dot{x}_1dt \quad (3.2.10a)$$

$$\frac{1}{2}m_2\dot{x}_2^2 + \frac{1}{2}k_2x_2^2 + \frac{1}{2}k_cx_2^2 - k_cx_1x_2 = \int -r_2\dot{x}_2^2 - \Theta(-a_2)c_2\left(\frac{a_2}{2l}\right)\dot{x}_2dt \quad (3.2.10b)$$

The left hand side of Equation 3.2.10 represents the kinetic and potential energy of the system, while the right hand side represents the dissipation due to the dampers, airflow dependent forcing effects, and nonlinear restoring force present during the collision of the vocal folds.

As an initial step, we obtained mass displacement and energy conservation plots for parameter sets for a reduced the system of equations which correspond with two-uncoupled Simple Harmonic Oscillators (SHO). These are obtained when the coupling coefficient (k_c) and nonlinearities resulting from dissipation and forcing terms are set to zero in Equation 3.2.10.

Additional terms were added to check the energy balance for various cases using Equation 3.2.10. Finally, the full two mass system was examined. Figure 12 shows the SHO case. The displacements and energies for mass 1 are shown in blue and for mass 2 in magenta.

Figure 12. SHO – (i) Without Dissipation and (ii) Without Forcing (Displacement).

The SHO model system oscillates continuously, as neither the dissipation nor forcing terms are included. This is reflected in Figure 12 by the periodicity of the displacements for x_1 and x_2 displacements.

Figure 13 shows the solution of Equation 3.2.10 with the coupling coefficient (k_c) present and the absence of dissipation and forcing terms.

The waveform of the two mass system (Figure 13) is periodic, although it appears more complex since it is not purely harmonic when compared with the SHO (Figure 12).

The energy calculations for the SHO (Figure 12) for both mass 1 and mass 2 shows that energy is conserved (Figure 14). The small scale oscillations in the energy waveforms are due to truncation and round off associated with solving the equations numerically. However, at $O(1)$ the fluctuations are not visible.

We plot the magnitude of the difference of the initial energy, E_0 , from its instantaneous value, E , divided by this initial energy value (absolute error) in the solver. Numerical test show that smaller tolerances reduce this quantity.

Similarly, Figure 15 shows that the two mass system with the nonlinear terms conserves energy.

Figure 16 shows the solution of Equation 3.2.10 for the SHO with the dissipation term set to zero subject to forcing.

Figure 16. SHO – (i) Without Dissipation and (ii) Forcing (Displacement).

In the case of an absent dissipation term with nonlinear forcing added to the SHO, the displacements of the two uncoupled masses is relatively small for mass 1 when compared to mass 2 (Figure 16). The forcing term used was extracted from Steinecke's model and describes the force due to the air pressure and the collision of the vocal folds. The displacement of mass 2 continued to increase as time progressed due to the lack of the coupling coefficient (k_c) during the interaction of the forcing terms affiliated with pressure and collision of the vocal folds.

Figure 17 shows the results of the two mass system in the absence of dissipation, but with the inclusion of the nonlinear forcing terms.

Figure 17. Two Mass System – (i) Without Dissipation and (ii) Forcing (Displacement).

The forcing case without dissipation for the two mass system shows that the behavior of mass 1 and mass 2 appears periodic, due to the addition of the coupling coefficient (k_c). The waveforms were also erratic in the sense that the duration of time present for the opening and closing phase of the vocal folds were not equal as shown in Figure 17. There was a shorter time duration for the closing phase, than the opening phase, and this could be related to the inclusion of the forcing terms.

Figure 18 shows the energy balance results for the SHO for the case without dissipation but with the addition of the forcing terms.

Figure 18. SHO – (i) Without Dissipation and (ii) Forcing (Energy).

Figure 19 shows the energy balance results for the nonlinear two mass system using the condition that excludes the dissipation term, but includes the forcing terms.

Figure 19. Two Mass System – (i) Without Dissipation and (ii) Forcing (Energy).

For the case for without dissipation, but with the addition of forcing as illustrated by Figure 18 and Figure 19, the energy model showed that energy was

conserved. Once again, the effect of the coupling coefficient (k_c) as a parameter that controls the motion of the vocal folds is illustrated by comparing the energy plots of Figure 18 and Figure 19. The SHO (Figure 18) lacks this coupling coefficient (k_c) while the two mass system (Figure 19) includes this parameter.

Figure 20 illustrates the motion of the masses with the inclusion of dissipation and the absence of forcing. In this case, the uncoupled masses will oscillate and motion will become completely damped out as time progresses.

This result is reflected in the phase plot (Figure 21) where the curves approach a fixed point as the system reaches equilibrium.

The presence of dissipation and lack of forcing was applied to the nonlinear coupled two mass system. Figure 22 shows that a similar behavior and result was obtained as that of the SHO displacement (Figure 20) for the same case. The masses for the two mass system ceased to oscillate after a sufficiently long time period.

Figure 23 shows the energy balance results for the SHO for the case of dissipation and the lack of forcing terms.

Figure 24 shows a similar energy balance results for the two mass system in the presence of dissipation and absence of nonlinear forcing terms as the SHO (Figure 23).

Figure 24. Two Mass System – (i) Dissipation and (ii) Without Forcing (Energy).

The behavior of the masses for the case that includes dissipation, but excludes the forcing terms as is depicted in Figure 23 and Figure 24, the displacement results shows a cessation of the mass motion which is reflected in the energy model.

Figure 25 shows the mass displacements results of the SHO with dissipation and the nonlinear forcing terms present. The motion of the uncoupled masses in the SHO displacement model (Figure 25) in the presence of dissipation and forcing shows the importance of the coupling term in the motion of the vocal folds.

Figure 26 shows the behavior of the masses for the two mass system in the presence of dissipation and the nonlinear forcing terms. Note the differences due to the nonlinearities compared to Figure 25.

The addition of the coupling term in the presence of dissipation and forcing regenerated the model described by Steinecke et. al. (1995) as seen in Figure 26. The dissipation term reduces the amplitude of the displacement following voice onset and the presence of forcing sustained the oscillation in time, following the reduction of amplitude.

Despite the absence of the coupling coefficient, our control, the SHO with the addition of dissipation and forcing conserved energy as is shown in Figure 27.

Similarly, Figure 28 shows that the two mass system also conserved energy when dissipation and nonlinear forcing terms are present in Equation 3.2.10.

Figure 28. Two Mass System – (i) Dissipation and (ii) Forcing (Energy).

Indeed, the comparison of our results for various combinations of springs, dampers, and forcing terms using the energy profiles for the SHO (control) and two mass model system, showed that the numerical solutions were accurate and an energy balance was achieved. With this knowledge, we were confident that we could modify the parameters of the model to produce results that will be useful in understanding the pathophysiology of certain vocal fold abnormalities in conjunction with a comparison with clinical data.

3.3 Extension of Two Mass Model - Parkinson Case

Zhang et. al. extended Steinecke et. al.'s two mass model to include high stiffness coefficients along with nonlinearities in the stiffness expressions to describe vocal fold abnormalities in Parkinson's disease patients³⁰. Reduced amplitude, incomplete vocal closure, increased phonation pressure, glottal tremor, and irregular vocal fold vibrations characterize Parkinson's disease³⁰.

These attributes are the resultant effects of abnormal rigidity in the musculature of the larynx. With the inclusion of additional terms to describe the stiffness coefficients and its corresponding nonlinearity components, the model had the following form.

$$m_1 \ddot{x}_1 + r_1 \dot{x}_1 + Kk_1 (x_1 + \eta_{1k} x_1^3) + \Theta(-a_1) Kc_1 \left[\frac{a_1}{2l} + \eta_{1h} \left(\frac{a_1}{2l} \right)^3 \right] + Kk_c [(x_1 - x_2) + \eta_{ck} (x_1 - x_2)^3] = ldP_1 \quad (3.3.1a)$$

$$m_2 \ddot{x}_2 + r_2 \dot{x}_2 + Kk_2 (x_2 + \eta_{2k} x_2^3) + \Theta(-a_2) Kc_2 \left[\frac{a_2}{2l} + \eta_{2h} \left(\frac{a_2}{2l} \right)^3 \right] + Kk_c [(x_2 - x_1) + \eta_{ck} (x_2 - x_1)^3] = 0 \quad (3.3.1b)$$

The variables and constants used in Equation 3.3.1 are the same as that of Equation 3.2.1. The additional parameters introduced in the extended model are as follows.

$i = 1, 2$	1- lower mass, 2- upper mass
η_{1k}	nonlinear spring coefficient
η_{ck}	nonlinear spring coefficient
η_{1h}	nonlinear collision coefficient
K	normalized stiffness coefficient which governs the dynamics of the vocal fold system
d	vocal fold thickness where $d = d_1 + d_2$

The typical parameter values for Equation 3.3.1 are $\eta_{1k} = 100$, $\eta_{ck} = 100$, $\eta_{1h} = 500$, and $K = 3.2^{30}$.

The published results of Zhang et. al.'s extended two mass model are shown in Figure 29, in which we reproduce in Figure 30³⁰. The additional parameter values

are $r_1 = r_2 = 0.01$, $a_0 = 0.025$, $k_c = 0.04$, and $P_s = 0.012$. All units are presented in centimeters, grams, and milliseconds and their corresponding combinations³⁰.

Figure 30 displays our results following the use of Parkinson's model (Equation 3.3.1) developed by Zhang et. al.

Figure 29 and Figure 30 correspond with voice irregularities inherent in voice disorders of Parkinson's disease patients. Despite the use of approximate initial conditions in Zhang's model, the amplitude of the Parkinson waveform is similar and the aperiodicity characteristic of increased stiffness is reflected in our model.

As with the Steinecke et. al model, energy conservation equations were derived for the Zhang model and similar numerical tests using this model in limiting cases were conducted. The results showed that energy was conserved and an increase in the tolerances improved the accuracy of the solution.

CHAPTER 4

ACQUISITION OF CLINICAL DATA

4.1 Introduction to the Acquisition of Voice Samples

A key diagnostic feature for further understanding of abnormalities in voice functioning is an evaluation of vocal fold vibration patterns. These patterns are called waveforms¹⁷.

Voice samples were collected from female patients and volunteers between the ages of 19 and 83 from the Voice Clinic at University of Wisconsin - Madison Division of Otolaryngology by Dr. Diane Bless and coworkers. These were provided to Dr. Yan as part of their collaborations. The adult female voice is considered a better standard to use in experimentation than the male adult voice because the average fundamental frequency for females range between 200-250 Hz¹. On the contrary, the fundamental frequency of the adult male voice exhibits a nonlinear behavior in the span of four decades following the end of puberty. Figure 31 illustrates this event. Therefore, the analysis proposed and conclusions in this thesis are only applicable to female voices and not for male voices⁸.

During an examination, the female patients phonate the vowel /i/ at npnl (normal pitch normal loudness) corresponding to a fundamental frequency of approximately 250 Hz for two seconds. Vowels are usually used for phonation exercises because vowels cause the vocal folds to vibrate and convert the steady streams of air from the lungs into periodic pulses⁹. These pulses, also called glottal flow, correlate with the oscillatory behavior of the vocal folds. For this reason, the vowel /i/ at npnl was used to acquire waveforms, which characterize the vocal fold behavior. We describe some prototype techniques used for these examination and data collection.

4.1.1 Glottal Area Waveform (GAW)

The High Speed Laryngeal Imaging (HSLI) involves using an endoscope to obtain video recordings and images of the vocal folds at a sampling rate of 2000 frames per second³¹. HSLI also provides a visual representation of the vocal folds. The visual representation of the vocal folds enables the clinician to see abnormalities such as cysts and/or warts (Papillomas). Other disorders, which may not be visually apparent, such as abnormal oscillation features, may also be captured using this technique. Data, such as the glottal area (GA), that is, the area between the two vocal folds are extracted and measured using segmentation techniques developed by Yan et. al.^{31,32}. For given periods, the GA is then reconstructed to produce the glottal area waveform (GAW). Typical vibrating folds are given in Figure 32 below.

Figure 32. Vibrating vocal folds for 24 frames (0.012 sec).

Comparison of this waveform obtained from patient data with the two mass model will be discussed in Chapter 5.

The HSLI technique is a direct method for obtaining the frequency ranges of the vocal folds (0-2000 Hz) with minimal artifacts due to frequencies that are include

from the result of vocal tract effects. Unfortunately, the HSLI method is quite invasive. This is because images are captured by first inserting an endoscope, with a camera attached, into the patient's throat. Most patients often gag as a response to the endoscope making the examination difficult³³. In addition, it is quite difficult to obtain a two-second recording of the images, where the vocal folds are centered in the frame. Often times, the endoscope moves slightly or the patient moves during examination, which causes shifts of the images that are captured. This makes the acquisition of the GA difficult since parts of the area may not be in the frame. Thus, the time sequence used in the signal analysis is dependent upon the length of useful frames. Consequently, the results displayed in this chapter and subsequent chapters may have varying time spans.

Figure 33 shows the GAW obtained from a normal 36-year-old female patient.. The waveform was generated by using the GA captured from the HSLI method and examining the changes of area over a time interval.

Figure 33. Example of GAW for a normal 36-year-old female patient.

4.1.2 Acoustic Waveform (AW)

Another technique for collecting vocal fold data is the acoustic method. A microphone is used to capture sounds produced by the larynx. The sampling rate used in the Kay high-speed system that incorporates simultaneous acquisition of the GA was 50,000 Hz for the acoustic signals³¹. This sampling rate is adequate to capture and reproduce the sound captured by microphone when the patient phonates.

The disadvantage of using the acoustic system is that it is an indirect technique of capturing the frequency ranges of the vocal folds³⁴. Factors such as turbulence in the air stream of a constricted region of the vocal tract at times affect the quality of signals. Our results show that the acoustic approach produces waveforms, which exhibit additional frequency components that are not in the vocal fold range, but may be of use in characterization. Although preprocessing methods exist to reduce, if not eliminate the additional frequencies, sometimes, vital frequencies that may be useful for the characterization of the vocal fold pathology are lost. For this reason, in this thesis, we will attempt to validate GAW data and compare our results with acoustic waveforms. Figure 34 shows the acoustic waveform for the same patient whose GAW is illustrated in Figure 33.

Figure 34. Example of Acoustic Waveforms (AW) for a normal 36-year-old female patient.

4.1.3 Other Acquisition Methods

Laryngeal videostroboscopy (LVS) is another technique that can be used to observe vocal fold vibrations³. It uses a strobe light flashing off-phase at integer multiples of the actual vibratory cycle to observe the motion of the vocal folds. A disadvantage of this technique is its inability to resolve the actual vocal fold vibration so the accuracy of the waveforms, especially pathological waveforms that are aperiodic, is low. However, this method is still widely used by clinicians to diagnose voice disorders³.

Electroglottography (EGG) is a device that produces low-intensity radiofrequency current that flows between a pair of electrodes³⁵. These electrodes are placed against the neck. As the vocal folds oscillate, the neck impedances (electrical resistance) are captured by the electrodes and form the EGG. Since EGG is an

external device, it provides only an approximate value of the glottis surface and some information on the period of vocal fold vibrations with respect to the vocal fold contact area³⁴.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Signal Processing Methods

In this Chapter, we describe the techniques used to analyze the results from the numerical simulation of the two mass model and the clinical data. Each clinical dataset (acoustic and GAW), contained two second sample signals of the patients phonating the vowel /i/. We use programs that were written in MATLAB® to analyze various data. Multi-spectral techniques, such as the Fast Fourier Transforms (FFT), Short Time Fourier Transform (STFT), and the Wavelet Transform (WT) were used to isolate unique characteristics from the ensuing signals. First, we review and outline these methods.

5.1.1 Fast Fourier Transform (FFT)

The FFT was used to transform discrete signals into sums of sinusoids of varying frequencies, amplitudes, and phases³⁶. The FFT algorithm uses the following pair of discrete Fourier transform equations.

$$X(\omega_k) = \sum_{n=0}^{N-1} x(t_n) e^{-j\omega_k t_n}, \quad k = 0, 1, 2, \dots, N-1 \quad (5.1.1)$$

$$x(t_n) = \frac{1}{N} \sum_{k=0}^{N-1} X(\omega_k) e^{j\omega_k t_n}, \quad n = 0, 1, 2, \dots, N-1 \quad (5.1.2)$$

In Equations 5.1.1 and 5.1.2, $x(t_n)$ is the discrete time signal, $X(\omega_k)$ describes the frequency components in the time signal, ω_k is the frequency, and N is the length of the signal sequence. Equation 5.1.1 allows a transform from the time to the frequency domain to be possible, while Equation 5.1.2 allows the inverse to be true.

Computing the inner product of the signal with complex exponential results in the decomposition of the signal into these series of sinusoids with varying magnitudes, phases, and frequencies (Equation 5.1.1)³⁶. The results of the FFT are displayed as amplitude versus frequency plot, such as given in Figure 35 for the corresponding sinusoidal waveform with frequency components at 50 Hz and 150 Hz (shown in Figure 36).

Figure 35. FFT example using a sine curve with a frequency component at 50 Hz and 150 Hz.

Figure 36. Waveform using a sine curve with frequency components at 50 Hz and 150 Hz.

5.1.2 Short Time Fourier Transform (Spectrogram)

In contrast to FFT, the STFT is used to reveal both the time and frequency information from the signal³⁷. This is accomplished by applying a fixed window to the signal of interest. The contents of the windows are finite-length portions of the given signal³⁶. The spectral characteristics of each window are analyzed over a given interval. Upon completion, the window is shifted and that section is then analyzed for its spectral content.

For nonstationary signals such as those encountered during acoustic analysis, the STFT technique seemed to work very well. These signals can be approximated as stationary over small time intervals, and can yield good spectrogram results³⁶. Indeed, for signals whose spectral characteristics vary rapidly, the choice of small window length is especially important because it provides a better time resolution and wide-band spectrogram at the expense of frequency resolution. Conversely, a wider window could be applied to a signal that does not have much variation in the spectral

characteristic. This leads to an increase in frequency resolution and a narrow-band spectrogram, at the expense of time-resolution³⁶.

There are many possible windows for analysis. The criteria for selecting the window are described by Mitra, and are outlined as below³⁶.

1. The main lobe width of the window is used to determine the ability of the window to resolve two signal components near each other.
2. The relative side lobe amplitude is used to control the degree of leakage of one component into a nearby signal component.

Since the signals we collected varied in time, it was necessary to choose a window that allows for small relative side lobe amplitude. Thus, wide-band spectrograms were used to estimate the changes in the vocal tract. The Hanning window was most effective for our purposes since it allowed for better frequency resolution than some of the other windows and provided reduced spectral leakage^{38,39}. Typically, the Hanning window is a good first choice for signals with spectral features that are unknown³⁸. These characteristic features of the Hanning window motivated the use of this window in the development of diagnostic techniques using spectral methods.

A Hanning window of length 512 was chosen as to allow the width to be as close to several pitch periods as possible³⁶. This was necessary to determine the formants and estimate the frequency spectrum. To establish the harmonics of the pitch frequencies, a narrow-band spectrogram with a window size of several pitch periods was necessary.

The STFT is described by the following equation.

$$X(\omega_k, t_n) = \sum_{m=0}^{N-1} x(t_n) w(m - t_n) e^{-j\omega_k m}, \quad k, n = 0, 1, 2, \dots, N \quad (5.1.3)$$

The expressions present in Equation 5.1.3 are as follows: $x(t_n)$ is the discrete time signal, $w(m - t_n)$ is the Hanning window, ω_k is the frequency, and N is the length of the signal sequence. When $w(m - t_n) = 1$, the STFT reduces to that of the FFT described previously by Equation 5.1.1.

The results of the STFT are displayed as a spectrogram, which is the magnitude of the squared modulus of the STFT function.

$$Spectrogram = |X(\omega_k, t_n)|^2 = \left| \sum_{m=0}^{N-1} x(t_n) w(m - t_n) e^{-j\omega_k m} \right|^2 \quad (5.1.4)$$

Figure 37 displays a schematic of a spectrogram for a signal composed of three different frequencies. The horizontal axis of the spectrogram represents time, while the vertical axis represents frequency. The third dimension represents the magnitude of the basis elements describing the local spectral content over a specified time and frequency ranges described by the dimensions of the window.

Figure 37 illustrates the idea of the spectrogram. The spectrogram uses a linearly tiled scheme⁶. The height and width of a tile represent the spectral and temporal widths of the basis element, respectively⁴¹. Thus, a change in the window width, as is illustrated by the blocks with different shading patterns, will increase or decrease the number of cycles/frequency components that are visualized inside the window. Low frequency components will not be well represented by the use of short width windows, which are used when a good time resolution is necessary, such as analyzing short pulses or signals with high frequency content. Conversely, short pulse/high frequency content is poorly localized in time using long width windows. As a result, the choice of a window length resulting in the optimal time or frequency

resolution depends upon the type of information that is needed for interpretation of the signal.

Figure 38 illustrates the use of the spectrogram on the waveform of Figure 36. A window length of 512 was chosen which provides good frequency resolution at the expense of good time resolution.

Figure 38. Spectrogram example using a sine curve with frequency components at 50 Hz and 150 Hz.

The results of the spectrogram provide the same information as that of the FFT (Figure 35) and is shown here as proof of concept. Frequency components at 50 Hz and 150 Hz are still displayed and are represented by the red color band corresponding to the largest magnitude.

The FFT and STFT decompose time signals into basis functions composed of sinusoidals. However, transient signals are difficult to detect using these methods⁴². Therefore, we use Wavelet Transform (WT) techniques for further analysis of the data. The beauty of the WT technique is that it provides basis functions that are better

suited for emphasizing transients in signals, which could be indicative of a vocal pathological condition³⁷. In addition, the WT helps to resolve the problem of the fixed resolution encountered from the use of the spectrogram. We briefly describe the technique.

5.1.3 Wavelets

We used the continuous wavelet transform (CWT) to analyze the two mass model and clinical data. The CWT pair is given by the following equations⁴³.

$$CWT(a, \tau) = \frac{1}{\sqrt{a}} \int x(t) \psi\left(\frac{t - \tau}{a}\right) dt \quad (5.1.5)$$

$$x(t) = \frac{1}{c_\psi} \int_{-\infty}^{\infty} \int_{a>0} CWT(a, \tau) \frac{1}{\sqrt{a}} \psi\left(\frac{t - \tau}{a}\right) \frac{1}{a^2} da d\tau \quad (5.1.6)$$

where $x(t)$ is the discrete time signal, $\psi(t)$ is the wavelet function, a is the scale, τ describes the translation/shift of the wavelet, and c_ψ is a constant, which depends only on $\psi(t)$.

We implemented the CWT method by computing the cross-correlation of the signal and the wavelet function ψ and shifting the wavelet across the signal. The correlation between the components of the signal and the wavelet were evaluated and the results were presented as coefficient values. Once this convolution process was completed, the wavelet was expanded and this expanded form corresponded to a different scale value. The cross-correlation between the expanded wavelet and the

discrete time signal was computed again and coefficients describing how well the wavelet correlated with the time signal were obtained.

Using the coefficients, pseudo-spectral information from the various scales, which are indirectly related to a particular frequency bands, were obtained. Low frequencies on the dependent axis corresponded to high scale values, and conversely, high frequencies correspond to low scale values. Figure 39 illustrates the idea of the wavelet coefficient plot, which uses a logarithmic tiling scheme⁶.

Once again, the height and width of a tile represent the spectral and temporal widths of the basis element, respectively⁴¹. A change in the window width represents a dilation or compression of the wavelet function⁴³. This results in a different scaled

pseudo-frequency and period for the wavelet function. The number of cycles/frequencies that are seen inside the window remains constant, unlike the STFT method (Figure 37).

The choice of the wavelet function $\psi(t)$ depended upon the time signal used as input and the characteristics of the signals that were anticipated. Daubechies wavelet number 9 (db9) was chosen because it characterized high frequency components in the signals better than Daubechies wavelet number 4 (db4) due to its composition. A comparison of db9 and db4 are given in Figures 40 and 41. It is observed that db9 provided energy bands, which correspond to good correlation between the scaled wavelet and the signal, with less “bleed through” or leakage effects as compared with db4. Therefore, db9 was used as our standard wavelet function, $\psi(t)$.

The multi-spectral tools aforementioned were used to obtain characteristic information about the signal in the transform (scale) domain, which was otherwise hidden when examining the signal in the time domain or frequency domain. We used MATLAB® package to perform the FFT, STFT, and Wavelet data analysis.

5.2 Spectral Analysis of Normal and Pathological Vocal Folds

The FFT, Spectrogram, and Wavelet methods were applied to both normal and pathological voices, which were obtained from female the patients. The waveforms were then compared to the two mass model described in Chapter 3. The normal, Parkinson, Papillomas, and Diplophonic cases were examined using both GAW and acoustic waveforms. The unique characteristics pertinent to the GAW describing the particular pathology were then extracted. Acoustic data were examined in the normal, Papillomas, and Diplophonic case since the data were

available to determine whether similar characteristics present in the GAW were visible, despite the addition of other frequency components that could occur due to vocal tract complications.

5.2.1 Waveforms

We present below the results for GAW data obtained from both the two mass model and clinical data. Acoustic signals and laryngeal images were simultaneously captured using the HSLI system from Kay Elementrics (Lincoln Park, NJ), each data set recorded for two-second durations. The GAW was derived from the HSLI data using the analysis system developed by Yan et. al.^{31,32}. GAW was also generated using the two mass model by Steinecke et. al. using the following set of equations²⁷.

$$a_1 = a_{01} + 2lx_1 \quad (5.2.1a)$$

$$a_2 = a_{02} + 2lx_2 \quad (5.2.1b)$$

x_1 and x_2 in Equation 5.2.1 are governed by Equation 3.2.1.

The GAW depicts the opening and closing mechanism of the vocal folds. For clarity in this exposition, only 50 ms of data will be presented. Both GAW obtained clinically and through the model was normalized to allow comparisons to be possible.

Figure 42 shows the acoustic waveform for a normal patient. Acoustic waveforms contain other frequency components, which are not characteristic of the glottal area region. As a result, acoustic waveforms appear more complicated when compared with the GAW for the same patient, such as shown in Figure 43.

Figure 44 displays the GAW for the normal voice of Figure 43 superimposed on the GAW obtained from the two mass model (magenta color).

The two waveforms, although not identical due to the lack of overlap between the crescents and troughs, appear to have similar frequencies. The lack of overlap, could be the result of the data acquisition tool used in extracting the GAW using the HSLI method. Possible patient and endoscope movements during data acquisition have invoked spurious data and noise and leads to the entire glottal area not being fully captured in the frame³¹. However, a more likely explanation is that the model parameters did not completely capture the true behavior of the vocal folds for the normal patient.

The two mass model allowed the GAW to be generated using only mass 1, only mass 2, or combined. Only mass 1 provided the best correlation for the GAW generated from clinical data. This could be because the clinical data obtained from

the HSLI do not exhibit significant mucosal wave behavior, which is aided by the addition of mass 2. Another possible reason may be due to the model being discrete and hence only serves as an approximation to the GAW obtained clinically. For this study, we assume mass 1 is a sufficient approximation.

Figure 45 shows a typical GAW for a Parkinson's disease patient.

Figure 45. Parkinson Patient - GAW.

Two characteristics of Parkinson are reduced amplitude and incomplete vocal closure. The reduction of amplitude is evident by comparing the amplitude of Figure 43 with Figure 45. Figure 45 shows a reduction of amplitude, above zero, of 0.2 units of the normalized glottal area. In addition, an incomplete vocal closure is reminiscent of flat troughs, such as observed in Figure 45, thus causing a leakage of air and the breathy characteristic of Parkinson³⁰.

Using the Steinecke et. al. model, model results are illustrated in magenta (Figure 46), the Parkinson case was reproduced by decreasing the mass ($\sim 60\%$), increasing the damping and stiffness (2-fold), and a reduction in pressure ($\sim 10\%$).

Figure 46. Parkinson Voice: Model Data (Magenta) vs. Clinical Data (Blue) - GAW.

The acoustic waveform for a patient with Papillomas disease (recurrent warts) is shown in Figure 47. Here, the characteristics pertinent to this pathology are difficult to interpret due to the presence of other frequencies that are not easily discernable.

Figure 47. Papillomas Patient (40 years old) - Acoustic Waveform.

The GAW for Papillomas patient (Figure 48) has a similar resemblance to the normal GAW shown in Figure 43. Distinguishing between the two through the waveforms is difficult. However, the FFT, Spectrogram, and Wavelet may provide additional information reminiscent of such signals. We present this analysis in Section 5.2.2-5.2.4.

Figure 48. Papillomas Patient (40 years old) - GAW.

Figure 49 compares the waveforms between the clinical data and model data (magenta) for a Papillomas patient.

Using the data described by the model in magenta (Figure 49), the Papillomas case was reproduced by decreasing the stiffness coefficient ($\sim 30\%$). The mass and pressure was also increased by ($\sim 22\%$ and $\sim 30\%$, respectively). We adjusted our parameters based upon work by Zhang et. al regarding polyps, which have similar characteristics to Papillomas⁴⁴. Our results show a good agreement between the model and the clinical data.

Figure 50 shows the acoustic waveform of a Diplophonic patient. The waveform appears very complex due to the multiple frequencies that are characteristic of this disorder.

Figure 50. Diplophonic female Patient (39 years old) - Acoustic Waveform (Pre-treatment).

Figure 51 shows the acoustic waveform of the same patient as Figure 50 following treatment of the Diplophonic condition. The acoustic waveform appears to resemble that of the normal GAW (Figure 43). This is an indicator that treatment did improve the voice; however, some additional frequencies remain. The improved Diplophonic condition is confirmed later on this Chapter when FFT and Spectrogram techniques are used to analyze the data.

Figure 51. Diplophonic female Patient (39 years old) - Acoustic Waveform (Post-treatment).

The contributions of frequencies outside the vocal fold range are minimal during the acquisition of the GAW using the more direct HSLI method. As a result, a less complicated description of the vocal fold oscillatory behavior for Diplophonic patients was obtained (Figure 52).

Figure 52. Diplophonic Patient – GAW (Pre-treatment).

The GAW post treatment (magenta color) for the Diplophonic patient shows that treatment appears successful at first inspection of Figure 53, which displays the result of an overlap with the normal GAW (blue color). These two waveforms appear to have similar shapes and characteristics. The FFT and Spectrogram can be used to validate this result and quantify these changes.

Using the data described by the two mass model, Figure 54 shows the Diplophonic case which was reproduced by increasing the stiffness resulting from the collision of the vocal folds (3-fold), increasing the thickness of the masses (20-fold), decreasing the glottal area (10-fold), and increasing the pressure (10-fold).

There appears to be good agreement between the model and clinical data. Analysis of both waveforms using spectral method (FFT, STFT, and Wavelets) will determine the correlation between the two and the power of these methods as diagnostic tools. We start with the FFT in the following section, which reveals the frequency components present in the GAW and acoustic signals.

5.2.2 FFT

Figure 55 shows the FFT for the normal acoustic patient. Three harmonics are visibly present with the primary harmonic stationed at 250 Hz. The frequency corresponds to the female patient who is phonating the vowel /i/ at npnl.

Figure 55. Normal Patient – Acoustic Waveform (FFT).

The FFT applied to the normal GAW (Figure 56) reveals the primary harmonic located at 250 Hz, as expected. The secondary harmonic also exists, even though it is not as prominent as that of the acoustic FFT result.

Figure 56. Normal Patient – GAW (FFT).

The FFT result using the waveform generated by the two mass model for the normal voice (Figure 57) only reveals the primary harmonic at 250 Hz. Multiple harmonic components observed in the clinical and model data may be the result of nonlinearities present in the motion of the vocal folds³¹. In the case of the model, perhaps a lack of higher order harmonics may be due to insufficient or minimal energy generated by the driving force and/or collision force. The inverse may be true if higher order harmonics are captured by the FFT plots.

Figure 57. Normal Voice (Model) – GAW (FFT).

The FFT for the GAW Parkinson data reveals frequencies that are approximately 200 Hz and 400 Hz as displayed in Figure 58. There is also a frequency component at approximately 8 Hz. This is characteristic of Parkinson and caused by the glottal tremor¹³. The bimodal behavior of the FFT spectral components in Figure 58 may be due to the vocal tremor effects characteristic of Parkinson.

Figure 59 shows the FFT result using the Parkinson model. The frequency components are present at 200 Hz, 400 Hz, and 600 Hz similar to the FFT result for the clinical data (Figure 58). Note that there are some other harmonics present in the data possibly due to the lack of a complete correlation of the peaks and troughs using the model data (Figure 46).

The FFT result from the acoustic Papillomas data reveals frequency components at 250 Hz and 500 Hz (Figure 60). This is expected since the acoustic waveform for Papillomas (Figure 47) is similar to that of the normal acoustic waveform (Figure 42).

The FFT of the Papillomas GAW (Figure 61) shows that the frequency components reside at 200 Hz and 400 Hz and not at 250 Hz and 500 Hz, as expected, for the normal GAW FFT (Figure 56). Other frequency components present in the Papillomas acoustic signal may have shifted the average frequency content of the signal analyzed to 250 Hz and 500 Hz.

Figure 61. Papillomas Patient (40 years old) – GAW (FFT).

The FFT result for Papillomas using the model (Figure 62) matched that of the clinical data (Figure 61). One peak is present at approximately 200 Hz. The amplitude of the peak was approximately eight units higher using the model. This result supports the idea that waveform obtained using the model adequately captured the characteristics that are unique to Papillomas.

The FFT result for the Diplophonic patients using acoustic data reveals multiple frequencies that are characteristic to this disorder (Figure 63)¹⁰. The primary harmonic at 250 Hz is still present along with other frequencies components that are higher and lower than this harmonic.

Following treatment, the FFT result for the Diplophonic patient using acoustic data still reveals the multiple frequencies (Figure 64). However, several frequency components have been eliminated and/or reduced in amplitude such as that located between 300 - 350 Hz. The primary harmonic is still present as is expected at 250 Hz. Although, lower frequencies are present in the patient's voice, the FFT visually confirms that treatment was successful.

Figure 64. Diplophonic Patient - Acoustic Waveform (FFT: Post-treatment).

The FFT result using the GAW (Figure 65) from the Diplophonic patient reveals multiple frequencies that are also observed for the acoustic FFT results using the pre-treatment case (Figure 63). The primary harmonic at 250 Hz was still visible although its amplitude was small. There are other frequency components that are below the 200 Hz mark, but are so small in amplitude and are hard to detect. As a result, the amplitude of the primary peak was scaled down as to reflect the smaller frequency components that are present.

Figure 65. Diplophonic Patient – GAW (FFT: Pre-treatment).

Following treatment, the FFT result for the Diplophonic patient using the GAW reveals only two frequency components (Figure 66). One occurs at 200 Hz and the other occurs at 400 Hz. The fact that the primary frequency component is now located at 200 Hz rather than 250 Hz could be due to side effects emanating from treatment.

Figure 66. Diplophonic Patient – GAW (FFT: Post-treatment).

The FFT result for the Diplophonic patient using the two mass model (Figure 67) reveals similar locations for the frequency components when compared with the clinical data shown in Figure 65.

Figure 67. Diplophonic Voice (Model) – GAW (FFT: Pre-treatment).

As parameters of the model were modified and the fit of the model waveform with the clinical waveform became more accurate, we noticed that the FFT results started to show good agreement. However, since the model is only an approximation, only the major spectral features were captured and not minor features such as the frequencies present in the tails of Figure 65 that adjoin the peaks.

5.2.3 Spectrogram

The following figures represent the results of the STFT, which are displayed as a spectrogram. This method allows information regarding the changes of the frequency components in time to be visualized.

The varying colors represent the intensity of those frequencies ranging from low to high. The frequency bands highlighted in red should be the focus when interpreting the spectrogram. The changes in the frequency components depicted in the FFT in time will be captured and reflected by using this type of plot for the acoustic data only. The spectrogram using the GAW did not provide any new information that was not already obtained using the FFT for the clinical data and model data. The GAW spectrogram results showed no changes of the frequencies in time over the time span being analyzed.

The spectrogram using the normal acoustic patient data (Figure 68) features three main bands as the FFT analysis revealed three major peaks (Figure 55). They were located at 200 Hz, 500 Hz, and 750 Hz and did not over the time interval. Other harmonics are also present, but they were not as prominent as the three peaks at 200 Hz, 500 Hz, and 750 Hz. These harmonics occurred at high frequencies and varied in time. For example, it is evident that a frequency of 1580 Hz does not appear for the entire two-second duration. Rather, this frequency band is faint in the beginning and becomes prominent after one second. It is at this point that the intensity or amplitude increases (see the side bar index).

Similarly, the spectrogram using the GAW (Figure 69) reveals the frequency components at approximately 250 Hz, 500 Hz, and 750 Hz. The bands are continuous across the selected periods that were analyzed hence do not show any variations in time. This continuous band behavior is present in all the GAW spectrogram results for the Diplophonic, Parkinson, and Papillomas disease. In this case, the use of the spectrogram as a diagnostic tool for the GAW does not provide any additional information that will aide in the characterization.

Figure 69. Normal Patient – GAW (Spectrogram).

The spectrogram using the acoustic Papillomas waveform (Figure 70) reveals frequency components at approximately 250 Hz and 500 Hz. It is interesting to note with this figure that at approximately 0.2 seconds, the patient's voice at the 500 Hz band reduces in pitch by ~30 Hz before increasing again. In addition, other higher order harmonics can be seen at intermittent intervals. The red-orange bands that appear and disappear as time progresses reflect this phenomenon.

Figure 70. Papillomas Patient (40 years old) - Acoustic Waveform.(Spectrogram).

The spectrogram results for the Diplophonic patient using the acoustic waveform (Figure 71) reveals the multiple frequency components that are characteristic of this disorder as is confirmed also by the FFT (Figure 63)¹⁰. It is interesting to note that for Diplophonia, not all these frequencies occur simultaneously or at the same amplitude in time. Low and high frequency components alike have intermittent traits. The behavior of these frequencies in time was not captured by the FFT and hence this example highlights the benefits of the spectrogram as a diagnostic tool.

Following treatment, the number of frequency components from the acoustic Diplophonic data has reduced (Figure 72). However, the intermittent behavior of the frequencies is still apparent. Nevertheless, the spectrogram does confirm the results of the acoustic waveform and FFT result that treatment did improve the patient's condition.

The spectrogram results using the acoustic waveforms provided valuable information describing the behavior of the frequency components in time. These results showed the importance of acoustic analysis in the development of characteristic information regarding the pathology.

The spectrogram results for the GAW did not provide any additional information due to the poor resolution in time, which did not show any changes for the given time span. As an alternative, wavelets may be used to obtain some characteristic information that may be a useful for diagnostic purposes. The results of the wavelet analysis transform the signal into a scale versus time representation, rather than a frequency versus time representation. This may display features present in the signal that may otherwise be missed using the spectrogram technique. We outline the results of wavelets in the following section.

5.2.4 Wavelets

The following figures show the results of the wavelets, which are displayed as a wavelet coefficient plot or a scalogram. The varying colors represent how well the wavelet, when scaled and translated, correlated with the time signal. High intensity (energy) colors (reddish) represent a strong correlation between the wavelet and the acoustic signal or GAW. The resolution of the scalogram was better than the spectrogram for the GAW and as a result, fluctuations of the pseudo frequency (scale) in time could be generated. The scales representing the highest energy/strongest

correlation were noted, along with the bandwidth of the energies, and the texture of the plot due to the improved resolution.

The wavelet coefficient plot for the normal acoustic data showed the highest energy location had a bandwidth that was between the scales of approximately 2 and 22 (Figure 73). The mean bandwidth was located at scale 12. There appear to be to separate bands though within the overall scale range of 2-22. They can be differentiated by their different colors. Scale 2-10 feature coefficient plots that did not correlate well with the Clinical data due the coloration deviating from the red-burgundy color. The second energy band is between the scales ranging from 12-22. These coefficient values correlated well with the acoustic data. The texture of the energy band appears rippled with a bowing effect.

The coefficient plot of the normal GAW data shows a distinct energy band between the scales of 4 and 8 (Figure 74). This energy band displays fluctuations with a repeating pattern of alternating blue and red intensity colors. This may be related to the crescents and trough overlap of the wavelet on the signal.

Figure 74. Normal Patient – GAW (Wavelet).

The wavelet coefficient plot of the normal voice using the two mass model contains an energy band between scales 4-8 (Figure 75). The pattern shown is repeating similar to Figure 74 which uses the GAW. There appears to be a secondary band at a scale of approximately 9.

The coefficient plot using the GAW data from the Parkinson patient reveals the highest energy concentrations being placed between scales 5 and 12 (Figure 76). In addition, these energies are intermittent and increase/decrease in intensity following every 100 samples.

The most visible energy band for the Parkinson voice using the model data spans between scales 5-13 as is seen in Figure 77. There is a weaker band that spans a scale ranging from 2-6. In addition, scale 14-32 features yellow intermittent bands that are more pronounced using the model data than using the data describing the GAW obtained from the HSLI method (Figure 76).

Figure 77. Parkinson Voice (Model) – GAW (Wavelet).

The wavelet coefficient plot for the acoustic Papillomas waveform (Figure 78) reflects a similar behavior as the spectrogram for the acoustic case (Figure 70). At approximately 0.2 seconds, which corresponds to approximately 1100 samples, we noted that the spectrogram bands reduced in frequency by ~30 Hz before increasing again. In the wavelet coefficient plot, this is reflected by an increase in scale and then a reduction, recall that high scale implies low frequency.

The energy band for the scalogram spans the scale from 5 to 27. There was a strong correlation between the wavelet and the acoustic Papillomas signal in this range with the exception of scale 12, which showed moderate correlation.

The wavelet coefficient plot using the GAW Papillomas data contained the strongest energy with a frequency bandwidth ranging from approximately 5-10 (Figure 79). This bandwidth is larger than the normal GAW scalogram case and can be a feature that will aid in the characterization.

The wavelet coefficient plot using the two mass model to describe Papillomas disease (Figure 80) was very similar to that of the clinical GAW data that was depicted in Figure 79. These results should not be surprising considering the waveform, FFT, and Spectrogram result between the model and clinical data for this pathology all correlated well.

Figure 80. Papillomas Voice (Model) – GAW (Wavelet).

The acoustic Diplophonic pre-treatment data showed a wavelet coefficient plot with the energy bandwidth ranging from approximately 10-32 (Figure 81). This larger bandwidth is characteristic of acoustic data. The highest energy was contained between scales 15-32. The texture of this plot appears very fine due the intermittent behavior occurring over a smaller sample width as compared with the normal acoustic waveform of Figure 73.

Figure 81. Diplophonic Patient - Acoustic Waveform (Wavelet: Pre-treatment).

The wavelet coefficient plot for the acoustic Diplophonia post-treatment data (Figure 82) was similar to that of the normal acoustic data (Figure 73) if examining the texture, coloration, and shape of the overall energies. The energies appear rippled. The energy location had a bandwidth that was between the scales of approximately 2 and 30. There also appear to be to separate bands though within the overall scale range of 2-30. They can be differentiated by their different colors. Scale 2-12 feature coefficient plots that did not correlate well with the clinical data due the coloration deviating from the red-burgundy color. The second band is at the range of 13-30. These coefficient values correlated well with the acoustic data.

Figure 82. Diplophonic Patient - Acoustic Waveform (Wavelet: Post-treatment).

The energy of the coefficients for the Diplophonic pre-treatment GAW data ranges from scale 5 to scale 13 (Figure 83). There is an overall intermittent behavior of the energy bands. However, this intermittent behavior of the energy band, ranging from approximately 13-28 is loosely spaced. This may be due to the multiple frequencies characteristic of Diplophonia.

Figure 83. Diplophonic Patient – GAW (Wavelet: Pre-treatment).

The wavelet coefficient plot of the Diplophonic voice post treatment using the GAW contains an energy band between 5-13 Hz (Figure 84). This energy band also reveals a repeating pattern of intermittent energies across the number of samples. The pattern of the energy band is similar to the normal case (Figure 74); however, it ranges between scales 5-13, as opposed to the normal case which ranged from scale 4-8.

Figure 84. Diplophonic Patient – GAW (Wavelet: Post-treatment).

The wavelet coefficient plot of the Diplophonic voice using the two mass model contains two separate energies (Figure 85). The main energy is in the scale range from 5-13. The secondary energy band present is in the range from 14-26, and appears more pronounced in the model data than the clinical data in the same scale range. These are possibly due to the lack of the subharmonic frequencies present in the FFT plot (Figure 65) that are characteristic of Diplophonia. These were not adequately captured by the model as it serves as an approximation of the clinical data.

5.2.5 Model Predications - Diplophonia

The predictive value of the two mass model was examined, using Diplophonia as an example, to test the usefulness of this method in characterizing and understanding vocal fold pathophysiology. The effect of the FFT and scalogram using GAW data for a patient with deteriorating Diplophonic condition might be visible with this analysis. This state is achieved by the introduction of increased nonlinearities into the two mass model. Specifically, the parameters that affect the collision of the vocal folds were increased by 50% from that of Figure 85 to visually confirm that a difference in the wavelet coefficient plot was present. Through experimentation, an increase of 50% was sufficient to provide enough variance to achieve this increased nonlinearity behavior. The following figures display the FFT and scalogram results of this study.

Increasing the nonlinearity using the two mass model for Diplophonia yielded FFT results (Figure 86) with frequency components that were larger in magnitude than the FFT results of Figure 67. The exception is with the fundamental frequency, which experienced a reduction in amplitude.

Correspondingly, the increased amplitudes are reflected in the wavelet coefficient plot by an increase in the energies describing the correlation between the wavelet and the GAW generated by the model (Figure 87). The same characteristic of the scalogram that was generated by Figure 85 is seen in Figure 87; however, it is accentuated further by the increased nonlinearities parameters of the model.

Figure 87. Diplophonic Voice (Model) – GAW (Wavelet: Increasing nonlinear parameters).

5.2.6 Model Behavior Using Spectral Analysis

The two mass model, in addition to time-frequency and time-scale methods such as the FFT, Spectrogram, and Wavelets provides information regarding the etiology of the pathology. An understanding of how the parameters of the vocal folds deviate from the norm and what needs to occur from a physiological standpoint to improve or worsen the condition in question could be investigated. These spectral analysis techniques may eventually be useful to the clinician as a diagnostic tool.

For example, waveforms that exhibit multiple frequencies that are reflected by the FFT and Spectrogram by multiple peaks/bands display an intermittent yellow ripple effect in the wavelet coefficient plot (Figure 85). The bandwidth for this type of effect ranges from approximately scale 13-25.

The following are general characteristics that were observed by the modification of the parameters (mass, damping, and stiffness) in the model using the normal case as the control.

1. An increase in mass decreases the fundamental frequency and increases the location of the scales corresponding to the highest energy. The bandwidth of this energy also increases.
2. A decrease in mass increases the fundamental frequency. The bandwidth of highest energy decreases and is located at a lower scale corresponding to a range of approximately 3-6.
3. Multiple harmonics characteristic of Diplophonia produces intermittent yellow ripple effects in the wavelet coefficient plot. This is always present if the peaks of the waveform are not pointy but rather have an “M” shape (Figure 50 and Figure 52).
4. Waveforms with peaks and troughs that have a “W” shape exhibit a primary energy band with a secondary energy band with a bandwidth that is at least 2 scales.
5. An increase in damping produces erratic waveforms and multiple frequencies. The ripple effect characteristic of Diplophonia is seen in the scalogram, however, the presence of the strong energy bands decrease.
6. A decrease in damping also produces erratic waveforms at times with multiple frequencies. This may be due to the accentuation of the nonlinear springs present during collision of the vocal folds, which is incorporated into the

equation describing the collision force. In addition, the location of the fundamental frequency increases.

7. An increase in stiffness produces intermittent behavior in the highest energy band with gaps that are at least 50 samples in length for the scalogram. The gaps between the ripples and the energy associated increases.
8. A decrease in stiffness increases the location of the highest energy band by shifting it up in scale. In addition, intermittent behavior of the ripples in this energy band is witnessed. The fundamental frequency component also reduces from 250 Hz. The magnitude of the reduction is dependent on the magnitude of the stiffness parameter decreased relative to the norm.
9. An increase in the coupling stiffness increases the fundamental frequency. Correspondingly, the location of the primary energy band in the scalogram is shifted down in scale and generates a narrower bandwidth. A secondary energy band with energies comparable to the primary band is witnessed. Both bands combined span a scale approximately 2-6 with a gap at scale 4 separating the two.
10. A decrease in the coupling stiffness decreases the fundamental frequency. A primary and secondary energy band is created. However, the gap separating the two is relatively small and may be difficult to detect due to the intermittent behavior that expands over the gap and appears to connect with the two energy bands.
11. An increase in the stiffness due to the collision of the vocal folds results in an increase in the fundamental frequency. The bandwidth of the highest energy

is decreased. There is a secondary energy band present ranging from a scale 6-8. There is a gap of scale 1 separating the two. In addition, both energy bands exhibit intermittent behaviors that are 50 samples in length. There is alternating high/low energy present in this behavior.

12. A decrease in the stiffness due to collision of the vocal folds shows the same characteristics as the normal voice when examining the FFT, with the exception of the wavelet coefficient plot. The scalogram appears to display three energy bands that are separated by gaps that are a width of approximately half a scale. The widths of all three bands, including the gaps, are comparable to the normal case.

The characteristics aforementioned describe the general traits of the mass, springs, and dampers when the values are increased or decreased from the normal parameter set. Similar patterns are observed when examining the pathological voice and increasing or varying the masses, springs, and dampers to see the effect of the FFT and scalogram on the pathological condition.

Zhang et. al. proposed a model for Parkinson (Equation 3.3.1) which was an extension of the model by Steinecke et. al (Equation 3.2.1). This Parkinson model included an increase in stiffness and the introduction of nonlinearities in the form of cubic expressions associated with the spring stiffness and collision for more accurate representation of the GAW for Parkinson. Zhang et. al did not do a direct comparison of their model with clinical data. Our results show that their model does not generate the Parkinson waveform similar to our clinical dataset. This was confirmed visually

by examining the waveforms and was further validated following the results of the FFT and scalogram. The simpler model provided by Steinecke was sufficient to capture the Parkinson waveform without the presence of the nonlinear terms and increased stiffness values.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

6.1 Conclusion

We investigated a refined two mass model (Steinecke et. al., 1995) for vocal fold vibrations and used it analyze vocal fold abnormalities. The two mass model is akin to a mechanical oscillator in which masses, springs, and dampers represent different vocal fold structures. This mechanical oscillator model described the glottal area vibratory dynamics. Furthermore, we investigated a model by Zhang et. al that was specifically adapted to study Parkinson's disease. The difference between these two models was the extent to which nonlinearities in the stiffness components were modeled. Acoustic signals and glottal area waveforms, acquired clinically from patients with normal and pathological voices, were obtained through standard techniques, such as acoustic and endoscopic HSLI. These signals contained unique information on the dynamic properties of the vocal system and correlated with specific voice conditions. The Steinecke et. al model was able to reproduce some of the oscillatory signals associated with vocal fold dynamics and vocal fold abnormalities related to the pathophysiology of vocal fold tissues from the clinical data. The methods presented in the thesis could potentially be used by clinicians to facilitate, more accurately, the diagnosis and treatment of vocal fold abnormalities

A comparison of the acoustic waveform with the GAW revealed that both methods contained the same frequency content. However, the GAW obtained from the HSLI method, proved to be more efficient than the acoustic waveform which

contained minimal spurious frequencies. The use of the GAW from female patients both with normal and pathological voices, together with Steinecke et. al's two mass model revealed characteristic information that might be used as a diagnostic tool. These characteristics, which can be difficult to interpret solely from the GAW, such as when comparing the normal voice with that of Papillomas, become evident following spectral analysis. A combination of the FFT, Spectrogram, and Wavelet analysis was implemented. The FFT described the frequency components in the signal while the Spectrogram showed how these frequencies varied in time. Results from the Wavelet analysis provided additional information about the waveforms, such as the location of the highest energies and their corresponding bandwidth and texture.

The power of the time-frequency and time-scale methods in characterizing vocal fold disorders and understanding the pathophysiology became evident when the GAW were compared with those simulated through the modification of parameters such as mass, damping, stiffness, and pressure in the two mass model. The response of the time-frequency and time-scale methods to changes in the parameters of the model as they converged to values characteristic of pathologies, indicated that this technique may allow for a more vigorous characterization of the pathophysiology for vocal fold disorders than relying solely on the audio and video files alone.

6.2 Future Work

The waveforms generated from the model were obtained by varying the parameters manually, through minimal parameter sensitivity analysis. This technique was time consuming. The range of parameter choices had physiological correlates. Experiments to measure vocal fold material and physical properties for various normal and pathological cases need to be conducted. This will allow for an accurate estimation of properties used in the models²³. Extensive studies of comparison between the model and clinical GAW data using more parameters that are accurate will need to be implemented to advance this research.

Future work could also involve the use of extensive sensitivity analysis and optimization techniques to determine parameter ranges that will generate given waveforms with physiological relevance. The possibility of using Hilbert-Huang Transform (HHT) techniques to analyze ensuing data in the time-frequency domain is also envisioned in future work⁴⁵. The HHT method provides a more accurate analysis of nonlinear and non-stationary data, such as that of our voice samples⁴⁵. By testing and improving the predicted value of the model with the results from FFT, STFT, and Wavelet analysis, improved diagnostic schemes might be developed.

REFERENCES

- [1] I. R. Titze, Principles of Voice Production, Iowa: National Center for Voice and Speech, 2nd printing, 2000.
- [2] "Voice Disorders: The Relationship Between Your Medical Health and Your Voice," Greater Baltimore Medical Center, 1999, <<http://www.gbmc.org/voice/disorders.cfm>>.
- [3] R.E. Hillman, "Improving the assessment and treatment of voice disorders: emerging technologies," *Proceedings of From Sound to Sense*, B173-B178, 2004.
- [4] M. M. Johns, "Update on the etiology, diagnosis, and treatment of vocal fold nodule, polyps, and cysts," *Current Opinion in Otolaryngology & Head and Neck Surgery*, 11(6):456-461, 2003.
- [5] J. Xin, J.M. Hyman, and Y-Y Qi, "Modeling vocal fold motion with a continuum fluid dynamic model: I. Derivation and Analysis," *Nonlinear Sciences*, 1(0108039):1-23, 2001.
- [6] S. Qian and D. Chen, Joint Time-Frequency Analysis, New Jersey: Prentice Hall PTR, 1996.
- [7] G. S. Goding, Jr. and D.D. Michael, "About the Voice: How does the voice work?" Lions Voice Clinic, 2004, <<http://www.lionsvoiceclinic.umn.edu/page2.htm>>.
- [8] M.C.L. Greene and L. Mathieson, The Voice and Its Disorders, London: Whurr Publishers Limited, 5th edition, 1989.
- [9] B. Story, "An overview of the physiology, physics and modeling of the sound sources for vowels," *Acoustical Science & Technology*, 23(4):195-206, 2002.
- [10] J. Lim, S. Lim, S. Choi, J. Kim, K. Kim, and H. Choi, "Clinical characteristics and voice analysis of patients with mutational dysphonia: clinical significance of diplophonia and closed quotients," *Journal of Voice*, Article in Press, 2006.
- [11] K. Hong and H. Kim, "Diplophonia in unilateral vocal fold paralysis and intracordal cyst," *Otolaryngology-Head and Neck Surgery*, 121(6):815-819, 1999.
- [12] J. P. Thomas, Voice Doctor, 2004, <<http://www.voicedoctor.net/media/cases/mucosal/polyp1.html#treat>>.
- [13] C. Leydon, J. Bauer, and C. Larson, "The role of auditory feedback in sustaining vocal vibrato," *Journal of Acoustical Society of America*, 114(3):1575-1581, 2003.
- [14] C. Dromey, P. Warrick, and J. Irish, "The influence of pitch and loudness changes on the acoustics of vocal tremor," *Journal of Speech, Language, and Hearing Research*, 45(5):879-890, 2002.
- [15] A. E. Aronson, L. O. Ramig, W. S. Winholtz, and S. R. Silber, "Rapid voice tremor, or "flutter," in amyotrophic lateral sclerosis," *The Annals of otology, rhinology, and laryngology*, 101(6):511-518, 1992.

- [16] L. Sulica, "Papilloma," Voice Medicine, 2004, <<http://www.voicemedicine.com/papilloma.htm>>.
- [17] B.H. Story and I.R. Titze, "Voice Simulation with a body-cover model of the vocal folds," *Journal of Acoustical Society of America*, 97(2):1249-1260, 1995.
- [18] T. Koizumi, S. Taniguchi, and S. Hiromitsu, "Two mass models of the vocal cords for natural sound voice synthesis," *Journal of Acoustical Society of America*, 82(4):1179-1192, 1987.
- [19] H. Herzel, D. Berry, I. Titze, and I. Steinecke, "Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling," *Chaos*, 5(1):30-34, 1995.
- [20] J. W. Van den Berg, "Myoelastic-aerodynamic theory of voice production," *Journal of Speech and Hearing Research*, 1(3):227-244, 1958.
- [21] J.C. Lucero, "Dynamics of the Vocal Fold Oscillation," Plenary talk at the XXVII National Congress on Applied and Computational Mathematics, 2004.
- [22] J. Flanagan and L. Landgraf, "Self-oscillating source for vocal tract synthesizers," *IEEE Transactions Audio Electroacoustics*, AU16(1):57-64, 1968.
- [23] M.E. Smith, G.S. Berke, B.R. Gerratt, J. Kreiman, "Laryngeal Paralysis: Theoretical Considerations and Effects on Laryngeal Vibration," *Journal of Speech and Hearing Research*, 35(3):545-554, 1992.
- [24] K. Ishizaka and M. Matsudaira, "Fluid mechanical considerations of vocal cord vibration," SCRL Monograph #8, Speech Communications Research Laboratory, 1972.
- [25] K. Ishizaka and N. Isshiki, "Computer simulation of pathological vocal-cord vibration," *Journal of Acoustical Society of America*, 60(5):1193-1198, 1976.
- [26] D. H. Klatt, "Review of text-to-speech conversion for English," *Journal of the Acoustical Society of America*, 82(3):737-792, 1987.
- [27] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," *Journal of Acoustical Society of America*, 97(3):1874-1884, 1995.
- [28] K. Ishizaka and J.L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *The Bell System Technical Journal*, 51(6):1233-1268, 1972.
- [29] J. Lucero, "Oscillation hysteresis in a two-mass model of the vocal folds," *Journal of Sound and Vibration*, 282:1247-1254, 2005.
- [30] Y. Zhang, J. Jiang, and D.A. Rahn III, "Studying vocal fold vibrations in Parkinson's disease with a nonlinear model," *Chaos*, 15(3):033903, 2005.
- [31] Y. Yan, K. Ahmad, M. Kunduk, and D. Bless, "Analysis of vocal-fold Vibrations from High-speed Laryngeal Images Using a Hilbert Transform based Methodology," *Journal of Voice*, 19(2):161-175, 2005.
- [32] Y. Yan, X. Chen, and D. Bless, "Automatic Tracing of the Vocal-fold Motion from High-speed Laryngeal Image Sequence," *IEEE Transaction on Biomedical Engineering*, Article in Press, 2006.
- [33] "Speech and language therapy endoscopy for voice disordered patients", Royal College of Speech and Language Therapists, 2004, <<http://www.rcslt.org/docs/free-pub/endoscopy.pdf>>.

- [34] S. Granqvist, S. Hertegardrd, H. Larsson, and J. Sundberg, "Simultaneous analysis of vocal fold vibration and transglottal airflow: Exploring a new experimental set-up," *Speech, Music and Hearing*, KTH, Stockholm, Sweden TMH-QPSR, KTH, 45:35-46, 2003.
- [35] A. Ghio, "Electroglottography (EGG)," 2005.
<<http://www.lpl.univ-aix.fr/~ghio/pedago-EggUK.htm>>.
- [36] S. Mitra, Digital Signal Processing: A Computer Based Approach, New York: The McGraw-Hill Companies, Inc., 3rd edition, 2006.
- [37] Full set of MATLAB® manuals, releases and other documentations in PDF is available on-line at:
<<http://www.mathworks.com/access/helpdesk/help/fulldocset.shtml>>.
- [38] "Windowing: Optimizing FFTs Using Window Functions," National Instruments™, 2006, <<http://zone.ni.com/devzone/conceptd.nsf/webmain/E0F0B1FFE86A14A986256F2E0005AF4D#4>>.
- [39] P. Wickramarachi, "Effects of Windowing on the Spectral Content of a Signal," *Sound and Vibration*, 10-11, January 2003.
- [40] "Electronic Warfare and Radar Systems Engineering Handbook: Transforms/Wavelets," Naval Air Warfare Center Weapons Division, 2006, <<https://ewhdbks.mugu.navy.mil/wavelet.htm>>.
- [41] P. Schniter, "Short-time Fourier Transform," *Connexions*, 2005, <<http://cnx.org/content/m10417/latest/>>.
- [42] O. A. S. Youssef, "New Algorithm to Phase Selection Based on Wavelet Transforms," *IEEE Transactions on Power Delivery*, 17(4):908-914, 2002.
- [43] Y.T. Chan, Wavelet Basics, Massachusetts: Kluwer Academic Publishers, 1995.
- [44] Y. Zhang and J.J. Jiang, "Chaotic vibrations of a vocal fold model with a unilateral polyp," *Journal of Acoustical Society of America*, 115(3):1266-1269, 2004.
- [45] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, and H.H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings Royal Society London A*, 454:903-995, 1998.