

**Acoustical Analysis of the Swallowing Mechanism for Diagnosis of
Dysphagia**

by

Samaneh Sarraf Shirazi

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Biomedical Engineering

University of Manitoba

Winnipeg, Manitoba, Canada

Abstract

Swallowing dysfunction (dysphagia) is a common disorder in patients with neurological impairments, head/neck injuries or brain-stem stroke. The main objectives of this thesis were to use acoustical analysis of swallowing and breath sounds for 1) understanding the swallowing mechanism and the main cause of dysphagia, and 2) developing a noninvasive diagnostic technology to detect swallowing aspiration (the entry of bolus into airway); thus, identifying individuals at high risk of severe dysphagia.

As the first objective of the study, swallowing mechanism modeling in two groups of healthy individuals and dysphagic patients (due to cerebral palsy or stroke) was approached by using two different assumptions to relate the swallowing sounds either to the pharyngeal response or to the neural activities that initiate the swallow. The results showed that the model with the assumption of neural activities as the cause of dysphagia was a better fit to the available data.

As the second main objective of the study, we analyzed breathing and swallowing sounds of 50 dysphagic individuals during the fiberoptic endoscopic evaluation of swallowing (FEES) or the videofluoroscopic swallowing study (VFS). The results showed 91% sensitivity and 85% specificity in identifying patients with severe aspirations. Also, the algorithm was able to detect the silent aspiration among the swallows of each patient.

The proposed methods led to development of a non-invasive and reliable diagnostic/screening tool as an aid to the clinical examination of swallowing. The proposed acoustic method can be performed at the patients' bedside to determine the appropriate further assessment or a dietetic treatment; thus, reducing the health care cost by prioritizing the patients' referrals to the VFS/FEES tests.

Acknowledgments

This research would not have been successful without the supervision of my supervisor, guidance of my committee members, help from friends, and support from my family.

First and foremost, my utmost gratitude goes to my supervisor, Prof. Zahra Moussavi for her supervision, inspiration, and encouragement to complete this project. The joy and enthusiasm she has for her research has always been a great motivation for me in completing my research.

It is a pleasure to thank Prof. Miroslaw Pawlak, Prof. Sherif Sherif, and Prof. Neil Popplewell for their helpful inputs and comments during my Ph.D studies.

I would like to thank the speech language pathologists at Riverview Health Center, Reesa Daun, Laura Lenton, Caitlin Buchel, and at Deerlodge Health Center, Gillian Barnes, Colleen Braun-Janzenfor, Jamie Conde, Debra MacGill, Monique Piatt, Lexa Ramsey, Carrie Sulkers, Kelly Tye Vallis, for their invaluable help in recruiting patients and collecting data. I am most grateful to Andrea Meakin, the regional manager speech-language pathology, who always welcomed the collaboration in both Riverview and Deer Lodge health centers. My research would not have been possible without their help.

This study was supported financially in part by Natural Sciences and Engineering Research Council (NSERC) of Canada, Riverview Health Centre (Winnipeg), MITACS in collaboration with SagaTech company. I would like to thank the research manager of Riverview health center, Dr. John Bond for his support and encouragement. I also acknowledge Dr. Ron Platt for providing me the opportunity to carry on my research at SagaTech.

I am grateful for being a member of the biomedical engineering group. The group has been a source of friendship as well as wonderful ideas. I would also like to extend huge, warm thanks to all my dearest friends in Winnipeg, who have always been there for me. I am indebted to them for their support, care, and precious friendship.

I wish to express my love and gratitude to my beloved family; for their encouragement and endless love.

Dedication

To my parents,
for their love and support

Table of Contents

Abstract	i
Acknowledgments	iii
Dedication	v
List of Tables	viii
List of Figures	ix
List of Copyrighted Material for which Permission was Obtained.....	xii
Glossary	xiii
Chapter 1 - Introduction.....	1
1.1 Objectives	2
1.2 Organization of the Thesis.....	3
Chapter 2 - Background.....	5
2.1 Swallowing Disorders	8
2.2 Swallowing Assessment Techniques.....	10
Chapter 3 - Acoustical Modeling of Swallowing Mechanism.....	19
3.1 Introduction.....	19
3.2 Method.....	20
3.2.1 Data.....	21
3.3 Results.....	28
3.4 Discussion	32
3.5 Summary.....	36
Chapter 4 - Non-invasive and Automatic Detection of Patients at High Risk of Swallowing Aspiration	37
4.1 Introduction.....	37
4.2 Methods.....	39
4.2.1 Study Subjects	39
4.2.2 Data Recording.....	40
4.2.3 Signal Analysis.....	41
4.3 Classification	43
4.4 Results.....	44
4.5 Discussion	46
4.6 Summary.....	49
Chapter 5 - Detection of Swallows with Silent Aspiration using the Swallowing and Breath Sounds Analysis.....	50
5.1 Method.....	51
5.1.1 Participants	51
5.1.2 Signal Processing	51
5.1.3 Feature Extraction.....	53
5.1.4 Unsupervised Classification.....	54
5.2 Results.....	56

5.3	Discussion	59
5.4	Summary.....	62
Chapter 6	- Conclusions and Recommendations for Future Work.....	63
6.1	Conclusions.....	63
6.2	Future Work Recommendation.....	65
	Bibliography	67
	Appendix A Investigating the Statistical Properties of the Swallowing Sound ..	76
A.1	Method	77
A.1.1	Data.....	77
A.1.2	Statistical Analysis.....	77
A.2	Results.....	82
A.3	Discussion	85
	Appendix B Characteristics of the Swallowing Sounds Recorded in the Ear, Nose and on Trachea.....	87
B.1	Method.....	88
B.1.1	Experimental Data.....	88
B.1.2	Signal Analysis.....	90
B.2	Results and discussion	92
B.2.1	Qualitative Observations of Swallow Signal Quality	92
B.2.2	Analysis of the Peak (f _{peak}) and Maximum Frequency (f _{max})	93
B.2.3	Analysis of Average PSD Magnitude over Octave Frequency Bands 93	
B.2.4	Analysis of Wavelet Coefficients.....	95
B.2.5	Study Limitation	97
B.3	Conclusion	97
	Appendix C Comparison Study of Other Time Frequency Representations.....	99
C.1	Introduction.....	99
C.1.1	Cohen's Class of Time-Frequency Representation	99
C.1.2	Desired Properties Related to the Kernel.....	101
C.1.3	Interference Properties.....	101
C.2	Method.....	106
C.3	Results.....	106
C.4	Discussion	112
	Appendix D Frequency Response of the Microphone.....	114
D.1	Introduction.....	114
D.2	Measurement Procedure	115
D.3	Results.....	116
D.3.1	Direct Connection of the Microphone to the Speaker.....	116
D.3.2	Microphone in front of the Speaker at 25cm Distance	117
D.3.3	Frequency Response.....	117
D.4	Conclusion	120

List of Tables

Table 3-1. The classification results at 8 levels of decomposition.	30
Table 5-1. Classification results of swallows for 22 patients having aspiration....	58
Table A-1. The difference between the theoretical and estimated values of the interquartile range averaged for each group.....	84
Table C-1. The properties of some TFR kernels	104
Table C-2. The classification results of aspiration detection by using the features obtained from two time-frequency representation methods: spectrogram and Born-Jordan.	111
Table D-1. The frequencies used to test the microphone.....	115

List of Figures

Fig. 2-1. (a) Oral phase (b) pharyngeal phase, and (c) esophageal phase of the swallowing [2]. (part of Figure 2.13: The Normal Swallow. From Evaluation and Treatment of Swallowing Disorders, 2 nd Ed. (p. 28), by J. A. Logemann, 1998, Austin, TX: PRO-ED. Copyright 1998 by PRO-ED, Inc. Reprinted with permission. This diagram is prohibited from any further duplication)...	8
Fig. 2-2. FEES procedure (a) feeding by the SLP during the test, and (b) the images of the anatomical areas seen on the screen.	13
Fig. 2-3. (a) Normal, non-aspirated swallow; (b) an aspirated swallow as observed by FEES. The bolus in the airway as a result of aspiration is indicated by the arrow.....	13
Fig. 2-4. The VFS setup for swallowing assessment. The individual is given barium coated food or liquid while the x-ray images are displayed and recorded.....	14
Fig. 2-5. (a) Normal, non-aspirated swallow; (b) an aspirated swallow as observed by VFS. The bolus in the airway as a result of aspiration is indicated by the arrow in (b).	15
Fig. 2-6. A typical swallowing sound signal.	18
Fig. 2-7. (a) A typical swallowing sound signal with the before/after breath in time domain. (b) the spectrogram of the same signal.	18
Fig. 3-1. Model of swallowing sound production [36].....	19
Fig. 3-2. The model [39] considered in the first approach to represent the IDS part of swallowing sound.....	24
Fig. 3-3. PSD and its fitted Bode plot of the IDS segment of the swallowing sound of a control subject.	25
Fig. 3-4. PSD and its fitted Bode plot of the IDS segment of the swallowing sound of a dysphagic subject.	26
Fig. 3-5. The time domain response of the Bode approximation models derived for the control (—) and dysphagic (--) groups.	26
Fig. 3-6. The Mean Square Error (MSE) between the PSD of each swallowing sound signal of control/dysphagic subjects (from both the datasets) and the Bode approximated transfer function for control/dysphagic models as shown in x/y axes, respectively (The bisector line of the plane is plotted for better presentation of the results).	29
Fig. 3-7. The approximation coefficients at 8 scales (a)-(h) for a typical swallowing sound of a non-dysphagic subject.	31
Fig. 3-8. The approximation coefficients at 8 scales (a)-(h) for a typical swallowing sound of a dysphagic subject.....	31
Fig. 3-9. The scatter plot of the energy of the control and dysphagic data at two levels of wavelet decomposition averaged among each subject's swallows. The values of x-axis and y-axis belong to the 2 nd and 3 rd level, respectively. The filled marker shows the values of the first dataset while the blank ones represent the second.....	32
Fig. 4-1. An illustration of the pattern of airflow in (a) a clear trachea and, (b) a trachea with a food particle in it.	39
Fig. 4-2. Spherical coordinate in the phase-space.....	43

Fig. 4-3. The 3-D phase-space plot of the two breath signals (following swallows) of two patients: (a) without, and (b) with aspiration. The asterisks show the points that fall outside the ellipsoid.	44
Fig. 4-4. The time-domain breath sound signals corresponding to the graphs shown in Fig. 4-3: (a), (b) without, and (c) and (d) with major aspiration. Data points that fell outside the ellipsoid are shown in red.	45
Fig. 4-5. Scatter plot of the ellipsoid feature for all the participants. The triangle marker (Δ) represents the patients in the severe aspiration group, and the asterisk (*) shows the group of patients with mild or no aspiration.	46
Fig. 5-1. The comparison of the PSD of breath sound signals after the swallow (simultaneously recorded with FEES) for the swallows (with/without aspiration) of a dysphagic individual.	53
Fig. 5-2 Scatter plot of the three average power features of the breath sounds after swallows of a patient having aspiration.	54
Fig. 5-3. Flowchart for aspiration detection.	56
Fig. A-1. The bispectrum of the swallowing sound of (a) a control and (b) a dysphagic subject.	84
Fig. A-2. The test of linearity results: the difference between the estimated and the theoretical interquartile range for all the 10 control and 10 dysphagic data (averaged over all the swallows of each individual).	85
Fig. B-1 Setup for Swallowing Experiment.	90
Fig. B-2 A typical normalized swallowing and breath sounds signal as marked by the solid arrow followed by breath sounds as indicated by the dashed arrow; the signals are shown in time domain, and recorded simultaneously from trachea, nose, and ear. Au= arbitrary unit.	91
Fig. B-3 Typical spectra of the IDS segment of a swallowing sound recorded at the ear, nose and trachea of one subject. Each segment was normalized to its total energy before spectral estimation.	92
Fig. B-4. f_{peak} of the IDS, BTS and the expiration segments for all locations averaged among subject's data.	93
Fig. B-5. The average power of the IDS, BTS and the expiration segments calculated over the octave frequency bands. The values are averaged among subject's data.	95
Fig. B-6. The mean and standard error of the energy of wavelet coefficients averaged for the IDS segments of all subjects a) 2 nd order decomposition, b) 3 rd order decomposition. The value of the standard error shows the variation within each subject.	96
Fig. C-1 (a) The test signal consisting of two frequencies, 12 Hz and 20 Hz in time domain, (b) spectrogram of the signal, (c) the Wigner representation, and (d) Born-Jordan time-frequency representation of the signal.	105
Fig. C-2 (a) The spectrogram of a breath sound after a non-aspirated swallow, (b) the Born-Jordan distribution of the same signal.	107
Fig. C-3 (a) The spectrogram of a breath sound following an aspiration, (b) the Born-Jordan distribution of the same signal.	108
Fig. C-4 (a) The PSD of a breath sound after a non-aspirated swallow, (b) the frequency marginal of Born-Jordan distribution of the same signal.	109
Fig. C-5 (a) The PSD of a breath sound following an aspirated swallow, (b) the frequency marginal of Born-Jordan distribution of the same signal.	110

Fig. C-6. Scatter plot of the three average power features of the breath sounds after swallows of a patient having aspiration by using (a) PSD (b) the frequency marginal of Born-Jordan distribution of the same signal.....	111
Fig. C-7 The frequency marginal of Born-Jordan distribution of the breath sound following (a) a non-aspirated swallow as shown in Fig. C-4 and (b) an aspirated swallow as shown in Fig. C-5.....	113
Fig. D-1. Spectrogram of the sound signals played through the speaker.	116
Fig.D-2. Spectrogram of the sound signal, recorded by the reference microphone, Sony ECM 77-B, placed on the speaker.	117
Fig. D-3. Spectrogram of the sound signal, recorded by the Sony ECM 77-B microphone, hanging in front of the speaker at the distance of 25 cm.....	118
Fig. D-4. Three measurements of the frequency response of the Sony ECM 77-B microphone hanging in front of the speaker at the distance of 25 cm.....	119

List of Copyrighted Material for which Permission was Obtained

The following list provides the title of each copyrighted item, its source, and the page on which it appears in the thesis.

- Fig. 2-1. (a) Oral phase (b) pharyngeal phase, and (c) esophageal phase of the swallowing [2]. (part of Figure 2.13: The Normal Swallow. From Evaluation and Treatment of Swallowing Disorders, 2nd Ed. (p. 28), by J. A. Logemann, 1998, Austin, TX: PRO-ED. Copyright 1998 by PRO-ED, Inc. Reprinted with permission. This diagram is prohibited from any further duplication) 8

Glossary

Anterior

Anterior refers to the structures at the front of the human body.

Cranial nerves

Cranial nerves emerge directly from the brain in contrast to the spinal nerves, which emerge from the spinal cord.

Epiglottis

Epiglottis is a flap of tissues located behind the tongue and in front of the larynx. At rest, it has an upright vertical position to allow breathing. During the swallowing, it folds down to a more horizontal position to cover the larynx and prevents food from going into the trachea.

Hyoid bone

It is located at the root of the tongue between the mandible and the larynx. It serves as an anchoring structure for many muscles such as the tongue, larynx, pharynx and epiglottis.

Larynx

The larynx is an apparatus made up of cartilage, ligaments, muscles, and mucous membrane that connects the lower part of the pharynx with the trachea. Vocal folds are located within the larynx.

Medulla

A part of the brainstem that is responsible for the control of breathing and swallowing.

Nasopharynx

The nasal part of pharynx, which extends from base of skull to the soft palate.

Nerve plexuses

A nerve plexus is a network of connected nerve fibers that link spinal nerves with different areas of the body.

Oropharynx

The oral part of pharynx, which extends from hard palate to hyoid bone.

Pharynx

The pharynx is the common channel for swallowing and breathing. It extends from the base of the skull to the lower border of the cricoid cartilage, where it becomes the esophagus.

Posterior

Posterior refers to the structures at the back of the human body.

Pyriform sinus

The pyriform sinus lies on either side of the larynx.

Upper Esophageal Sphincter (UES)

UES refers to the superior portion of the esophagus. The swallowing reflex triggers the UES opening.

Vagus cranial nerve

The vagus nerve is the tenth cranial nerve, which is the longest of all cranial nerves.

Chapter 1- Introduction

Swallowing is one of the most complex mechanisms in the human body; it involves approximately fifty paired muscles [1]. The timing and coordination of the swallowing events are very important since any slight mismatch in the process may result in aspiration (the food is drawn into the airway below the level of true vocal folds). A normal swallow usually occurs in less than a second; it requires intricate control and coordination of the three swallowing phases. Any impairments such as abnormalities in food processing, delays in initiating swallows, difficulties in swallowing liquid, and inefficient oral and pharyngeal clearance of swallowed material may cause swallowing disorders (dysphagia) [2]. Dysphagia can occur as a result of congenital abnormalities, structural damage, cerebral palsy, head trauma, spinal cord injury, and damage after treatment for the oral, oropharyngeal and laryngeal cancer [2]. Also affected are individuals who suffer from cerebrovascular accidents in the form of complete or partial swallowing disorders. Dysphagic individuals are at risk of pneumonia, malnutrition, and dehydration [2].

There are different methods used to evaluate the specific characteristics of the swallowing mechanism such as the timing and movement pattern of anatomical

structures. Each method has its own advantages and limitations. Currently, among all the methods, videofluoroscopy (VFS) and fiberoptic endoscopy (FEES) are known as the gold standard assessment techniques [2, 3]. However, significant limitations of VFS are the subjects' exposure to x-ray radiation, and that mild dysphagic individuals may not aspirate during the short VFS assessment. The limitations of FEES include the inability to view the oral cavity and the striated esophagus, as well as the events that may occur while the view is obstructed during the swallowing act (whiteout period). Furthermore, both VFS and FEES require patient's cooperation, and are not portable and cannot be performed in a typical eating environment. Therefore, a non-invasive method to assess the risk of dysphagia that can be applied in any setting with minimal interference on normal eating procedure of the patient would be advantageous and of great interest for clinicians.

Acoustical analysis of swallowing mechanism has received considerable attention in the last two decades. Its application as a monitoring tool for swallowing assessment [4], [5], [6] as well as gaining more insight about the swallowing mechanism and its pattern of occurrence within the respiratory cycle in both healthy and dysphagic individuals at different age groups [2], [7], [8] were investigated. However, the acoustical study of the swallowing is still at its infancy, and needs to be paired with the currently used methods such as VFS or FEES for validation. It should be noted that the VFS or FEES does not have any impact on the swallowing sounds.

1.1 Objectives

The goal of this thesis is to investigate the application of acoustical analysis in modeling the swallowing mechanism as well as identifying the individuals at high risk of severe aspiration. The objectives are:

- Developing a mathematical model for swallowing mechanism that can distinguish between dysphagic and healthy (non-dysphagic) individuals, and possibly the cause of dysphagia.
- Developing an algorithm to detect swallowing aspirations; thus, developing a screening technique to identify patients at high risk of severe aspiration to prioritize patients' referrals to the VFS/FEES tests.

1.2 Organization of the Thesis

This thesis is focused on the acoustical analysis of the swallowing for both swallowing mechanism modeling and as a tool to detect dysphagia. The thesis has two major parts: 1) modeling the swallowing mechanism to identify dysphagic individuals and also possibly the cause of dysphagia, and 2) detecting aspirated swallows independent of, but with comparable accuracy to, imaging assessments.

In Chapter 2, the swallowing mechanism and the assessment techniques are explained. In Chapter 3, a mathematical linear model for swallowing sound generation and transmission is derived and followed by a pilot study on the use of the model for identifying patients at risk of dysphagia. This part of the work has been published in [9]. The nonlinearity characteristics of swallowing sounds are presented in Appendix A [10].

In Chapter 4, a novel acoustic swallowing assessment is proposed to identify patients at high risk of aspiration [11]. In Chapter 5, the feasibility of the acoustical analysis for detection of silent aspiration is investigated. The classification algorithm is developed based on the time-frequency features obtained from the spectrogram of the breath sound signal. This part of the work has been published in [12], [13]. Appendix D contains the study, in which we

tested the frequency response of the microphone to ensure it did not affect the acoustic characteristic of the swallowing and breathing sound within the frequency range of our interest.

Investigation of other time-frequency techniques are discussed in Appendix C, where the classification results of the aspiration detection method using the features extracted from Born-Jordan time-frequency method are compared with those described in Chapter 5. Chapter 6 presents the overall conclusions, the main contributions of this thesis and recommendations for future work.

One major limitation with acoustical diagnostic techniques for dysphagia is the difficulty of recording a good quality sound signal if the patient has loose skin over the trachea. Therefore, in Appendix B, we explore the viability of using the ear and nose as alternative recording locations for recording breath and swallowing sounds for the purpose of dysphagia and aspiration detection [14].

Chapter 2- Background

Swallowing occurs in the oral cavity, pharynx, larynx and esophagus. The oral cavity includes the lips, the teeth, soft palate, hard palate, mandible, floor of the tongue, and faucial arches [2].

The pharynx is a fibro-muscular tube that extends from the base of the skull to the inferior part of the cricoid cartilage at which point it becomes the esophagus. The pharyngeal structure involved in the swallowing is divided into nasal, oral and laryngeal parts. The nasal part, which lies behind the nose, is entirely for respiration while the oral and laryngeal parts contribute to the pathway for both food and air. The muscular wall of the pharynx is comprised of the three pairs of pharyngeal constrictors named superior, medial and inferior constrictors which their fibers are attached anteriorly to the soft palate, the tongue base, the mandible, hyoid bone, thyroid and cricoid cartilage.

The larynx is a valve connecting the pharynx to the trachea. The larynx is suspended from the hyoid bone by the thyrohyoid muscle and thyrohyoid ligament. The hyoid bone, which serves as the foundation of the tongue, is suspended by the muscles of the floor of the mouth. The pharynx ends with the Upper Esophageal Sphincter (UES), and becomes the collapsed muscular tube called the esophagus which extends to the stomach.

Based on the anatomic structure and the timing of the events, swallowing is divided into three distinct phases: oral, pharyngeal and esophageal.

1- The oral phase: During this phase, which is the only voluntary phase of swallowing, the food is mechanically formed, chewed (masticated) and mixed with saliva in preparation to be swallowed and pass smoothly through the pharynx and esophagus. The facial muscles and the muscles of the tongue are involved to seal the oral cavity, open/close the jaw, and propel the bolus posteriorly. This phase is shown in Fig. 2-1(a).

The sensory receptors in the oropharynx and tongue are stimulated when the tongue propels the bolus posteriorly. The decoding of the sensory information which occurs in the medulla triggers the pharyngeal phase [2].

2- The pharyngeal phase: Events during the pharyngeal phase, which is shown in Fig. 2-1(b), consist of different activities:

- sealing of the nasopharynx by the elevation of soft palate to prevent the chewed food from entering the nasal cavity, and to develop positive air pressure around the pharynx to drive the bolus for the rest of the swallowing process;

- elevation and anterior movement of the hyoid bone and the larynx that contribute to the closure (protection) of the airway. Anterior movement of the hyoid bone and larynx along with the contraction of the thyrohyoid muscle that connects the hyoid bone and the thyroid cartilage, provides the force to pull the UES open;

- closure of the larynx that prevents any penetration of the food into the airway at the level of the true vocal cords. This closure occurs by elevation of larynx and movement of the epiglottis that covers the trachea. The epiglottic movement occurs in two steps: Horizontal movement which is induced by hyoid

elevation and downward movement as a result of the contraction of the thyroepiglottic muscle;

- UES Opening: Anatomically, UES is the location of the cricopharyngeal muscle (CP) and the inferior muscle (IP) at the cricoid cartilage. At rest, the UES is kept closed by the contraction of the CP and IC muscles. At the end of the pharyngeal stage, these muscles relax to allow the bolus to enter the esophagus.

As the bolus enters the esophagus, the tongue, hyoid bone, larynx and epiglottis are restored to their original positions. The upper esophageal sphincter closes when the tail of the bolus passes through the UES into the cervical part of the esophagus.

Sensory information that guides the pharyngeal stage is conveyed by the ninth and the tenth vagus cranial nerves to the medulla. Any sensory loss in the pharynx or larynx results in the impairment of swallowing and penetration of the food into the airway.

3- The esophageal phase: As shown in Fig. 2-1(c), in this stage the food is transported along the esophagus by peristaltic contraction waves generated by the sequential contraction of the esophageal wall muscle. The esophageal wall consists of two layers of muscle, a circular inner layer and the longitudinal outer one. Each layer has a different type of muscle, i.e. the upper third of the esophagus is dominated by the striated muscle, while the lower third is made up of smooth muscle. A combination of both types is found in the middle third of the esophagus.

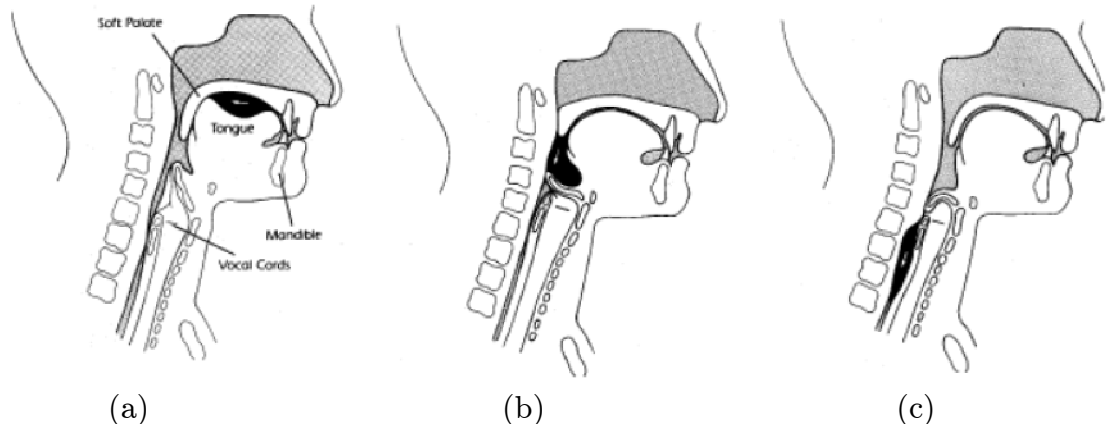


Fig. 2-1. (a) Oral phase (b) pharyngeal phase, and (c) esophageal phase of the swallowing [2]. (part of Figure 2.13: The Normal Swallow. From Evaluation and Treatment of Swallowing Disorders, 2nd Ed. (p. 28), by J. A. Logemann, 1998, Austin, TX: PRO-ED. Copyright 1998 by PRO-ED, Inc. Reprinted with permission. This diagram is prohibited from any further duplication)

The contraction of the circular muscles along with local shortening of the longitudinal muscle transports the bolus from the pharynx to the stomach. The negative intrathoracic pressure occurring at the end of inspiration can help the propagation of the radially symmetrical waves of muscle contraction and relaxation (peristalsis) in the esophagus.

The peristaltic mechanism is controlled by the central nervous system, vagus nerve, the intrinsic esophageal neurons and nerve plexuses. The relative role of each part of the nervous system depends on the type of the muscle, i.e. the upper third of the esophagus is mainly controlled by the vagus nerve while in the lower portion the role of the vagus nerve is to facilitate the autonomous intrinsic peristaltic wave.

2.1 Swallowing Disorders

Difficulty in moving food from the mouth to the stomach is called dysphagia. As mentioned in the introduction, dysphagia can occur as a result of congenital abnormalities, structural damage, cerebral palsy, head trauma, spinal cord injury,

and damage after treatment for the oral, oropharyngeal and laryngeal cancer [2]. Also affected are individuals who suffer from cerebrovascular accidents in the form of complete or partial swallowing disorders.

In general, swallowing abnormalities can be categorized in four groups: Penetration, Residue, Backflow and Aspiration [2]. They are summarized below; aspiration is more elaborated due to its importance.

- Penetration is associated with the entry of food into the larynx at a level higher than the true vocal cords. Failure of laryngeal elevation, laryngeal closure and epiglottis tilt may cause penetration [2].
- Residue is the oropharyngeal disorder, in which the food is left behind in the pharynx as the result of the inadequate push applied by the tongue to the bolus.
- Residue Backflow disorder happens if the food moves back either from the pharynx into the nasal cavity or from the esophagus into the pharynx [2].
- Aspiration is defined as the entry of the food into the airway below the true vocal folds. Poorly coordinated swallowing, as a result of any physiological or anatomical disturbances such as those in neurological disorders or in patients with stroke, may cause aspiration. It occurs if the protective mechanism of the glottic/supraglottic fails or does not happen in coordination with the bolus movement through the pharynx [15].

Aspiration is the most important symptom of dysphagia; it can occur before, during and after the opening of UES. Aspiration before the swallow results from the damaged tongue function or the delayed triggering of the swallow reflex. The reduced control of the tongue during the oral preparatory phase may let a part of the bolus roll into the trachea causing aspiration. On the other hand, the delayed

triggering of the swallow happens due to any pause in the bolus movement in the transition between the consecutive oral and pharyngeal stages.

Aspiration during the swallow relates to reduced laryngeal closure, the mechanism designed to protect the airway during the swallow. Aspiration after the swallow may have different causes: the remaining residue in the pharynx after the swallow may be aspirated or inhaled into the trachea. Other possible reasons may include reduced laryngopharyngeal sensation, the reduced pharyngeal peristalsis, pharyngeal dysfunction, reduced laryngeal elevation, or cricopharyngeal dysfunction.

Aspiration can be either accompanied [16] by a clinical indication (such as coughing /choking and a wet vocal quality) or be silent. It has been reported that 40% of dysphagic individuals aspirate silently. The majority of the patients with silent aspiration are not diagnosed until they develop pneumonia. The frequency of aspiration after a stroke has been reported to be between 51% and 73% [17], and 38%-41% after acquired brain injury (ABI) [18]. Also, frequent aspiration has been found in patients with head and neck cancers and advanced dementia [19].

2.2 Swallowing Assessment Techniques

The instrumental techniques to assess swallowing disorders include ultrasound, electromyography, pharyngeal manometry, VFS, FEES, scintigraphy, auscultation and acoustical analysis.

Ultrasound

Ultrasound can examine the oral phase of swallowing such as tongue function and temporal relationships between the events during this phase [20]. The temporospatial pattern and the quantitative analysis of the hyoid bone movement can be obtained by this method [21]. Although this non-invasive method provides

multiple plane views of the oral phase, its inability to detect penetration or aspiration (due to the shadows cast by the pharyngeal structure) significantly limits its application.

Electromyography (EMG)

Electromyography (EMG) is used to determine which muscles are active during swallowing, and to evaluate the strength and the duration of their activity during different phases of the swallowing. For example, the contraction properties of the muscles of the mouth, larynx, and pharynx during swallowing have been studied using EMG technique [22]. Furthermore, the EMG method can help to get an understanding of the physiology of the lower cranial nerves [23]. Most of the studies of the pharyngeal stage of swallowing focused on the activity of a single muscle or muscle pair such as cricopharyngeus, superior pharyngeal constrictor, or levator veli palatini. EMG studies of swallowing cannot help identifying neither aspiration, nor residual problems of swallowing disorders.

Manometry

Manometric data are recorded in the pharynx by inserting a catheter transnasally into the pharynx, UES and the cervical part of the esophagus [24]. As the bolus passes through those regions, the pressure changes, which result from the muscle contractions, are recorded at several holes along the catheter. In addition, pharyngeal manometry can be used to quantify the strength of the muscle contraction, the intra bolus pressure and the UES relaxation.

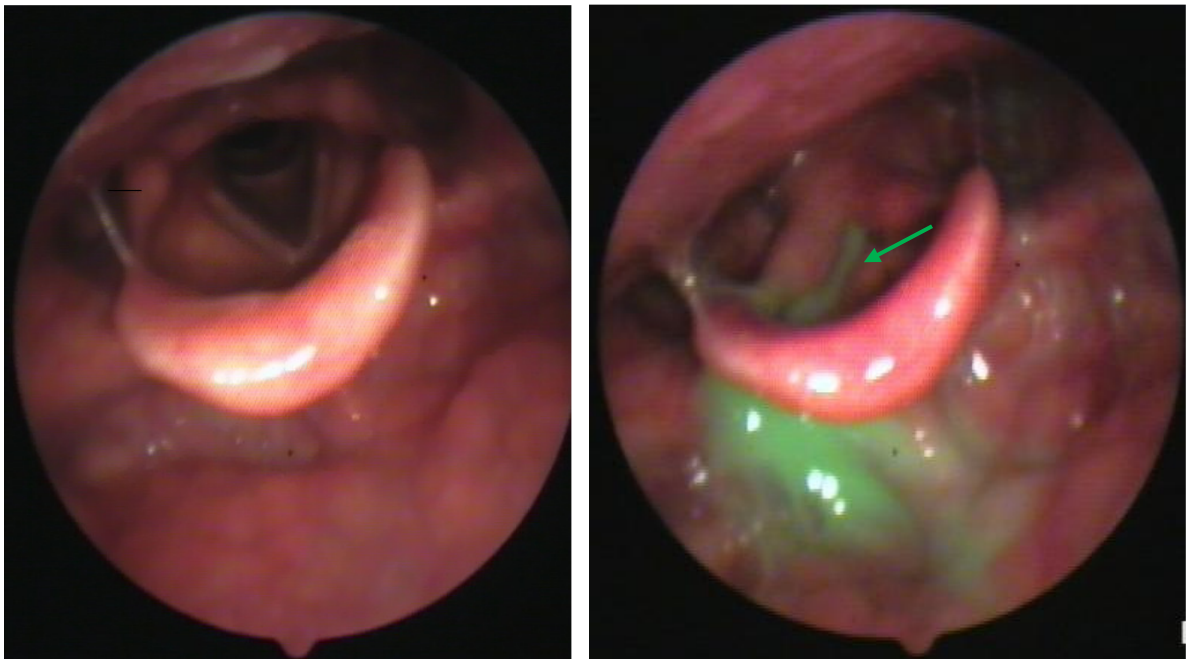
In order to correlate specific characteristics of the recorded space-time pressure with physiological events during the swallowing, pharyngeal manometry is typically combined with VFS. However, the invasive nature of this method limits its clinical application.

Fiberoptic Endoscopic Evaluation of Swallowing (FEES)

Fiberoptic Endoscopic Evaluation of Swallowing (FEES) is used to provide a dynamic view of the anatomical structure of the oral cavity, pharynx and larynx before and after swallowing. FEES is performed by inserting a small fiber-optic endoscope transnasally down to the level of the soft palate. FEES can visualize aspiration of the saliva secretion, post-swallow residue in the pharynx and vocal fold abnormalities. Compared to other swallowing assessment tools, it is more sensitive in detection of residue and its location. However, the observations obtained by FEES are limited. The interaction between the oral, pharyngeal and upper part of the esophagus cannot be visualized during the “whiteout” period when endoscope’s tip contacts the swallowing structures. This obstructs the image during the period when the most important swallowing events such as the epiglottis closure occur. Moreover, FEES is an invasive method that might be uncomfortable for the patients. The way aspiration is detected by FEES is shown in Fig. 2-3 in two frames of a recorded video during FEES test; the left image shows the anatomical structure of swallowing before the swallow, and the right one shows the frame after swallowing; the aspirated bolus is colored, and pointed out by an arrow, to be seen easily.



Fig. 2-2. FEES procedure (a) feeding by the SLP during the test, and (b) the images of the anatomical areas seen on the screen.



(a)

(b)

Fig. 2-3. (a) Normal, non-aspirated swallow; (b) an aspirated swallow as observed by FEES. The bolus in the airway as a result of aspiration is indicated by the arrow.

Videofluoroscopy Study (VFS)

VFS and FEES are considered to be the gold standard tools to assess the structure and mechanism of the swallowing. During VFS examination, subjects are fed small amounts of a barium-coated bolus, and X-ray waves are used to

track the bolus and capture the swallowing events in real time. Figure 2-8 shows the VFS recording equipment. It enables visualization of the movement of the oral cavity structures, larynx, hyoid bone, tongue base, pharyngeal walls and cricopharyngeal region, along with monitoring the bolus position. Also, it permits accurate measurement of the temporal relationships among swallow events, the opening ability of the upper esophageal sphincter (UES) and detection of the mechanisms causing aspiration [25]. The way aspiration is detected by VFS is shown in Fig. 2-5.

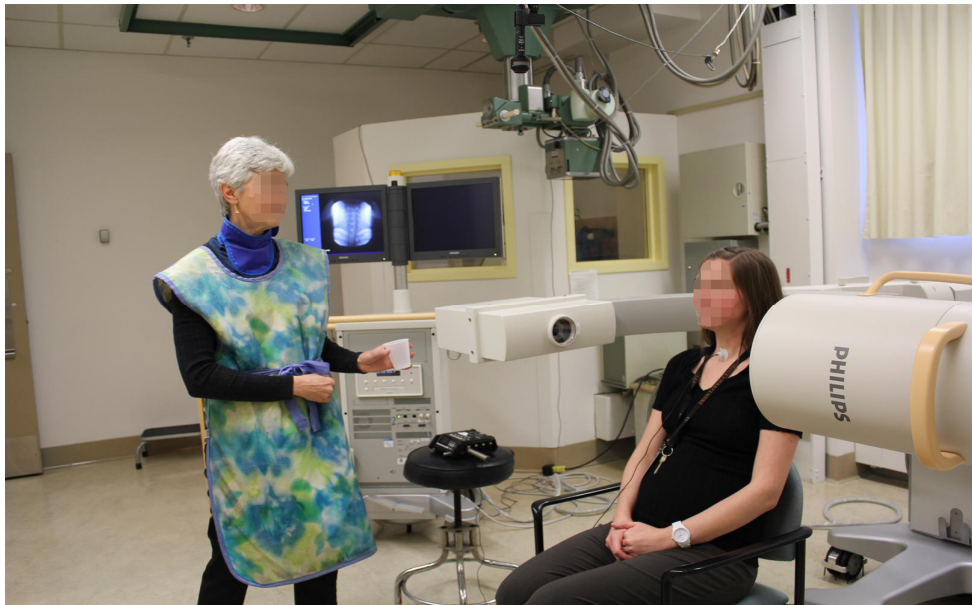


Fig. 2-4. The VFS setup for swallowing assessment. The individual is given barium coated food or liquid while the x-ray images are displayed and recorded.

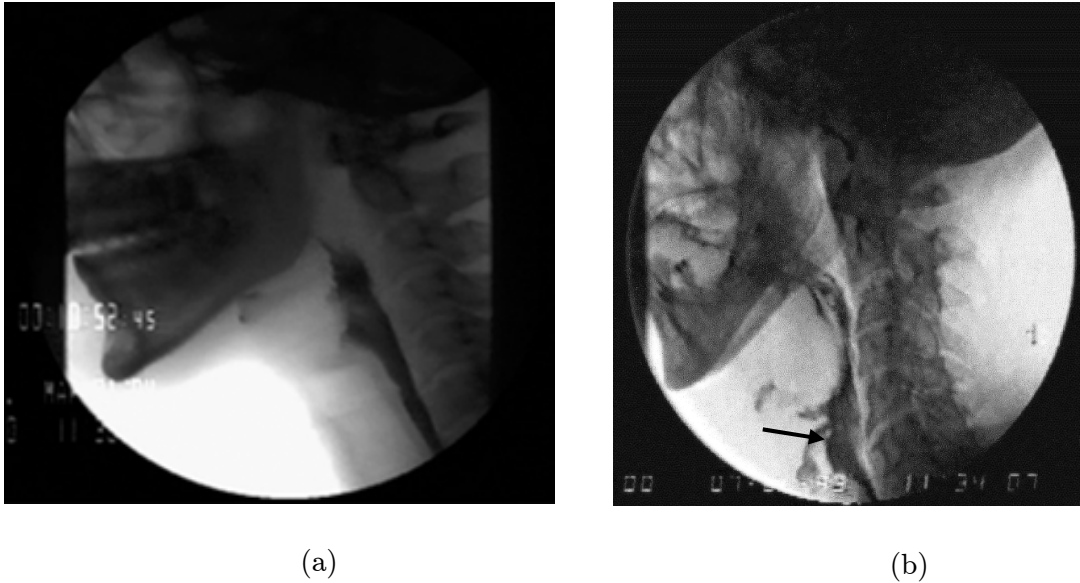


Fig. 2-5. (a) Normal, non-aspirated swallow; (b) an aspirated swallow as observed by VFS. The bolus in the airway as a result of aspiration is indicated by the arrow in (b).

The most important limitation of VFS is the subjects' exposure to radiation as radio-sensitive structures, the ocular lens and the thyroid gland, are in the imaging field. In addition, the test is not portable, cannot be performed in a typical eating environment, and requires patient's cooperation.

Scintigraphy

Scintigraphy examination is a nuclear medicine test, in which the person swallows a specific amount of radioactive bolus, and a gamma camera records the images during the swallowing. The advantage of this method, in addition to the identification of the silent aspiration of saliva, is its ability to quantify the aspiration and to track aspirated materials. Although the scintigraphy and VFS methods are complementary, scintigraphy is more sensitive for detecting aspiration below the glottis [26]. However, the exact cause of the disorder cannot be identified by using this technique. Scintigraphy is an expensive and invasive test.

Auscultation

Cervical auscultation has been used by clinicians and researchers who investigate acoustical patterns and temporal relations of the swallowing mechanism. It was shown that cervical auscultation can be used to identify the inspiration and expiration phases happens before/after the swallow, the moment of the pharyngeal swallow, the pharyngeal secretions in the airway before/ after the swallow, changes in the levels of secretion [27].

It was suggested that the respiratory pattern during swallowing is different for individuals with dysphagia. For example, a more variable respiratory pattern, a less consistent swallow apnea and the more frequent occurrence of inspiration after the swallow are observed in those patients [28],[29]. Thus, it was hypothesized that the clinical examination including cervical auscultation with a stethoscope would be able to distinguish patients who aspirate from those who do not [27]. However, swallowing assessment by auscultation highly depends on the examiner's skills [30]. Therefore, in recent years, acoustical analysis of swallowing has received considerable attention as a complementary tool for cervical auscultation.

Acoustical Analysis

Swallowing sound signals are listened and recorded by placing a microphone or an accelerometer on the subject's neck, either laterally or over the suprasternal notch of trachea. It is commonly known that the swallowing sound consists of two distinct segments: Initial Discrete Sound (IDS) and Bolus Transit Sound (BTS) [4]. IDS is related to the opening of the upper esophageal sphincter, while BTS is the gurgle sound associated with the transmission of the bolus into the esophagus. A Typical swallowing sound signal is shown in Fig. 2-6.

Swallowing occurs within the breath cycle. There is a deglutition apnea in breathing when swallowing happens. The microphone (accelerometer) collects both the swallowing sound and the breath sounds before and after the swallow. The swallowing and breath sounds are considerably different and have to be analyzed separately. The time-frequency characteristics of the breath and swallowing are different. The spectra of the swallowing sound signal contain higher frequency range compared to that of the breath sound. A typical waveform and spectrogram for a sequence of swallowing and breath sounds, recorded by a microphone placed over the suprasternal notch, is shown in Fig. 2-7.

Early acoustical swallowing sound analysis was mainly focused on the timing of the swallowing events [31]. Later, with the application of digital signal processing, it was used for diagnostic purposes, i.e. detecting swallowing disorders [5]. Some work has been done to automatically detect the swallowing sound segments from the breath sounds, as well as classifying normal and dysphagic swallowing sounds [32]. Different techniques including nonlinear dynamic analysis, recurrence quantification analysis (RQA) [32], hidden Markov modeling (HMM) [32], and multiresolution wavelet analysis [32] were applied to detect characteristic features of swallowing sounds. Other researchers [33], [34] investigated whether there are any differences between normal and dysphagic swallowing sounds.

This work of this thesis is built upon the knowledge gained by previous studies to enhance dysphagia and aspiration detection by acoustical analysis. The focus of previous studies were on swallowing segmentation and dysphagia prediction by swallowing sounds analysis; the focus of this work is on the

detection of aspiration in a record of swallowing and breath sounds as well as investigating the cause of swallowing disorder.

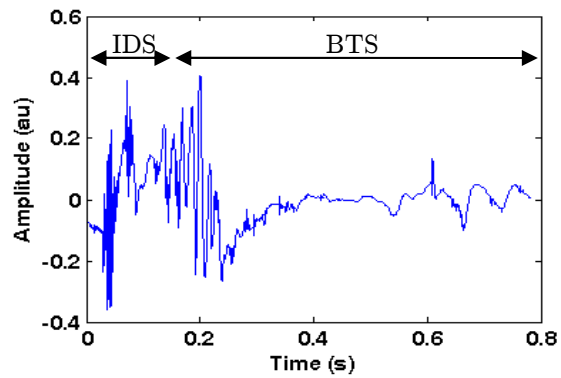


Fig. 2-6. A typical swallowing sound signal.

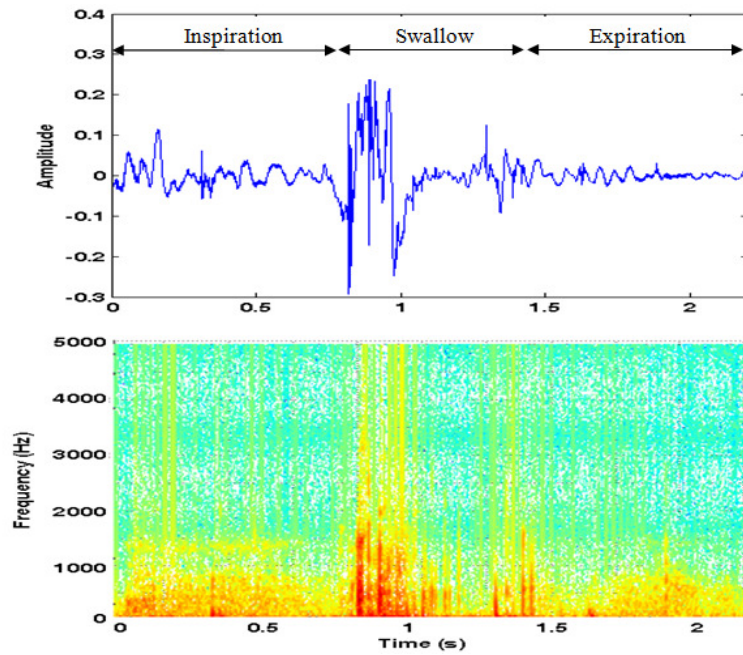


Fig. 2-7. (a) A typical swallowing sound signal with the before/after breath in time domain. (b) the spectrogram of the same signal.

Chapter 3- Acoustical Modeling of Swallowing Mechanism

3.1 Introduction

Swallowing sounds have been analyzed to find plausible differences between the control and dysphagic groups with a high accuracy [5], [35]. However, mechanism of swallowing sound generation and transmission has not been addressed adequately; thus, the exact mechanism of swallowing sounds has remained a challenge [36]. In one of our group member's previous works [36] a model was introduced in an attempt to explain swallowing sound generation. In that model, swallowing sounds were assumed to be produced by exciting the pharyngeal wall structure and tissue with an impulse train coming from the pharynx. Having assumed the excitation source and the pharyngeal wall structure are independent, the linear model shown in Fig. 3-1 was proposed to represent the swallowing sound generation.

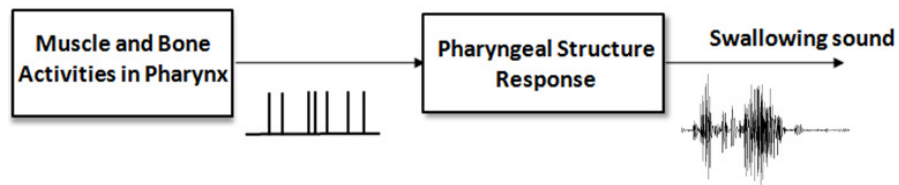


Fig. 3-1. Model of swallowing sound production [36].

Swallowing mechanism, as a sequential and rhythmic motor behavior is an interesting model for the neuro-physiological analysis of motor activities. However, because of the complexity of the motor pattern along with the great number of muscles and nerves involved, it has received less attention than other motor activities such as locomotion, mastication, or respiration [37]. In addition, it is difficult to initiate the swallowing sequence in anesthetized animals [37]. Understanding the underlying neural pathways that govern the act of swallowing is crucial in order to correctly identify swallowing disorders.

Basically, the swallowing mechanism initiates with a train of impulses coming from the brain cortices that produce a temporal summation on the brainstem swallowing center [38]. Consequently, the swallow is triggered via the cranial nerves, which directly control the muscles involved in the swallowing. In a normal swallow, if any residual material remains in the pharynx, it will be detected by the sensory receptors that initiate a second swallow to clear the pharynx.

As a pilot study for modeling the swallowing sound, it was attempted to find a mathematical model that describes the swallowing sound generation of the pharyngeal phase. The model, which is based on the model introduced in [36], proposes the components that may account for normal or abnormal swallowing sounds. This section is mainly focused on the model components and how they change as the characteristics of the swallowing sound change between individuals with and without swallowing disorders.

3.2 Method

Considering the swallowing sound as a result of a linear time invariant system excited by a train of impulses, pharyngeal swallowing sound modeling can be approached by two different assumptions. In one approach it is assumed the

impulse train, representing the neural activities to trigger swallow, is the same for both groups of healthy and dysphagic and it is the pharyngeal model that accounts for the difference between the two groups. On the other hand, in the second approach, it is assumed the pharyngeal response is the same for both groups but the neural activities to initiate the swallow are different between the two groups.

3.2.1 Data

Two different datasets of swallowing sounds were used for the pilot study. Each dataset included swallowing sound recordings of 5 dysphagic patients and 5 individuals without any swallowing disorder (controls). However, since the sensor and sampling rate of the two datasets were different, each dataset was analyzed separately. Dataset 1 was adopted from our previous study [7], in which swallowing sounds of controls and dysphagic (patients with cerebral palsy) were recorded by a Siemens (EMT25C) accelerometer placed over the suprasternal notch of the trachea and digitized at 10240 Hz. Dataset 2 included swallowing sounds recorded by a Sony (ECM-77B) microphone placed over suprasternal notch of trachea, and digitized at 44 kHz. The controls' data in dataset 2 was adopted from [36], while the dysphagics' data were recorded at Health Sciences Centre, Winnipeg, with the same microphone and sampling rate from stroke patients. The experimental protocol was the same for all data recordings: each participant was fed 5 boluses of a thin liquid texture (i.e. juice) with a 5ml spoon, resulting in 50 swallowing sounds for each dataset. The study has been approved by the Biomedical Ethics Board of the University of Manitoba and all participants or their legal guardian signed a written consent prior to experiments.

Since the sampling rates of the two datasets were different, the recorded sounds of dataset 2 were down-sampled by a factor of 4 to match those of dataset

1. All signals were band-pass filtered between 100-2500 Hz (as it was the filter's bandwidth of the adopted signals from [7], and normalized to their maximum. IDS and BTS parts of the swallowing sound signals were separated manually by an expert by auditory and visual inspection of the signals in the time and frequency domain. Only the IDS part of the swallowing sound, which is related to the pharyngeal phase, was considered in this study.

Modeling

The swallowing sound can be thought to be the output of a linear and time invariant (LTI) system representing the oropharynx muscle and tissue responses to the neural activities that initiate swallow, and are represented by an impulse train (Fig. 3-1). Since both the input (neural activities) and the transfer function (pharyngeal response) can affect the output (IDS), the modeling is approached by investigating the changes between the swallowing sounds of the two groups of control and dysphagic as a result of a change either in the pharyngeal response or the system input, each being considered separately.

Approach 1: Pharynx Transfer Function

In the first approach, it is assumed that the transfer function of the pharynx changes as the tissue characteristics vary between the individuals with and without dysphagia. In this approach, similar to the model suggested for the electromyogram (EMG) signals, the impulses are considered to be with the same amplitude, arriving at random intervals and having Poisson distribution [39].

Muscles involved in swallowing are driven by several motorneurons, through which the neural signals are transmitted to the muscles [37]. Similar to the EMG model in [39], it is assumed that the recorded swallowing sound signal is a weighted sum of the neural pulses (the impulse train) passing through the throat muscle fibers. The swallowing sound is assumed to be a weighted sum of the

impulse responses of each active muscle fiber's response, h_1, \dots, h_n , as depicted in Fig. 3-2.

For simplicity, it is assumed that the input impulse trains are uncorrelated. Also, each impulse train is considered as a renewal point process with identical interval statistics. In addition, we may safely assume the fiber types of the muscles involved in swallowing are of the same type. Therefore, if the number of active motoneurons is constant, n , then the initial part of swallowing sound signal (IDS) can be assumed to be a stationary process, and its power spectrum density (PSD) may be expressed as[39]:

$$\Gamma_{yy}(\omega) = NE\{k_j^2\} |H(j\omega)|^2 \Gamma_{xx}(\omega) \quad (3-1)$$

where k_j is the coefficients of the weighted sum, $E\{\}$ represents statistical mean, $H(j\omega)$ is the Fourier transform of $h(n)$, and $\Gamma_{xx}(\omega)$ is the PSD of the input impulse train.

If $\Gamma_{xx}(\omega)$ is relatively constant compared to $\Gamma_{yy}(\omega)$, it can be replaced by a constant term (k_a) and (3-1) can be rewritten as:

$$\Gamma_{yy}(\omega) = NE\{k_j^2\} |H(j\omega)|^2 k_a. \quad (3-2)$$

The above approximation used in (3-2) was shown to be valid regardless of the interval density of the incoming impulse train [39]. Moreover, the individual neural impulse trains may be considered to be Poisson point processes with the distribution parameter (mean value) of $\lambda = k_a$.

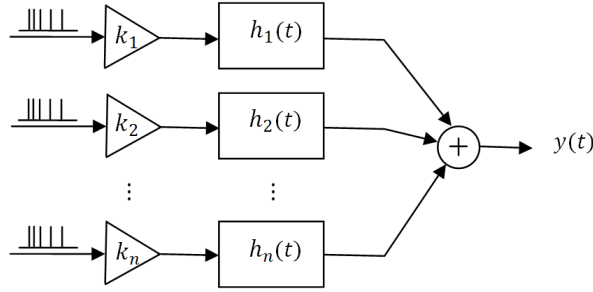


Fig. 3-2. The model [39] considered in the first approach to represent the IDS part of swallowing sound.

In this approach of modeling, the k_j coefficients represent the intensity of the neural impulses to initiate swallow. Given that the bolus size and texture are fixed in our data, we can simplify the model further and assume the $E\{k_j^2\}$ is constant, then $\Gamma_{yy}(\omega)$, representing the IDS segment of the swallowing sound is a scaled version of $H(j\omega)$. Therefore, $H(j\omega)$ (or equivalently $h(t)$) can be obtained by fitting a suitable analytical form to the PSD of the IDS segment of the swallowing sound signal $\Gamma_{yy}(\omega)$.

To evaluate the above model, the PSD (using Welch method [40] with 50% overlap between the successive windows) of the IDS segment of the swallowing sound was fitted using a Bode approximation with a resulting transfer function of

$$H(s) = G \frac{(s+a)^p}{(s+b)^q} \quad (3-3)$$

where a , b , and G are the zero of order p , pole of order q , and gain of the system, respectively.

To find the transfer function for each group of the control and dysphagic swallowing sounds, the Bode approximation as presented in (3-3) was fitted to the IDS segment of every swallowing sound, and the Bode parameters were averaged among the signals of every subject in each study group. The swallowing sound characteristics do vary among the individuals and within each one due to

different bolus size and texture. In this study we have kept the bolus size and texture the same for all study subjects. Furthermore, we fed every study subject in the same manner in order to keep the pharynx and upper airway posture the same for everyone. Hence, we assumed that a specific model may represent the swallowing sounds of healthy and dysphagic groups. Figure 3-3 shows the PSD of a typical swallowing sound of the control group along with its fitted Bode curve. The same plot for a typical dysphagic swallowing sound is shown in Fig. 3-4.

The time-domain transfer function signals were found by calculating the inverse Fourier Transform of the approximated transfer functions found for each group of data (Fig. 3-5).

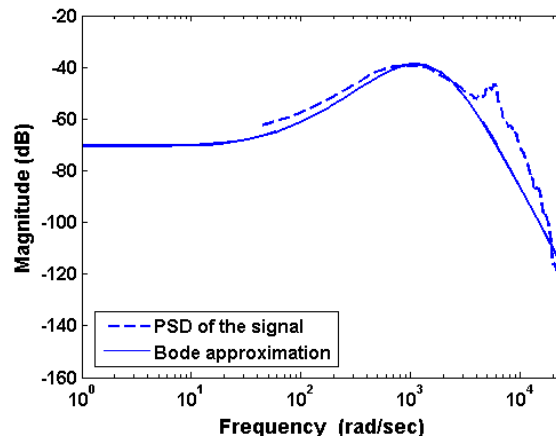


Fig. 3-3. PSD and its fitted Bode plot of the IDS segment of the swallowing sound of a control subject.

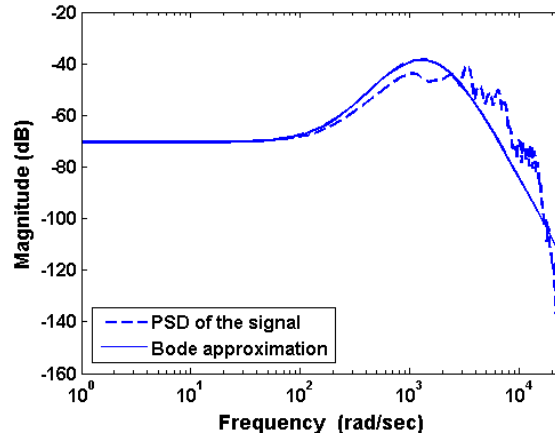


Fig. 3-4. PSD and its fitted Bode plot of the IDS segment of the swallowing sound of a dysphagic subject.

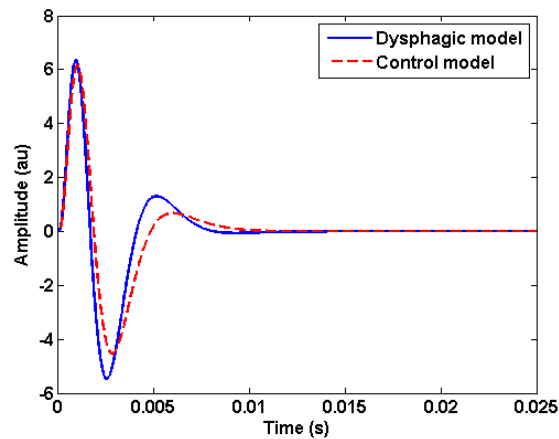


Fig. 3-5. The time domain response of the Bode approximation models derived for the control (—) and dysphagic (--) groups.

To validate the fit of the models for each group of data set, the mean-square-error (MSE) was calculated. To determine MSE, the error between the actual PSD of the IDS segment and the corresponding values of the Bode curve was calculated, squared, added up for all data points, and divided by the number of points. The smaller the MSE, the closer the fitted curve is to the data.

Approach 2: Pharynx Neural Inputs

In this approach, it was hypothesized that the transfer function of the pharyngeal walls to the initiating neural pulses is the same for both normal and dysphagic swallow but it is the neural pulses, i.e. the input of the system that accounts for the difference between the two groups. Therefore, it was assumed that the pharyngeal response (the system transfer function) can be suitably represented by a wavelet basis function at a fixed pre-defined level. The selection of the type of the basis function is described below.

To investigate the above hypothesis wavelet decomposition was applied to the IDS parts of the swallowing sounds of both groups. In wavelet analysis a basic finite function is defined as the basis function for describing a set of functions or signals. Then, the signal is described in some scales by which the components of the signal in the subspace spanned by the specified basis function [41].

$$f(t) = \sum_{k=-\infty}^{\infty} c_{j_0}(k)\phi_{j_0,k}(t) + \sum_{k=-\infty}^{\infty} \sum_{j=j_0}^{\infty} d_j(k)\psi_{j,k}(t) \quad (3-4)$$

Where c_{j_0} and d_j are the approximation and detail coefficients, respectively. The first term shows the components of the signal $f(t)$ in the subspace spanned by $\phi_{j,k}$ and the second term represents the detailed components of the signal in higher scales.

In the wavelet analysis, the discrete wavelet approximation coefficients at a specific level of decomposition were examined. The Symlet wavelet of order 8 was chosen to be consistent with previous studies [36]. It is speculated that the energy of the coefficients at the specific level can mimic the impulses applying to the system.

To find the level of decomposition that can best distinguish between the healthy and dysphagic swallows, a simple classification has been done. For each swallowing sound, the energy of the approximation coefficients was calculated at 8 levels of decomposition. The best level was chosen according to the specificity and sensitivity of the classification calculated for each level.

3.3 Results

In the first approach, different models were obtained for the control and dysphagic subjects. The control group's model was approximated by a second order zero at 16 Hz, and a pole of the order 7 at 255 Hz. On the other hand, the dysphagic group's model was approximated by a third order zero at 32 Hz and a pole of the order 6 at 230 Hz. Both models shared the same gain. To evaluate how well each model represented its assigned group (control/ dysphagic) the Mean Square Error (MSE) between the actual PSD of the signal and each of the two Bode approximated models were calculated for each of the swallowing sounds. Therefore, two sets of MSE values were obtained for each signal: the MSE of the signal in its own model and the MSE in the other model (Fig. 3-6).

As can be seen in Fig. 3-6, the MSE of the control individuals in the control model is less than that in the dysphagic model (the MSE lie in the left side of the bisector line). This result was desired; however, it was not consistent for the dysphagic group, i.e., the points were spread in the region of plot, in which the MSE of the dysphagic data in the control model is less than that in the dysphagic model. In other words, if the models were the best candidates to represent their own data, the MSE plot would have been perfectly separated by the bisector line.

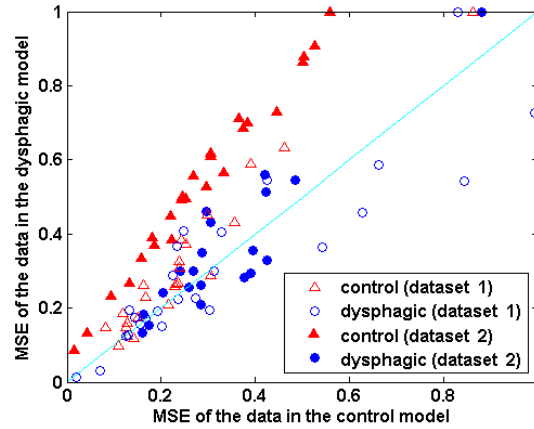


Fig. 3-6. The Mean Square Error (MSE) between the PSD of each swallowing sound signal of control/dysphagic subjects (from both the datasets) and the Bode approximated transfer function for control/dysphagic models as shown in x/y axes, respectively (The bisector line of the plane is plotted for better presentation of the results).

As for the second approach using the wavelet analysis, the approximation coefficients for all levels of decomposition were calculated; they are shown in Fig. 3-7 and Fig. 3-8 for the control and dysphagic signals, respectively. To determine the scale, at which the wavelet basis function describes the pharyngeal response of the swallowing sound the best, we used the ability of the model to classify the two control and dysphagic groups as the criterion. A simple linear classification technique was used as the linear ones are more robust in heterogeneous populations. At each level of decomposition, the mean values of the energy of the wavelet approximation coefficients were calculated for each group, the average of which sets a threshold for a simple classifier for that level. The classification results are shown in Table 3-1. The accuracy is determined by calculating the percentage of correctly classified control and dysphagic data according to:

$$\text{accuracy} = \frac{\text{true control} + \text{true dysphagic}}{\text{total number of control and dysphagic}}$$

To further evaluate the performance of each level in representing the pharyngeal response, the specificity and sensitivity of the classifier were

calculated as follows:

$$\text{specificity} = \frac{\text{true dysphagic}}{\text{true dysphagic} + \text{false dysphagic}}$$

$$\text{sensitivity} = \frac{\text{true control}}{\text{true control} + \text{false control}}$$

As demonstrated in Table 3-1, the 2nd and the 3rd level coefficients can discriminate between the swallows of the control and dysphagic groups with higher accuracy in comparison to the other levels. Therefore, they were chosen for further investigation.

Table 3-1. The classification results at 8 levels of decomposition.

level	Accuracy		Specificity		Sensitivity	
	D1	D2	D1	D2	D1	D2
1	78%	84.44%	71.88%	76%	88.89%	95%
2	80%	84.44%	74.19%	76%	89.47%	95%
3	82%	84.44%	75%	76%	94.44%	95%
4	80%	80%	72.73%	72%	94.12%	95%
5	76%	68.89%	67.57%	61.54%	100%	78.95%
6	66%	71.11%	59.52%	64%	100%	80%
7	64%	51.11%	58.14%	47.06%	100%	63.64%
8	64%	57.78%	58.14%	51.43%	100%	80%

The energy of the best scale coefficients was then calculated for all swallows of each individual. To display the classification results for all swallows of the two groups, the energy of the 3rd level coefficients were plotted versus that of the 2nd level coefficients for both datasets (Fig. 3-9). The values were averaged among the swallows for each subject. Since two different datasets were used for this

study, their energy levels were different; hence, to show both datasets on the same plot, we normalized the values in each dataset with respect to the maximum value within the same dataset.

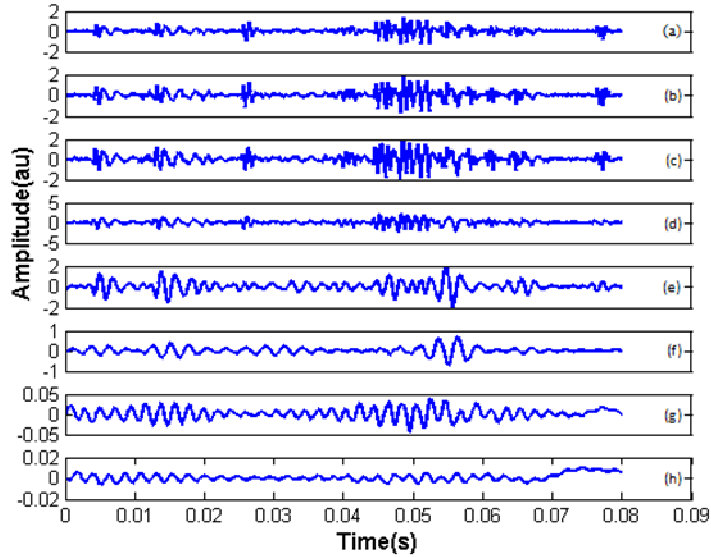


Fig. 3-7. The approximation coefficients at 8 scales (a)-(h) for a typical swallowing sound of a non-dysphagic subject.

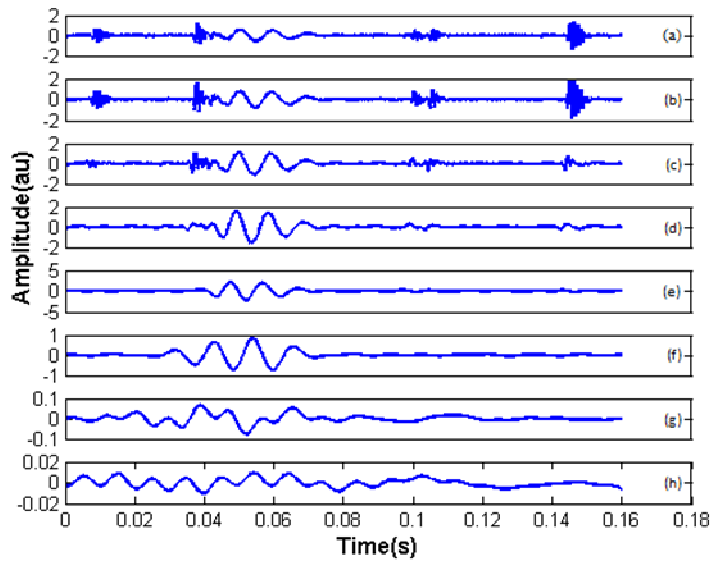


Fig. 3-8. The approximation coefficients at 8 scales (a)-(h) for a typical swallowing sound of a dysphagic subject.

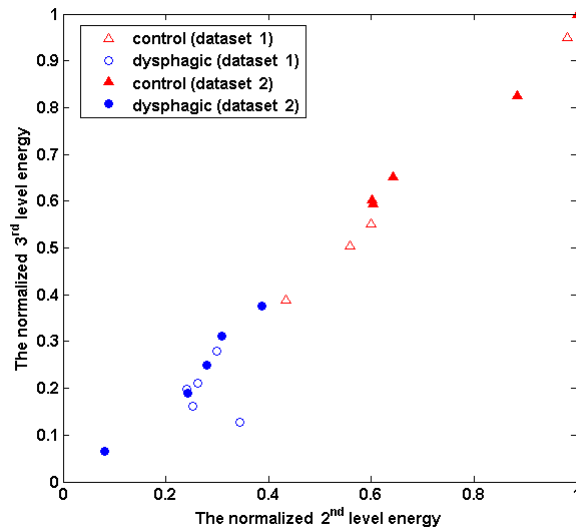


Fig. 3-9. The scatter plot of the energy of the control and dysphagic data at two levels of wavelet decomposition averaged among each subject's swallows. The values of x-axis and y-axis belong to the 2nd and 3rd level, respectively. The filled marker shows the values of the first dataset while the blank ones represent the second.

3.4 Discussion

In this part of study, based on the assumption that swallowing sound is a result of a linear time invariant system excited by a train of impulses, two hypotheses for modeling the swallowing sounds were investigated. In the first hypothesis it is assumed the neural activities are the same and it is the pharyngeal response (model) that is different between the two groups of the swallowing sounds. On the other hand, in the second hypothesis it is assumed the impulse train representing the neural activities to trigger swallow is the same for both groups of healthy and dysphagic and it is the pharyngeal model that accounts for the difference between the two groups.

In the first hypothesis it is assumed the neural activities are the same for all swallows despite the potential disorder, and it is the pharyngeal response (the model) that is different between the two groups of the healthy and dysphagic swallowing sounds. On the other hand, in the second hypothesis it is assumed the

pharyngeal response is the same for both groups but the neural activities to initiate the swallow are different between the two groups.

To investigate the first hypothesis, the pharyngeal tissue response to the neural activities was modeled by a linear time invariant model. The model was derived by fitting a single zero and pole with multiple orders Bode approximation to the PSD of the IDS part of the swallowing sound signals in each group. As a result of the Bode fitting, two different models were obtained for the healthy and dysphagic groups.

In a previous study, swallowing sound was assumed to be generated as the result of convolving two signals in time domain: a train of impulses as the neural activities and the linear system that represents the pharyngeal wall. To separate the convolutional components from the resultant signal, cepstrum analysis was used [36]. By applying a low-pass filter to the cepstrum result, the low frequency component which resembles the model of the pharyngeal response was obtained and called the principal curve. The time domain response of the pharyngeal system, shown in Fig. 3-5, is similar to the principal curve introduced in [36].

The idea of fitting a Bode curve to the PSD of the signal was borrowed from [39] in modeling electromyogram (EMG) signals. The EMG model was considered to be composed of the neural pulse train inputs passing through an LTI system representing the motor unit action potential. In spite of the same description for the two biological signals, EMG and swallowing sound, the outcome was not similar. While the method was successfully applied in EMG modeling, it was not found to be a good tool for modeling swallowing sound.

The results of the first approach, shown in Fig. 3-6, imply that some dysphagic swallows were fit better by the normal model than their own model. Hence, one may question the validity of the first hypothesis at least for this

group of dysphagic patients that their data was used in this study. Given the fact that the participants in our dysphagic group were all cerebral palsy and stroke patients, the second hypothesis that considers a deficiency in neural activities for swallow initiation in dysphagic group, sounds more reasonable.

The wavelet analysis may provide another piece of the modeling puzzle relating to the role of the neural activities. The main idea behind the wavelet analysis in this study was to find an estimate of the impulse train input to the swallowing model. Since the principal curve was similar to the Symlet wavelet of the 8th order, the symlet8 wavelet was used for the wavelet analysis in [36] and the present study. The timing and amplitude of the arriving impulses are inherent in the wavelet coefficient at a particular order of the decomposition. Choosing the 3rd level is shown to yield the best classification results between the two groups; the energy of the 3rd level wavelet approximation coefficients was lower for the dysphagic individuals than those for the control subjects.

It is interesting to note that the 3rd level of the decomposition, which yielded the best classification result, is in agreement with the best features obtained in [34]. It was shown that the frequency domain features, extracted from the IDS segment, that were corresponding to the frequencies below 700 Hz, represented the main characteristics of the signal [34]. The frequency range of the dyadic wavelet transform corresponding to the 3rd level wavelet resolution can be obtained by dividing the Nyquist frequency by 2^3 which is calculated as 640 in this study.

Swallow, as a sequential and rhythmic motor behavior, can be thought to be the result of a neuro-physiological model involving all levels of the nervous system; it can be triggered by stimulating a peripheral nerve and also involves some regions of the central nervous system. The model is stimulated by a pattern

of excitation in pharyngeal and laryngeal muscles appeared as seen in the form of the spikes in EMG studies of the swallowing sounds [35]. The exact role of the central mechanisms in generating the impulses is still unknown. However, the timing and amplitude of the impulses have been found to be responsible for the sound generation in swallowing [37].

The role of the input signal in the model of the swallowing can be highlighted in cerebral palsy (CP) disorder which was the case in the patient group of this study. It has been shown that cortical dysfunction, as in CP, may result in swallowing impairment [28]. In CP patients, a major output pathway from the motor cortex to the brainstem and cranial nerve is damaged. Also, the damage can occur in the neural circuitry that mediates the processing of the signals coming from the anatomic structures involved in swallowing such as oral cavity, pharynx, and esophagus. These signals contain the information regarding the characteristics of a bolus as it passes through the swallowing pathway. Therefore, either types of damage influence the excitation signal that stimulates the muscles. This can be due to the improper information conveyed to the structures controlling swallowing or the impaired pathway from the control center.

It can be speculated that both assumptions for modeling the swallowing mechanism of the two groups of control and dysphagic are valid depending on the type of dysphagia. As the patients participating in this study suffered from CP, the wavelet modeling seems to be a better representation of the swallowing sound generation. These individuals lack a controlled, coordinated swallow due to neuromotor impairments that may interfere with efficient food processing, delays in initiating swallows, and inefficient oropharyngeal clearance of the swallowed material [7]. Therefore, the model in the second approach that takes into account the effect of the input is congruent with physiological cause of dysphagia that occurs in conjunction with the neurologic problem such as CP. However, if

dysphagia is due to muscle atrophy or degeneration, the first model may be a better fit to the swallowing mechanism.

The proposed modeling and the model evaluation in this part of the study were the first attempts toward modeling the swallowing mechanism. Moreover, the statistical properties of the swallowing sound are discussed in this study. This knowledge is important to select an appropriate type of the system (i.e. linear vs. nonlinear) for modeling. For this purpose, the gaussuanity and nonlinearity of the system are investigated and discussed in Appendix A.

3.5 Summary

In the pilot study presented in this chapter, a mathematical modeling of the swallowing sound generation is presented. To evaluate the model, its application on swallowing disorder diagnosis is discussed. As a starting point, a simple linear time invariant model is assumed to represent the pharyngeal wall and tissue excited by a train of impulses. The modeling is approached by two different assumptions. In one approach it is assumed the impulse train, representing the neural activities to trigger swallow, is the same for both groups of control and dysphagic and it is the pharyngeal model that accounts for the difference between the two groups. On the other hand, in the second approach, it is assumed the pharyngeal response is the same for both groups but the neural activities to initiate the swallow are different between the two groups. The results show the second approach complies better with the physiological characteristics of swallowing mechanism as it provides a much better discrimination between the swallowing sounds of control and dysphagic groups of this study. Though, it should be noted that our dysphagic group subjects were cerebral palsy and stroke patients. Hence, the model accounting for initiation of neural activities is reasonable to show better results.

Chapter 4- Non-invasive and Automatic Detection of Patients at High Risk of Swallowing Aspiration

4.1 Introduction

Aspiration represents a problem in the airway protection during the swallow. Swallowing and respiration are reciprocal functions, in which many muscles and structures have dual roles. Thus, a safe swallow requires a highly coordinated interaction with respiration.

In most normal individuals, expiration occurs at the end of the swallowing, and clears the larynx and pharynx from small residues of materials. Expiration after swallowing may help clear the airway from any material that may have penetrated [15]. It was shown that the number of swallows followed by inspiration increases as the level of consciousness decreases [7], [20]. This observation emphasizes the role of the central nervous system in controlling the breath-swallow interaction. The central nervous system modulates the timing and coordination of the interaction using the feedback sensory information about the presence of the food. Thus, neurological disorders, which affect the pathways involved in the swallow-breathing coordination, may increase the risk of aspiration as the result of changing the aforementioned pattern. For example, dysphagic individuals often show interruption in the inhalation phase, which may lead to aspiration.

When intervening with an individual with dysphagia, the first goal is to determine whether they are aspirating. Acoustical detection of aspiration has clinical value as a large population of the patients demonstrates silent aspiration, which does not show any clinical indications such as coughing or choking.

Identifying dysphagic individuals does not draw attention as a diagnostic tool for speech-language pathologists (SLP) or physicians as they make the diagnosis by using the clinical symptoms and a quick examination. What is important for clinicians is the accurate aspiration detection. Thus, the rest of the thesis is dedicated to aspiration detection by analyzing the breath sounds after a swallowing event.

In this chapter, we investigated the application of acoustical analysis of breath and swallowing sounds for identifying patients at high risk of aspiration [11]. We hypothesized that the existence of a small particle of food or liquid in the airway (due to aspiration) would change the pattern of breathing airflow by creating vortices (Fig. 4-1). Thus, we proposed a novel method based on phase-space analysis of breath sounds immediately after swallows followed by support vector machine classifier (SVM) [42] as a diagnostic aid for identifying patients with high risk of aspiration.

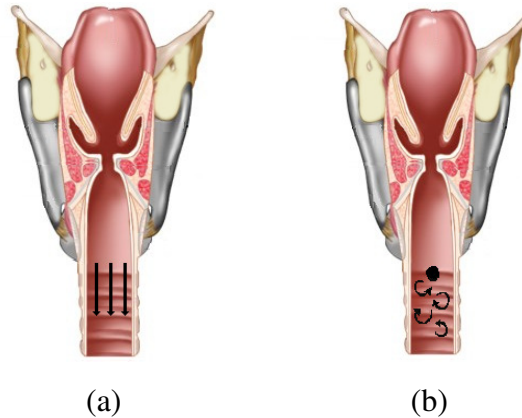


Fig. 4-1. An illustration of the pattern of airflow in (a) a clear trachea and, (b) a trachea with a food particle in it.

4.2 Methods

4.2.1 Study Subjects

Data were collected from 50 dysphagic adult patients (age 60 ± 20 years, 11 females), who were referred to either VFS or FEES assessment as part of their diagnostic routine. Individuals with stroke, acquired brain injury and neurodegenerative disorders such as Parkinson, Huntington and Amyotrophic Lateral Sclerosis (ALS) were included. The study participants were fed different types of a variety of solid and liquid food, which were presented by a cup, spoon, or straw. The type and order of the food were decided by the speech-language pathologist. The recording was finished as soon as enough information was obtained about the swallowing mechanism of the participant based on the speech pathologist's judgment.

Swallowing and breath sounds were recorded simultaneously with either VFS or FEES assessment at DeerLodge Centre, Riverview Health Centre, and Health Sciences Center, Winnipeg, Canada. The study was approved by the Biomedical Research Ethics Board of the University of Manitoba, and all participants gave a written consent prior to the experiments. The speech-pathologist marked the

patient's aspiration events using either VFS or FEES assessment. We considered patients as severely dysphagic if they aspirated in more than 50% or more than 5 of their swallows during the swallowing assessment; 11 of the 50 dysphagic participants were in severely dysphagic category. Note that the aspirations were all identified by the speech pathologist in charge of imaging assessment independent of our analysis.

4.2.2 Data Recording

Tracheal sounds (breath and swallowing sounds) were recorded by a Sony (ECM-77B) microphone placed over the suprasternal notch of the trachea using double-sided tape. The sounds were recorded at 44.1 kHz sampling rate by a digital sound recorder (EDIROL R-44). The patients were fed different types of a variety of solid and liquid food decided by the SLP, who performed the VFS or FEES assessment. As the microphone collects both the swallowing and the breath sound signals, we separated the breath and swallowing signals by aural and visual examination of the signals in the time and frequency domain. Our group has studied and developed several automatic algorithms for breath and swallowing sounds separation; the most accurate one [43] achieved 93% accuracy. However, our recorded sounds included swallowing and breath sounds, the SLP's voice and also some other noise in the hospital that needed to be excluded from the segmented data prior to analysis. Therefore, to assure 100% accuracy at this stage in this study, we used manual separation. The breath sounds of the 1 to 3 breath phases (inspiration/expiration) immediately after each swallow and before the next swallow were selected for analysis. All the signals were band-pass filtered between 100-5000 Hz.

4.2.3 Signal Analysis

Inspired by the idea in “phase-space thresholding” used in the acoustical Doppler velocimetry [44]-[45], we plotted the time-domain breath sound data of each respiratory phase (after a swallow) versus its first and second derivatives, and fit an ellipsoid to encompass the data points. Then, we calculated the ellipsoid’s principal axes, and determined the data points of the time-domain breath sound data that fell outside the ellipsoid. The sum of the distances of those points from the center of ellipsoid was calculated, and normalized by the total energy of the signal in the time-domain. This was repeated for the 1 to 3 available breath phases sounds immediately after each swallow and before the next swallow. We considered the mean value of the distance of the data points outside the fitted ellipsoid as a characteristic feature for classification and testing our hypothesis as whether a foreign particle in the airway would change the flow of air and consequently change the generated breath sounds. The detail of the feature calculation is explained below.

The first and second derivatives of the time-domain signal (Δx_i and $\Delta^2 x_i$) were calculated as:

$$\Delta x_i = (x_{i+1} - x_{i-1}) / 2, \quad (4-1)$$

$$\Delta^2 x_i = (\Delta x_{i+1} - \Delta x_{i-1}) / 2. \quad (4-2)$$

The center of the ellipsoid, which represents the mean values of the time-domain signal, and its velocity and acceleration, is located at the origin of the phase-space domain after the mean values of each axis are removed. The linear least-square regression method [45] was used to calculate the rotation angles of ellipsoid’s principal axes. For Δx_i versus x_i and $\Delta^2 x_i$ versus Δx_i , the rotation angles of the principal axes are zero because of the symmetry in their

distribution. However, for x_i versus $\Delta^2 x_i$ the rotation angle of the principal axis is given by:

$$\alpha = \tan^{-1} \left(\frac{N \sum x_i \Delta^2 x_i - \sum x_i \sum \Delta^2 x_i}{N \sum x_i^2 - (\sum x_i)^2} \right), \quad (4-3)$$

where α is the rotation angle of the principal angle in $x-\Delta^2 x$ plane. Considering that the mean values of the time-domain signal is zero, the above equation simplifies to:

$$\alpha = \tan^{-1} (\sum x_i \Delta^2 x_i / \sum x_i^2). \quad (4-4)$$

The principal axes of the ellipse projected into the $x-\Delta^2 x$ plane, are a and b , as determined by equations (4-5) and (4-6), respectively; c is the major axis in $\Delta x-\Delta^2 x$ plane.

$$a = \sqrt{\frac{(\lambda \sigma_x \cos \alpha)^2 - (\lambda \sigma_{\Delta^2 x} \sin \alpha)^2}{\cos^4 \alpha - \sin^4 \alpha}}, \quad (4-5)$$

$$b = \sqrt{\frac{(\lambda \sigma_{\Delta^2 x} \cos \alpha)^2 - (\lambda \sigma_x \sin \alpha)^2}{\cos^4 \alpha - \sin^4 \alpha}}, \quad (4-6)$$

$$c = \lambda \sigma_{\Delta x}, \quad (4-7)$$

where $\sigma_{\Delta x}$ and $\sigma_{\Delta^2 x}$ are the standard deviations of the first derivative and second derivative respectively and λ is defined as $\lambda = \sqrt{2 \ln(N)}$, where N is the number of data points. In spherical coordinate system the ellipsoid is defined as:

$$\frac{1}{r^2} = \frac{(\sin \phi \cos \theta \cos \alpha + \cos \phi \sin \alpha)^2}{a^2} + \frac{(\sin \phi \cos \theta \sin \alpha + \cos \phi \cos \alpha)^2}{b^2} + \frac{(\sin \phi \sin \theta)^2}{c^2}, \quad (4-8)$$

where r , θ and ϕ are spherical coordinate, as shown in Fig. 4-2.

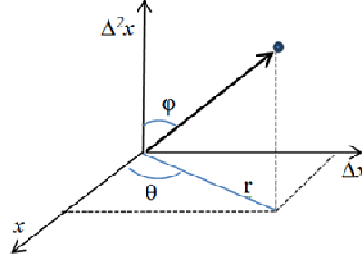


Fig. 4-2. Spherical coordinate in the phase-space.

4.3 Classification

We used the SVM classifier to evaluate the classification performance. SVM finds the decision boundary by minimizing the objective function in (4-9), subject to the constraint shown in (4-10):

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i^N \xi_i \quad (4-9)$$

$$Y_i(\mathbf{w}^T \mathbf{X}_i + b) \geq 1 - \xi_i \text{ for } i = 1, \dots, N; \quad \xi_i \geq 0, \quad (4-10)$$

where \mathbf{X}_i is the feature vector, Y_i is the label of the class, \mathbf{w} is the weight vector, b is the bias, C is the regularization parameter and ξ_i is called the slack variable, which measures how much the constraint is violated.

To remove any bias between training and testing, we used leave-one-out routine for testing the classification, in which data of all subjects—except one—were selected as the training data, and the left-out data was used as the test data; this routine was repeated till all subjects' data was used as a test data once. The training and testing accuracies were then calculated.

4.4 Results

Figure 4-3 shows the phase-space plots of the breath signals following the swallows of two typical dysphagic patients: one with severe aspiration and one without aspiration. As expected from our hypothesis, two distinct patterns are observed (Fig. 4-3). For the breath sound of the patient without aspiration, the majority of the data points can be encompassed by an ellipsoid. On the other hand, for the patients with severe aspiration more variations occur in the breath sound's phase-space, which results in more points outside the ellipsoid. The corresponding time-domain sound signals of Fig. 4-3 are shown in Fig. 4-4, where the points outside the ellipsoid are marked with red color to show the time occurrence of those points. Figures 4-4 (a) and (b) show the signal in the time domain. The variations of the signal (the first derivative) against the signal and time are shown in Fig. 4-4 (c) and (d).

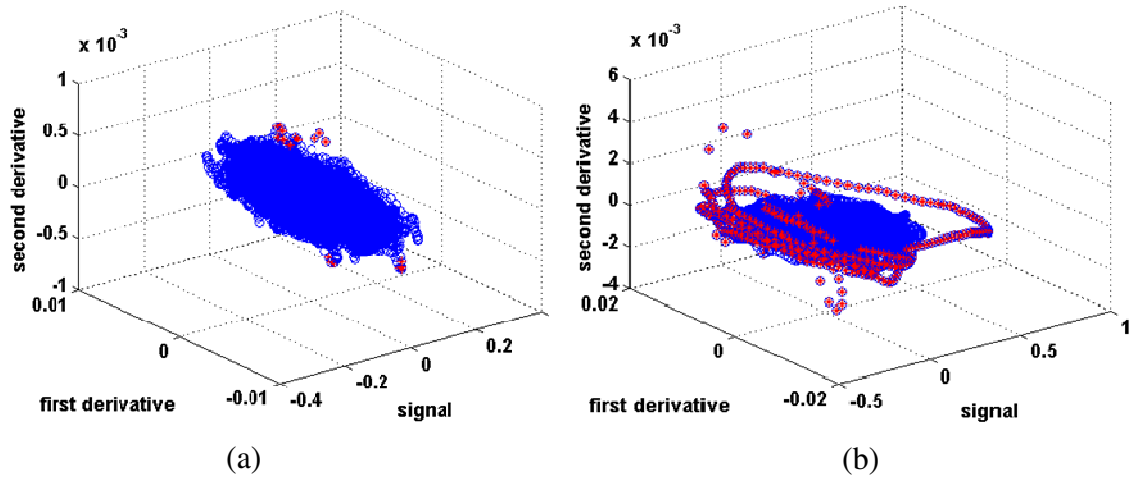


Fig. 4-3. The 3-D phase-space plot of the two breath signals (following swallows) of two patients: (a) without, and (b) with aspiration. The asterisks show the points that fall outside the ellipsoid.

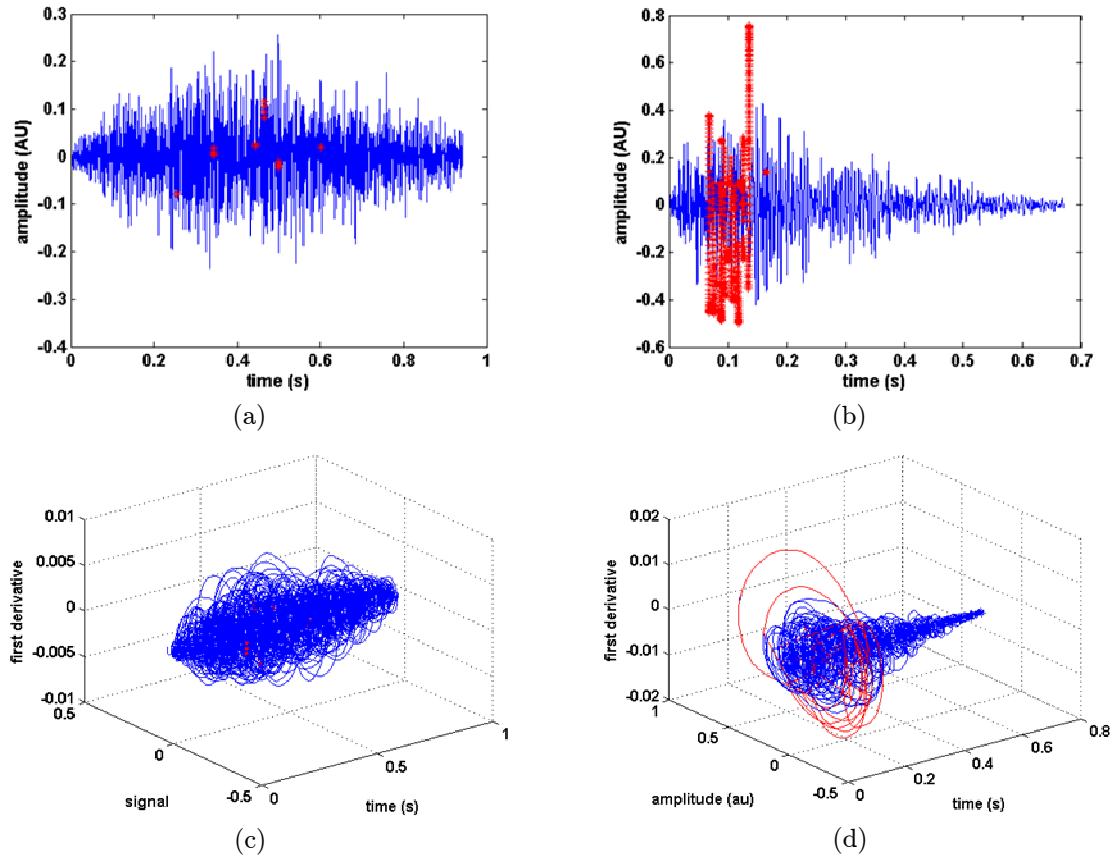


Fig. 4-4. The time-domain breath sound signals corresponding to the graphs shown in Fig. 4-3: (a), (b) without, and (c) and (d) with major aspiration. Data points that fell outside the ellipsoid are shown in red.

Figure 4-5 shows the scattered plot of the characteristic feature (sum of data points outside the ellipsoid) calculated for all the study subjects. The SVM classification resulted in 86% accuracy on the test dataset with 91% sensitivity (one false negative) and 85% specificity (6 false positive). The false positive cases are discussed in the discussion. Since we used leave-on-out routine, for every fold there was a training accuracy; thus we averaged them over the entire folds. The training accuracy, sensitivity and specificity were found to be $86\% \pm 0.01$, $91\% \pm 0.02$ and $84\% \pm 0.0$, respectively. As can be seen the average training accuracy is very close to test accuracy; that implies the robustness of the

classification results; it is also congruent with the observation that the standard deviation of the training accuracy between the folds is very small.

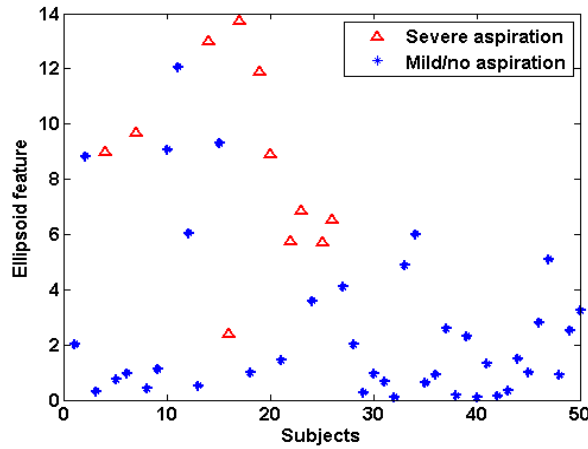


Fig. 4-5. Scatter plot of the ellipsoid feature for all the participants. The triangle marker (Δ) represents the patients in the severe aspiration group, and the asterisk (*) shows the group of patients with mild or no aspiration.

4.5 Discussion

We propose a novel acoustic swallowing assessment to identify patients at high risk of aspiration based on analysis of the breath sounds immediately after a swallow. The classification results show a high sensitivity (91%) and a reasonable specificity (85%) on the test dataset which was independent of the training dataset.

Since any foreign particle in the airway (i.e. due to aspiration) can change the flow of air, we studied the fluid mechanical characteristics of the air flowing in the trachea during breathing after each swallow. As mentioned in Introduction section, the trachea can be modeled as a tube through which the airflow passes [46]. The airflow in the trachea is highly likely to be turbulent [47]. Therefore, we hypothesize that any small particle in the airway as a result of aspiration (Fig. 4-1) would affect the size of the turbulence large scales; that should be reflected in the breath sound signals. Based on this theory and hypothesis, we expect the

existence of a foreign particle in the airway would cause a rattling sound during breathing. This analogy shows to be correct, when one considers the time occurrence of the points outside the ellipsoid of time-domain signal versus its velocity and acceleration. As can be seen in Fig. 4-3 and Fig. 4-4, those data points outside the ellipsoid are all close to each other in time; implying the rattling sound in that time period due to the foreign particle in the airway. In addition, Fig. 4-4(d) shows the decaying of the rattling sound's amplitude by time as the airflow passes and moves away from the particle in the trachea.

In our previous study [12] with a dataset of 21 patients, we developed a two-stage classification method for the same purpose as in this study. We tested the proposed method of our previous study on the dataset of this study as well; the two-stage classification on the dataset of this study showed 91% accuracy (1 of false positive and 1 of false negative) at stage 1, when identifying the aspirated patients, and 81% accuracy at stage 2 when identifying the aspirated swallows. These results of stage 1 are comparable with the accuracies achieved by the new proposed method, while the new method has the advantage of achieving the same accuracy or better with only one feature. The fact that the averaged training accuracy was similar to test accuracy is very encouraging as it implies the robustness of the classification results.

Our proposed method was able to identify all patients except one correctly but had 6 false positives. It is of interest to consider the conditions of false positives further. One of the false positive cases was a patient, who did not show any aspiration in his VFS but he had continuous involuntary head movements as a result of the Huntington's disease; that might have affected the sound recording. The other false positive cases were patients with mild aspiration based on their FEES/VFS assessments. However, it is possible that some aspirations were

missed by the FEES/VFS imaging during the FEES white-out period (when endoscope's tip contacts the swallowing structures) or when there was a pause in the VFS recording. Therefore, these 4 patients might have been indeed at high risk of aspiration (severely dysphagic) as our method predicts.

Other than our group's studies, to the best of our knowledge, there has not been any other study analyzing the tracheal breathing and swallowing sounds to particularly identify patients at risk of aspiration. Another group of researchers have used the accelerometer data to classify the healthy and abnormal swallows, which also included the swallows with aspiration; they used Euclidean linear discriminant classifier with an accuracy of 74.7% for aspiration detection on a population of 24 patients [48]. However, they did not report the sensitivity and specificity of their method. Our results of acoustic analysis of tracheal sounds significantly outperform the vibrational analysis of trachea during and after swallowing. Vibrational analysis only reflects the movements of the larynx, while the sounds are pseudo-pressure signals, and thus, representative of the physiology of the airway.

The modeling algorithm described in Chapter 3 was not applied to new data mainly because the new data were recorded with the aim of breath sound analysis. For that purpose, we had to increase the gain of the amplifier to improve the sensitivity of recording breath sounds at the cost of having the swallowing sounds mostly clipped.

The main contribution of this part of the thesis is proposing a novel, fast and robust acoustic analysis method that can identify patients with high risk aspiration. The method's main novelty is in its feature extraction that represents the physiology of aspiration. Using the proposed method, one can develop a device that will simplify swallowing clinical assessment significantly. Given the

non-invasive nature of the acoustic swallowing assessment, and also that it can be readily applied during a natural meal setting at the bed side of the patient, it would be very useful and economical to be utilized as an screening before referring the patient for either VFS or FEES assessment. Therefore, the patients' referrals to the VFS/FEES tests can be prioritized; thus reducing the health care cost significantly.

4.6 Summary

In this chapter, we investigated the application of acoustical analysis of breathing and swallowing sounds for identifying patients at high risk of aspiration. We propose a novel method based on phase-space analysis of breath sounds immediately after swallows followed by support vector machine classifier (SVM) as a diagnostic aid for identifying patients with high risk of aspiration. The classification result of the proposed method was compared with the FEES/VFS assessments provided by the speech-language pathologists; it showed 91% sensitivity and 85% specificity in detection of patients with severe aspiration (high risk dysphagia). The results are promising to suggest the proposed phase-space acoustical analysis method as a quick and non-invasive screening clinical tool to detect patients developing severe aspiration.

Chapter 5- Detection of Swallows with Silent Aspiration using the Swallowing and Breath Sounds Analysis

In this part of the study, we investigated the feasibility of detecting silent aspiration (those without any coughing or throat clearing after the aspiration) by sound analysis of the breath sounds immediately after a swallowing event. Theoretically speaking, the existence of an external particle in the airway, as a result of aspiration must change the sound of flow turbulence during breathing as shown in Fig. 4-1. Hence, we hypothesize that silent aspiration can be detected by sound analysis of the breath sounds immediately after swallowing event with aspiration.

To test our hypothesis, we recorded the breathing and swallowing sounds of dysphagic individuals, who were referred for the VFS or FEES swallowing assessment, simultaneously with VFS/FEES imaging. Our goal was to detect silent aspirations; thus, we excluded the aspirated swallows that were followed by coughing. We analyzed data of the patients who had at least 2 swallows with silent aspiration confirmed with either VFS or FEES. We analyzed the breath sounds immediately after the swallowing events, derived the characteristic features, and used a fuzzy clustering to classify the aspirated swallows from non-aspirated ones. The results were compared and validated using the VFS/FEES

assessment. The following sections detail the acoustical analysis method and discuss the results.

5.1 Method

5.1.1 Participants

Data were collected from 45 dysphagic patients (10 females), who suffered from stroke, and neurodegenerative disorders such as Parkinson, Huntington and Amyotrophic Lateral Sclerosis (ALS), or traumatic brain injury, and were referred to either VFS or FEES assessment. Patients with tracheostomy were not considered eligible for enrolment in this study as the tracheostomy changes the breathing and swallowing sounds in a way that it requires a separate group of individuals with enough number of subjects with the same condition.

The swallowing sounds were recorded simultaneously with the VFS or FEES assessment at DeerLodge Health Centre, and Riverview Health Centre, and Health Sciences Centre, Winnipeg, Canada. The study was approved by the Health Research Ethics Board of the University of Manitoba, and all participants or their legal guardian signed a written consent prior to the experiments.

Tracheal sounds were recorded by a Sony (ECM-77B) microphone placed over the suprasternal notch of the trachea. The sounds were recorded by a digital sound recorder (EDIROL R-44) at 44.1 kHz sampling rate. The patients were fed different types of a variety of solid and liquid food; the type and order of the food were decided by the speech-language pathologist (SLP) in charge, who performed either VFS or FEES assessments.

5.1.2 Signal Processing

Swallowing occurs within a breath cycle during the deglutition apnea (cessation of breathing). The microphone placed over the trachea collects both the

swallowing and breath sounds. In this study, only the breath sound signals after the swallowing events were analyzed. Therefore, the breath and swallowing sounds were separated manually by auditory and visual inspection of the signals in the time and frequency domain.

Each swallowing event of a patient was assessed by the SLP staff using the VFS/FEES images, and was marked as aspirated, or non-aspirated. Aspiration can occur during or after a swallowing; therefore, it is often difficult to find the exact timing of the aspiration. Thus, we analyzed 1-3 successive breath sounds after each swallow if the interval between the consecutive swallows permitted. Consequently, for any swallow marked as aspirated, for designing our classifier we marked the 1-3 proceeding breath sounds as aspirated. It should be noted that the swallows followed by coughing were excluded. Nevertheless, the cough is usually considered as a sign of aspiration. However, our goal is to detect the challenging silent aspirations.

Figure 5-1 shows the spectra (calculated using the Welch method with 50% overlap between the successive windows of 20 ms) of the breath sounds after all the swallowing events of a patient having aspiration. The spectra of the breath sounds after aspirated swallows are marked with red dashed color. It is notable that the main difference exists in low frequencies, where the spectra of the breath sounds after aspiration show higher magnitude.

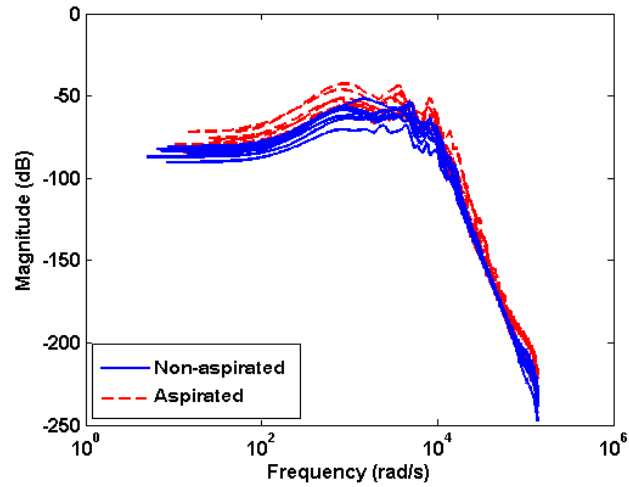


Fig. 5-1. The comparison of the PSD of breath sound signals after the swallow (simultaneously recorded with FEES) for the swallows (with/without aspiration) of a dysphagic individual.

5.1.3 Feature Extraction

We divided the frequency range below 500 Hz of each breath sound signal into 100Hz sub-bands (the first sub-band was between 60-100 Hz as the signals were previously high-pass filtered for above 60 Hz). Then, the average power of each subband was calculated from the PSD of each breath, resulting in 3 features for each breath. Figure 5-2 shows the scatter plot of these three features calculated for all breath sounds after the swallows of one patient having aspiration. The red squares belong to the breaths after aspirated swallows.

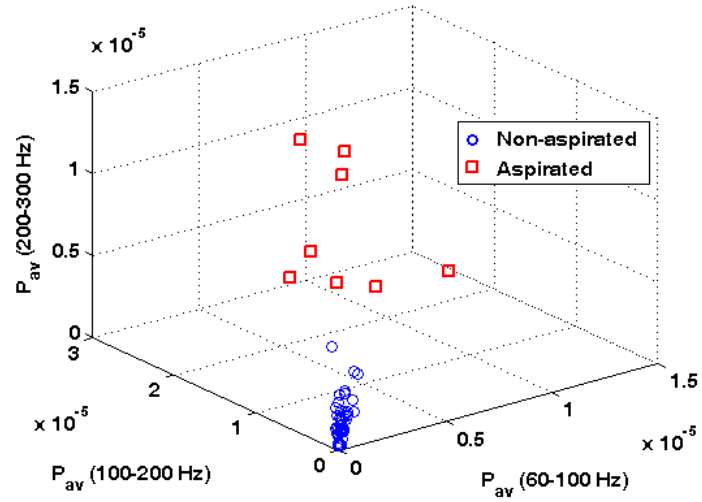


Fig. 5-2 Scatter plot of the three average power features of the breath sounds after swallows of a patient having aspiration

5.1.4 Unsupervised Classification

We used the fuzzy k-means clustering algorithm to find two clusters among the data points related to all the breath sounds of each patient. The fuzzy k-means clustering algorithm is commonly used for unsupervised classification [42]. In this method, each data point belongs to a group of clusters with a membership degree that is optimized through running the algorithm. The membership value is the probability function $\hat{p}(\omega_i | \mathbf{X}_j, \hat{\theta})$ whose parameter θ has to be updated by minimizing the cost function as (5-1):

$$J = \sum_{i=1}^k \sum_{j=1}^n [\hat{p}(\omega_i | \mathbf{X}_j, \hat{\theta})]^m \|\mathbf{X}_j - \mathbf{V}_i\|^2, \quad (5-1)$$

where \mathbf{X}_j is the data point or the feature vector to be classified, k is the number of clusters, \mathbf{V}_i represents the vector of cluster centers and m is the fuzziness index ($m > 1$) [49]. The role of m is to ignore the data points whose probability is low. The higher the value of m is, the lesser contribution of those data points to

the cost function [50]. In this study, we selected $\mathbf{X}_j = [P_{av}(60-100) P_{av}(100-200) P_{av}(300-400)]$, $k=2$ and $m=2$.

The algorithm begins by initializing the probability value of each data point indicating the degree it belongs to each cluster. The values should meet the condition expressed in (5-2):

$$\sum_{i=1}^k p(\omega_i | \mathbf{X}_j) = 1. \quad (5-2)$$

Then, the center of each cluster is calculated according to (5-3). Also, the probability of each cluster for every point is updated as (5-4).

$$\mathbf{V}_i = \frac{\sum_{j=1}^n p(\omega_i | \mathbf{X}_j)^m \mathbf{X}_j}{\sum_{j=1}^n p(\omega_i | \mathbf{X}_j)^m}, \quad (5-3)$$

$$p(\omega_i | \mathbf{X}_j) = \frac{\left(1/\|\mathbf{X}_j - \mathbf{V}_i\|^2\right)^{1/(m-1)}}{\sum_{l=1}^k \left(1/\|\mathbf{X}_j - \mathbf{V}_l\|^2\right)^{1/(m-1)}}. \quad (5-4)$$

Lastly, the procedures expressed (5-3)-(5-4) are repeated until the cost function reaches its local minimum. Consequently, the final values of the centers of clusters are obtained, and every data point is assigned to a cluster where its membership degree has the highest probability.

Having identified the two clusters of the breath sounds for each patient in the aspirated group, we designed a heuristic screening scheme to classify those breaths that were followed after an aspirated swallow as the following: If any of the breaths (up to 3) after a swallow belongs to the aspirated cluster, then that swallow is labeled as an aspirated swallow. The clustering and heuristic classification scheme is shown schematically in Fig. 5-3.

To calculate the accuracy, specificity and sensitivity, true positives were defined as the swallows marked as aspiration, and true negatives are the non-aspirated swallows.

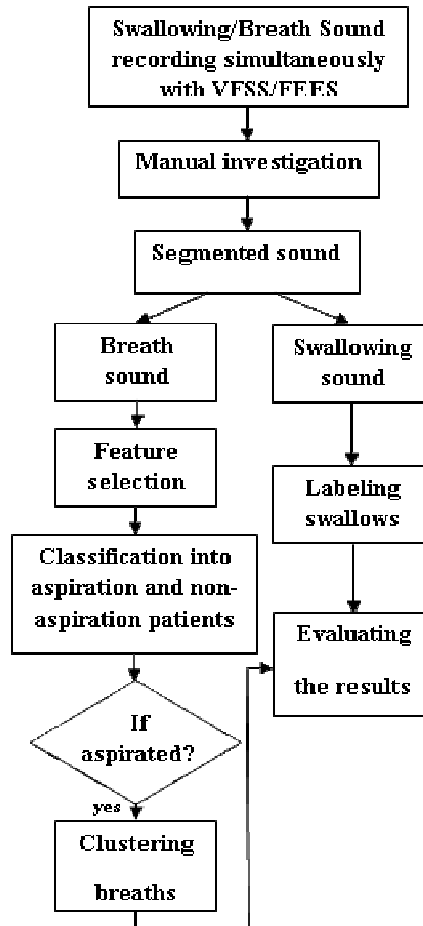


Fig. 5-3. Flowchart for aspiration detection.

5.2 Results

The spectral analysis showed in the breath sounds after the swallows formed two distinct clusters (Fig. 5-2); this pattern was consistent for all patients. Using the FEES/VFS assessment, it became apparent that the majority of the swallows of Cluster 1 belonged to non-aspirated swallows, while Cluster 2 had the majority of

the aspirated swallows. Clearly, the breath sounds after an aspirated swallow have significantly higher average power on the selected subbands.

The results of the unsupervised clustering along with the heuristic classification of the swallows of each patient are summarized in Table 5-1. For 8 out of 22 patients the sensitivity of the method was found to be 100% as all the aspirations were detected correctly. However, it should be noted that due to the very few number of aspirated swallows in some patients, i.e. Subjects, 2, 5, 8 and 9, the average accuracy may not a good representative of the method's accuracy.

Therefore, we have shown results in terms of the total number of swallows, the number of aspiration, the number of correctly classified, and the misclassified cases in Table 5-1. It should be noted that no result was reported for subject #16 in the table. The patient was on tube feeding, which is a device used to deliver nutrition directly into the stomach. The FEES was performed to evaluate his swallowing mechanism and decide if the tube can be removed. However, he could not complete even one swallow. There was a considerable accumulation of secretion in the pyriform sinuses that resulted in aspiration.

The overall accuracy, sensitivity, and specificity for the all the swallows of the patients are calculated as 80.0%, 79.9% and 81.3%, respectively.

Chapter 5- Detection of Swallows with Silent Aspiration using the Swallowing and Breath Sounds Analysis

Table 5-1. Classification results of swallows for 22 patients having aspiration.

subjects	Total no. of swallows	No. of swallows with aspiration	TP	FP	TN	FN	Gender /Age
1	10	3	3	2	5	0	M/70
2	20	2	1	0	18	1	M/60
3	11	4	4	1	6	0	F/23
4	20	6	4	1	13	2	F/23
5	20	4	3	2	14	1	M/71
6	9	7	6	0	2	1	M/63
7	14	3	3	4	7	0	M/60
8	14	2	0	2	10	2	F/70
9	21	2	1	2	17	1	F/57
10	16	4	4	3	9	0	M/62
11	25	2	2	2	21	0	M/52
12	22	9	9	2	11	0	M/70
13	21	5	4	2	14	1	M/81
14	19	10	8	2	7	2	M/56
15	9	4	4	3	2	0	M/90
16	0	0	0	0	0	0	M/40
17	15	6	4	2	7	2	M/73

18	21	5	4	4	12	1	M/85
19	19	7	5	4	8	2	M/91
20	29	8	7	9	12	1	M/20
21	27	10	8	5	12	2	M/20
22	13	4	3	2	7	1	F/57
Total	375	106	86	54	214	20	58.8±21.7

5.3 Discussion

Swallowing assessments are generally performed to ensure the early diagnosis of patients who are at high-risk of aspiration. As mentioned earlier, VFS and FEES are two instrumental methods used as the Gold Standard to detect aspiration. Both techniques have limitations such as being invasive (exposing the patient to radiation), the high cost of equipment and the need for specialized training. In addition they cannot be applied during normal eating setting. Therefore, a non-invasive, simple and yet accurate method with minimal interference on the normal eating procedure is of great interest. The results of this study show that the acoustical analysis of swallowing and breath sounds after the swallow has great potential to be used as a daily clinical tool to detect aspiration with high reliability and accuracy.

Given that the trachea can be modeled as a tube through which the air (expiration/inspiration) flows [51], then it is logical to study the fluid mechanics characteristics of the airway dynamics. It has been shown that the airflow in the trachea is highly likely to be turbulent [47]. Thus, any particle in the airway will increase the size of the turbulence large scales, which results in a higher power at

low frequencies [52]; thus, a lower pitch sound is expected to be heard after an aspirated swallow.

The particle in the trachea, as the result of aspiration, most probably changes the pattern of flow by creating vortices. Approximating the particle as a circular cylinder, we can obtain the frequency of vortex shedding in trachea, f , by the von Karman formula [53]:

$$\frac{fd}{U} = 0.198 \left(1 - \frac{19.7}{\text{Re}}\right),$$

where $\frac{fd}{U}$ is known as the Strouhal number, in which U is the instantaneous speed of airflow, d is the diameter of the particle, and Re is the Reynolds number. For a small particle with an average diameter of 1 mm, given that typical inspiratory speed is between 0.68 and 2.70 m/s [48] and Re varies between 800 in light breathing and 9300 in heavy breathing for the airflow in the trachea, the frequency of vortex shedding is calculated to be in the range of 136-540 Hz. Thus, the vortex frequency due to the particle in trachea can increase the energy of the sounds' power in frequencies lower than 540 Hz, depending on the size of the aspirated particle. This is congruent with the frequency range that the features are selected in this study.

It should be noted that in this study only the breath signals that were free of artifacts and vocal noise were analyzed. When analyzing the breath sounds after the swallow, there might be some breath sounds that mimic the breath sounds after aspiration. These breath sounds may mistakenly be interpreted as those following aspiration, while they might be related to a compensatory clearing mechanism [54]. This may increase the false positive error.

We aimed to detect the swallows during/after which the patients developed aspiration. The method was successful in detecting silent aspirations with 81% sensitivity. However, the sensitivity of the method for the patients with few number of aspirations was not as good as the sensitivity obtained for the patients with more aspirations. If we exclude the data of the patients with less than 3 aspirated swallows from Table 5-1 and calculated the sensitivity, it would increase; we can achieve 84% sensitivity. However, as we increase the sensitivity by ignoring the results of the subjects with few aspirations, we decrease the specificity results. The false negative results mainly exist due to the penetration and the residues that are left in the pharynx area after the swallow; they can increase the risk of aspiration.

Despite the fact that physicians have used cervical auscultation to detect aspiration, there has not been any study that analyzed the tracheal sounds to detect aspiration. Some studies used the accelerometer to detect tracheal vibration during swallowing act; they have reported 80.3% accuracy [55] for 94 aspirated and 100 normal swallowing samples. The same research group, using the same recording technique of tracheal vibration, classified the healthy and abnormal swallows (which also included the swallows with aspiration) by Euclidean linear discriminant classifier with an accuracy of 74.7% on a population of 24 patients [48].

The results of our proposed method for detecting aspiration are superior to the above studies (using tracheal vibration) and also those of the cervical auscultation [56], [57]. Our methods achieved 81% for the sensitivity and 80% specificity for silent aspiration detection compared to sensitivity and specificity values of 62% and 66% [57] and 94% and 70% reported by cervical auscultation [56]. Although there was an overestimation of aspiration (high false positive

error) in the studies mentioned above, cervical auscultation was suggested as a tool to identify patients with high risk of aspiration [57].

It should be noted that recording noise-free and good quality tracheal sounds for the purpose of detecting silent aspiration can be challenging in case of patients with loose skin and/or tracheostomy. In such cases, as an alternative solution, we suggest to record the sounds by a miniature microphone in the external ear or inside the nose, which was studied in [14] and explained in 0.

5.4 Summary

In this chapter, the application of acoustical analysis for detection of swallowing silent aspiration is investigated. We analyzed the breath sounds of dysphagic individuals, who demonstrated silent aspiration during their VFS/FEES feeding and swallowing assessments. We found that the low frequency components of the power spectrum of the breath sounds after a swallow show higher magnitude when there is aspiration. Thus, we divided the frequency range below 300 Hz into three sub-bands and calculated the average power of the breath sound signal in each sub-band as the characteristic features. We used the unsupervised fuzzy k-means clustering algorithm to label the breath sounds immediately after a swallow as aspiration or non-aspiration. The results were compared with those of the FEES/VFSS assessments provided by the speech language pathologists. The results are encouraging: more than 80% sensitivity in detection of swallows with silent aspiration.

Chapter 6- Conclusions and Recommendations for Future Work

6.1 Conclusions

This thesis describes the application of acoustical analysis in the detection of swallowing disorders. The first objective was to develop a mathematical model that represents the swallowing mechanism. The model was shown to be able to account for the differences between the pharyngeal phase of the swallowing act in the dysphagic and control individuals. The outcome of this part of study can be used to identify the individuals with swallowing disorders.

However, besides the research interest in understanding the complex series of neuromuscular events involved in swallowing, its application is limited to the research since identifying dysphagic individuals can be done by examining the clinical symptoms. On the other hand, identifying the patients, who are at risk of silent aspiration, is of great interest for the SLPs. Therefore, this research is mostly dedicated to aspiration detection.

The main goal of the swallowing assessment techniques is to determine if silent aspiration occurs. This information will help the SLP create an individualized treatment plan and a dietary modification to help patients improve their swallowing skills and reduce developing further complications such as pneumonia.

We hypothesized that the entry of food or liquid into the airway due to aspiration can change the time-frequency characteristics of the breath sound related to aspiration. Then, to detect aspiration, we analyzed 1-3 breath sounds after the swallows. First, the research proceeded by proposing a novel method based on phase-space analysis of breath sounds immediately after swallows followed by support vector machine classifier (SVM) as a diagnostic aid for identifying patients with high risk of aspiration. The classification result of the proposed method was compared with the FEES/VFS assessments provided by the SLPs; it showed 91% sensitivity and 85% specificity in detection of patients with severe aspiration (high risk dysphagia). The result was promising to suggest the proposed phase-space acoustical analysis method as a quick and non-invasive screening clinical tool to detect patients developing severe aspiration.

Also, we used a fuzzy k-means clustering to identify the aspirated swallows. The data set includes 22 patients who demonstrated aspiration during the assessments that was run simultaneously with VFS/FEES. The results showed an overall 80% accuracy in detecting the aspirated swallows. These results were encouraging on the use of the sound analysis of the breaths after the swallows for silent aspiration detection.

Overall, the contributions of this thesis can be summarized as the followings.

1. Development of a mathematical model for swallowing sounds generation to help understanding the main cause of dysphagia (Chapter 3)
2. Development of an algorithm for identifying patients at high-risk dysphagia. (Chapter 4)
3. Development of an algorithm for non-invasive swallowing aspiration detection (Chapter 5)
4. Investigating non-linear characteristics of swallowing sounds (Appendix A)

5. Investigating other sites for recording swallowing and breath sounds as an alternative to the trachea (Appendix B)
6. Investigating other time-frequency methods for analysis of swallowing sounds in comparison with the spectral and wavelet analyses (Appendix C)

6.2 Future Work Recommendation

The results of our research are encouraging; however, there are some improvements and works that can be considered in future studies as suggested below.

- Precise synchronization of the sound and imaging data: This can help improve the accuracy of the results. Also, the correlation between the swallowing/breath sounds and other swallowing disorders such as residue, can be studied. Thus, it will lead to better understanding of the swallowing mechanism.
- Automatic/semi-automatic segmentation of the breath sounds: This will facilitate using the algorithm in a medical device for swallowing assessment of dysphagic individuals by the SLPs in health centers.
- Using another microphone to record the ambient sounds, and using an adaptive filter to remove the undesired sounds such as the SLP's talking and preserve the swallowing and breath sounds to facilitate the automatic segmentation.
- Record in a controlled environment, where a specific protocol is defined for the type of food and liquid, and their order by which the patient is fed. Then, the effect of the type of food on the breath sound following aspiration can be studied.

- Study the depth of aspiration and quantify it by analyzing the swallowing/breath sounds.
- Extend the research by recording the pharyngeal pressure simultaneously with the VFS/FEES and swallowing/breath sounds: The correlation between the sound data and pressure profile of the swallowing can improve the modeling and diagnosis of dysphagia.

Bibliography

- [1] B. Jones, Normal and Abnormal Swallowing: Imaging in Diagnosis and Therapy, Springer, 2002.
- [2] J. A. Logeman, Evaluation and Treatment of Swallowing Disorders, Austin, Tx: Pro-Ed, 1998.
- [3] S. Langmore, Endoscopic Evaluation and Treatment of Swallowing Disorders [Hardcover], Thieme, 2000.
- [4] J. E. Dempsey, F. L. Vice and J. F. Bosma, "Combination of Cervical Auscultation and Videoradiography in Evaluation of Oral and Pharyngeal Dysphagia (abstract)," in *Symposium on Dysphagia*, Johns Hopkins Hospital, 1990.
- [5] L. Lazareck and Z. Moussavi, "Classification of Normal and Dysphagic Swallowing Sounds by Acoustical Means," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 12, pp. 2013-2112, 2004.
- [6] M. Aboofazeli and Z. Moussavi, "Analysis of Temporal Pattern of Swallowing Mechanism," 2006.
- [7] G. Rempel and Z. Moussavi, "The Effect of Viscosity on the Breath Swallow Pattern of Young People with Cerebral Palsy," *Dysphagia*, vol. 20, no. 2, pp. 108-112, 2005.
- [8] S. L. Hamlet, R. L. Patterson, S. M. Fleming and L. A. Jones, "Sounds of Swallowing Following Total Laryngectomy," *Dysphagia*, vol. 7, no. 3, pp. 160-165, 2006.
- [9] S. Sarraf-Shirazi and Z. Moussavi, "Acoustical Modeling of Swallowing

Bibliography

- Mechanism," *IEEE Trans. Biomed. Eng.*, vol. 8, no. 1, pp. 81-87, 2011.
- [10] S. Sarraf Shirazi and Z. Moussavi, "Investigating the Statistical Properties of the Swallowing Sounds," in *33rd IEEE EMBC Conf*, Boston, 2011.
- [11] S. Sarraf Shirazi, A. H. Birjandi and Z. Moussavi, "Non-invasive and automatic diagnosis of patients at high risk of swallowing aspiration," *Med. Biol. Eng and Comput.*, vol. 52, no. 5, pp. 459-465, 2014.
- [12] S. Sarraf Shirazi, C. Buchel, R. Daun, L. Lenton and Z. M. Moussavi, "Detection of swallows with silent aspiration using swallowing and breath sound analysis," *Med. Biol. Eng and Comput*, vol. 50, no. 12, pp. 1261-1268, 2012.
- [13] S. Sarraf Shirazi and Z. Moussavi, "Silent Aspiration Detection by Breath and Swallowing Sound Analysis," in *34th IEEE EMBC Conf*, San Diego, 2012.
- [14] S. Sarraf Shirazi, J. F. Baril and Z. Moussavi, "Characteristics of the Swallowing Sounds Recorded in the Ear, Nose and on Trachea.," *Med Biol Eng Comput.*, vol. 50, no. 8, pp. 885-890, 2012.
- [15] K. A. Kendall, R. J. Leonard and S. McKenzie, "Airway Protection: Evaluation with Videofluoroscopy," *Dysphagia*, vol. 19, pp. 65-70, 2004.
- [16] M. L. Splaingard, B. Hutchins, L. D. Sulton and G. Chaudhuri, "Aspiration in rehabilitation patients: videofluoroscopy vs bedside clinical assessment," *Arch Phys Med Rehabil.*, vol. 69, no. 8, pp. 637-640, 1988.
- [17] S. H. Lim, P. K. Lieu, S. Y. Phua, R. Seshadri, N. Venketasubramanian, S. H. Lee and P. W. Choo, "Accuracy of bedside clinical methods compared

Bibliography

- with fiberoptic endoscopic examination of swallowing (FEES) in determining the risk of aspiration in acute stroke patients," *Dysphagia*, vol. 16, no. 1, pp. 1-6, 2001.
- [18] L. E. Mackay, A. S. Morganz and B. A. Bernstein, "Swallowing Disorders in Severe Brain Injury: Risk Factors Affecting Return to Oral Intake," *Arch. Phys. Med. Rehabil.*, vol. 80, pp. 365-371, 1999.
- [19] B. R. Garon, M. Engle and C. Ormiston, "Silent Aspiration: Results of 1,000 Videofluoroscopic Swallow Evaluations," *J. of Neuro. Rehab.*, vol. 10, pp. 121-126, 1996.
- [20] M. A. Loughlin and G. M. Lefton-Greif, "Specialized Studies in Pediatric," *Dysphagia*, vol. 17, no. 4, 1996.
- [21] M. A. Cordaro and B. C. Sonies, "An Image Processing Scheme to Quantitatively Extract and Validate Hyoid Bone Motion Based on Real-Time Ultrasound Recordings of Swallowing," *IEEE Trans. on Biomedical Eng.*, vol. 40, no. 8, pp. 841-844, 1993.
- [22] A. L. Perlman, "Electromyography and the Study of Oropharyngeal Swallowing," *Dysphagia*, vol. 8, pp. 351-355, 1993.
- [23] J. B. Palmer, "Electromyography of the Muscles of Oropharyngeal Swallowing: Basic Concepts," *Dysphagia*, vol. 3, pp. 192-198, 1989.
- [24] R. B. Williams, A. Pal, J. G. Brasseur and I. J. Cook, "Space-time Pressure Structure of Pharyngo-eophageal Segment during Swallowing," *Am. J. Physiol. Gastrointest Liver Physiol*, vol. 281, p. 1290-1300, 2001.
- [25] D. Shaw, R. Williams, I. Cook, K. Wallace, M. Weltman, P. Collins, E.

Bibliography

- Mckay, R. Smart and M. Simula, "Oropharyngeal Scintigraphy: A Reliable Technique for the Quantitative Evaluation of Oral-Pharyngeal Swallowing," *Dysphagia*, vol. 19, p. 36-42, 2004.
- [26] K. H. Silver and D. Van Nostrand, "The Use of Scintigraphy in the Management of Patients with Pulmonary Aspiration," *Dysphagia*, vol. 9, pp. 107-115, 1994.
- [27] P. M. Zenner, D. S. Losinski and R. H. Mills, "Using Cervical Auscultation in the Clinical Dysphagia Examination in Long-Term Care," *Dysphagia*, vol. 10, pp. 27-31, 1995.
- [28] W. Selley, F. C. Flack and W. A. Brooks, "Respiratory Patterns Associated with Swallowing, part 2 Neurologically Impaired Dysphagic Patients," *Age Ageing*, vol. 17, pp. 173-176, 1989.
- [29] S. Langmore, "Coordination between Swallow and Respiration in Dysphagic Patients," 1991.
- [30] S. L. Hamlet, D. Penney and J. Formolo, "Stethoscope Acoustics and Cervical Auscultation of Swallowing," *Dysphagia*, vol. 9, pp. 63-68, 1994.
- [31] S. L. Hamlet, R. J. Nelson, R. L. Patterson and D. Michigan, "Interpreting the Sound of Swallowing: Fluid Flow through the Cricopharyngeus," *Ann Otol Rhinol Laryngol*, vol. 99, pp. 749-752, 1990.
- [32] M. Aboofazeli and Z. Moussavi, "Analysis and Modeling of Swallowing Sounds," Ph.D. dissertation, Dept. Elect. & Comp. Eng., Univ. of Manitoba, Winnipeg, MB, 2006.
- [33] L. J. Lazareck and Z. Moussavi, "Classification of Normal and Dysphagic

Bibliography

- Swallows by Acoustical Means,” M.Sc. dissertation, Dept. Elect. & Comp. Eng., Univ. of Manitoba, Winnipeg, MB, 2003.
- [34] A. Yadollahi and Z. Moussavi, ”Feature Selection for Swallowing Sounds Classification,” in *29th IEEE EMBS Conf*, 2007.
- [35] M. Aboofazeli and Z. Moussavi, ”Analysis and Classification of Swallowing Sounds Using Reconstructed Phase Space Features,” in *Proc. IEEE ICASSP*, 2005.
- [36] A. Yadollahi and Z. Moussavi, ”A Model for Normal Swallowing Sound Generation Based on Wavelet Analysis,” in *29th IEEE EMBS Conf*, 2008.
- [37] A. Jean, ”Brain Stem Control of Swallowing: Neuronal Network and Cellular Mechanisms,” *Physiological Reviews*, vol. 18, no. 2, 2001.
- [38] P. A. O’Neill, ”Swallowing and Prevention of Complications,” *British Medical Bulletin*, vol. 56, no. 2, pp. 457-465, 2000.
- [39] E. Shwedyk, R. Balasubramanian and R. N. Scott, ”A Nonstationary Model for the Electromyogram,” *IEEE Trans. on Biomed. Eng.*, vol. 24, no. 5, pp. 417- 424, 1977.
- [40] D. G. Proakis and J. G. Manolakis, *Digital Signal Processing*, New Delhi: Prentice Hall, 2007.
- [41] A. Mallat, *Wavelet Tour of Signal Processing*, Academic Press, 1999.
- [42] R. O. Duda, . P. E. Hart and D. G. Stork, *Pattern Classification*, Carrollton, TX: Wiley, 2000.
- [43] M. Aboofazeli and Z. Moussavi, ”Swallowing Sound Detection using Hidden Markov Modeling of Recurrence Plot Features,” *Journal of Chaos, Solitons*

Bibliography

- and Fractals*, vol. 39, no. 2, pp. 778-783, 2009.
- [44] D. Goring and V. Nikora, "Despiking Acoustic Doppler Velocimeter Data," *Journal of Hydraulic Engineering*, vol. 128, pp. 117-126, 2002.
- [45] L. Cea, J. Puertas and L. Pena, "Velocity Measurements on Highly Turbulent Free Surface Flow Using ADV," *Exp. Fluids*, vol. 42, p. 333-348, 2007.
- [46] I. Ayappa and D. M. Rapoport, "The Upper Airway in Sleep: Physiology of the Pharynx," *Sleep Medicine Reviews*, vol. 7, no. 1, pp. 9-33, 2003.
- [47] X. Y. Luo, J. S. Hinton, T. T. Liew and K. K. Tan, "LES Modeling of Flow in a Simple Airway Model," *J Med Eng Phys*, vol. 26, p. 403-413, 2004.
- [48] J. Lee, C. M. Steele and T. Chaua, "Classification of Healthy and Abnormal Swallows Based on Accelerometry and Nasal Airflow Signals," *Artificial Intelligence in Medicine*, vol. 52, pp. 17-25, 2011.
- [49] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum Press, 1981.
- [50] M. P. Windham, "Cluster Validity for the Fuzzy C-means Clustering Algorithm," *IEEE Trans Pattern Anal Mach Intell, PAMI-4* 4, pp. 357-363, 1982.
- [51] I. Ayappa and D. M. Rapoport, "The Upper Airway in Sleep: Physiology of the Pharynx," *Sleep Medicine Reviews*, vol. 7, no. 1, pp. 9-33, 2003.
- [52] S. P. Pope, *Turbulent Flows*, Cambridge: Cambridge University Press, 2000.
- [53] R. D. Blevins, *Flow-Induced Vibration*, 2nd edition, Van Nostrand Reinhold,

- 1990.
- [54] A. E. Stroud, B. W. Lawrie and C. M. Wiles, "Inter- and Intra-rater Reliability of Cervical Auscultation to Detect Aspiration in Patients with Dysphagia," *Clinical Rehab*, vol. 16, pp. 640-645, 2002.
- [55] J. Lee, S. Blain, M. Casas, D. Kenny, G. Berall and T. Chau, "A Radial Basis Classifier for the Automatic Detection of Aspiration in Children with Dysphagia," *J. Neuroengineering Rehabil*, vol. 3, no. 14, pp. 1-17, 2006.
- [56] C. Borr and M. Hielscher-Fastabend, "Reliability and Validity of Cervical Auscultation," *Dysphagia*, vol. 22, p. 225–234, 2007.
- [57] P. Leslie, M. J. Drinnan, P. Finn, G. A. Ford and J. A. Wilson, "Reliability and Validity of Cervical Auscultation: a Controlled Comparison using Videofluoroscopy," *Dysphagia*, vol. 19, no. 4, pp. 231-240, 2004.
- [58] J. M. L. Caillec and R. Garello, "Asymptotic Bias and Variance of Conventional Bispectrum Estimates for 2-D Signals," *Multidimensional Systems and Signal Processing*, vol. 16, p. 49–84, 2005.
- [59] M. J. Hinich, "Testing for Gaussianity and Linearity of a Stationary Time Series," *Journal of Time Series Analysis*, vol. 2, no. 2, pp. 169-176, 1982.
- [60] D. R. Brillinger and P. R. Krishnaiah, *Handbook of Statistics 3: Time Series in the Frequency Domain*, North-Holland, 1983.
- [61] J. K. Patel and C. B. Read, *Handbook of the Normal Distribution*, New York and Basel: Marcel Dekker Inc., 1982.
- [62] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate*

Bibliography

- Distribution, vol. 2, New York: Wiley, 1970.
- [63] K. Takahashi, M. E. Groher and K. Michi , "Symmetry and reproducibility of swallowing sounds," *Dysphagia* , vol. 9, pp. 168-173, 2004.
- [64] K. Takahashi, M. E. Groher and K. Michi , "Methodology for detecting swallowing sounds," *Dysphagia* , vol. 9, no. 1, pp. 54-62, 1994.
- [65] H. Firmin, S. Reilly and A. Fourcin, "Non-invasive Monitoring of Reflexive Swallowing," *Speech, Hearing and Language*, vol. 10, pp. 171-184, 1997.
- [66] G. A. Pressler, J. P. Mansfield, H. Pasterkamp and G. R. Wodicka, "Detection of Respiratory Sounds within the War Canal," *Proc. of the Second Joint EMBS/BMES Conf.*, vol. 2, pp. 1529-1530, 2002.
- [67] G. A. Pressler, J. P. Mansfield, H. Pasterkamp and G. R. Wodicka, "Detection of Respiratory Sounds at the External Ear," *IEEE Trans on Biomed Eng*, vol. 51, no. 12, pp. 2089-96, 2004.
- [68] A. E. Deddens, E. P. Wilson, T. H. J. Lesser and J. M. Fredrickson, "Totally Implantable Hearing Aids: The Effects of Skin Thickness on Microphone Function," *American Journal of Otolaryngology*, vol. 11, no. 1, pp. 1-4, 1990.
- [69] L. Cohen, "Time-Frequency Distributions-A Review," *Proceedings of the IEEE*, vol. 77, no. 7, p. 941-981, 1989.
- [70] P. J. Loughlin, J. W. Pitton and L. E. Atlas, "Bilinear Time-Frequency Representations: New Insights and Properties," *IEEE Trans. Signal Process*, vol. 41, no. 2, pp. 750-767, 1993.
- [71] B. Zhang and S. Shunsuke, "A Time-Frequency Distribution of Cohen's Class with a Compound Kernel and its Application to Speech Signal Processing,"

Bibliography

- IEEE Trans. Signal Process*, vol. 42, no. 1, pp. 54-64, 1994.
- [72] H. I. Choi and W. J. Williams, "Improved Time-frequency Representation of Multicomponent Signals Using Exponential Kernel," *IEEE Trans. Acoust., Speech, Signal Process*, vol. 37, pp. 862-871, 1989.
- [73] D. L. Jones and T. W. Parks, "A Resolution Comparison of Several Time-Frequency Representations," *IEEE Trans. Signal Process.*, vol. 40, no. 2, pp. 413-420, 1992.
- [74] J. Gramtorp and E. Frederiksen, "Frequency Response for Measurement Microphones – a Question of Confidence," Technical Review, Properties and Calibration of Laboratory Standard Microphones, Brüel & Kjær, Nov. 2001.
- [75] B. Jones, *Normal and Abnormal Swallowing: Imaging in Diagnosis and Therapy*, Springer, 2002.
- [76] R. Z. German and J. B. Palmer, "Anatomy and Development of Oral Cavity and Pharynx," 2006. [Online]. Available: <http://www.nature.com/gimo/contents/pt1/full/gimo5.html>. [Accessed 17 11 2013].

Appendix A Investigating the Statistical Properties of the Swallowing Sound

Much of what is known about physiological systems such as swallowing mechanism has been learned using linear and time-invariant (LTI) system theory. The main advantage of linear system analysis is the availability of analytical tools to deal with the modeling. However, many real systems have complex properties that cannot be studied by restricting them to linear techniques.

The input of the system is an important issue in modeling as it is neither known nor accessible for most physiological systems, such as swallowing. In such cases, it is convenient to make some assumption of the input signal, i.e. a random Gaussian noise signal. However, if the model is considered as an LTI, then it is not correct to assume a white Gaussian noise input while the output is not Gaussian. Hence, the statistical characteristics of the signal (the output of the system) should be studied before making any assumption of the input distribution and the type of the system. This section investigates the gaussianity and linearity of the swallowing sound signal, and whether it is different between the two groups of control and dysphagic individuals.

A.1 Method

A.1.1 Data

Data in this study included swallowing sound recordings of 10 dysphagic (stroke and/or head trauma patients) and 10 age-matched individuals without any swallowing disorder as the control subjects. The swallowing sounds were recorded by a Sony (ECM-77B) microphone placed over the suprasternal notch of the trachea and digitized at 44 kHz. The swallowing sounds of the dysphagic group were recorded simultaneously with the VFS or the FEES assessment at Health Sciences Centre, and Riverview Health Centre, Winnipeg, Canada. The experimental protocol was the same for all data recordings: Subjects were fed 5-8 boluses of a thin liquid texture (i.e. juice) with a 5ml spoon. The study was approved by the Biomedical Ethics Board of the University of Manitoba, and participants signed a written consent prior to experiments.

All signals were normalized to their maximum amplitude. IDS and BTS segments of the swallowing sound signals were separated manually by an expert by auditory and visual inspection of signals in the time and frequency domain. The IDS part of the swallowing sound was considered in this study.

A.1.2 Statistical Analysis

According to Wold decomposition theory, any weak or wide-sense stationary process $\{x(t)\}$ with innovations $\{\varepsilon(t)\}$ has a moving average (MA) representation as (5) [58]:

$$x(t) = \sum_{k=1}^{\infty} h(k)\varepsilon(t-k), \quad (\text{A-1})$$

where $\varepsilon(t)$ are independent identically distributed random variables with

$$E\{\varepsilon(t)\} = 0 \text{ and } h(0) = 1, \sum_{k=1}^{\infty} |h(k)|^2 < \infty.$$

In other words, a linear model can approximate $\{x(t)\}$ because the innovations are independent. In this case, any input with the Gaussian distribution will result in a Gaussian output. Thus, the 2nd order statistics, i.e. correlation and spectral analyses, can readily determine the statistical properties of the process and be used for system identification. However, if the innovations are not normal and $E\{\mathcal{E}^3(t)\} \neq 0$ or the signal comes from a nonlinear system, then higher order statistics become important. The 3rd order cumulant of a zero mean stationary process is defined as:

$$c_{xxx}(m, n) = E\{x^*(t)x(t+m)x(t+n)\}, \quad (\text{A-2})$$

which is nonzero since $E\{\mathcal{E}^3(t)\} \neq 0$. The cumulants of a Gaussian signal are zero at the orders higher than 3. Therefore, the Gaussianity of the process should be determined before making any assumption of the type of the system (i.e. linearity). It should be noted that the signal is assumed to be zero-mean in all the analyses.

Similar to the definition of the power spectrum as the Fourier transform of the second cumulant (the autocorrelation), the Bispectrum is defined as the two-dimensional Fourier transform of the 3rd order cumulant

(A-3). This is known as the indirect method for calculating the bispectrum:

$$B(\omega_1, \omega_2) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} c_{xxx}(m, n) e^{-j(\omega_1 m + \omega_2 n)} \quad (\text{A-3})$$

This function is periodic in both ω_1 and ω_2 with period 2π . The symmetry properties of the bispectrum of a real signal can be expressed as:

$$B(\omega_1, \omega_2) = B(\omega_2, \omega_1) = B(-\omega_1 - \omega_2, \omega_2) = B(\omega_1, -\omega_2 - \omega_2) \quad (\text{A-4})$$

Therefore, $B(\omega_1, \omega_2)$ can be determined in terms of the values inside a triangle whose vertices are located at $(0,0)$, $(1/2,0)$, $(1/3,1/3)$.

The indirect method is not a consistent estimate of the bispectrum, and needs some modification such as multiplying the time domain signal by a window function [58]. Also, this method is time consuming and has computational burden. This motivates the use of frequency domain analysis known as the direct estimation of the bispectrum assuming the linear representation shown in (A-5). The direct estimate of the bispectrum is calculated based on the Fourier transform of the signal ($X(\omega_j)$) is:

$$\hat{B}(j, k) = N^{-1} X(\omega_j) X(\omega_k) X^*(\omega_{j+k}), \quad (\text{A-5})$$

where N is the length of the signal, and $\omega_j = 2\pi j / N$ for $j = 0, 1, \dots, N$. Since $X(\omega_j)$ is periodic with N , the values of $\hat{B}(j, k)$ will be computed over the region inside a triangle formed by three lines as: $k = 0$, $j - k = 0$, and $2j + k = N$. It was shown that this estimation is not consistent [59]. Thus, the concept of averaging such as what is done in the power spectrum estimation methods is applied. The simplest approach is to average the values of $\hat{B}(j, k)$ over a 2D window of size $M \times M$. All the points must be located in the triangular region introduced above. It is shown that to have an asymptotically unbiased estimate, M should be an increasing function of N and satisfy the criteria in Eq. 10 [58].

$$\lim_{N \rightarrow +\infty} \frac{M(N)}{N} = 0, \text{ and } \lim_{N \rightarrow +\infty} \frac{M^2(N)}{N} = +\infty. \quad (\text{A-6})$$

Thus, $M(N)$ is proposed to be equal to N^c , where $0.5 < c < 1$. The area inside each window, over which the averaging is performed, consists of M^2 points centered at $\frac{(2j-1)M}{2}, (2k-1)M/2$, where $j = 0, \dots, k$ and $k \leq \frac{N}{2M} - \frac{N}{2}$. Therefore, the new estimator can be obtained as (A-7):

$$\hat{B}_{av}(j, k) = 1/M^2 \sum_{p=(j-1)M}^{jM-1} \sum_{q=(k-1)M}^{kM-1} \hat{B}(p, q). \quad (\text{A-7})$$

$\hat{B}_{av}(j,k)$ is an asymptotically unbiased, consistent estimate [60]. The approximate asymptotic distribution of the 2nd order spectra estimates was shown to be independent complex normal variables [61]. Therefore, a complex normal distribution with unit variance can represent the distribution of $Z(j,k) = \hat{B}_{av}(j,k) / \sqrt{\text{var}\{\hat{B}_{av}(j,k)\}}$.

The complex variance is calculated in [56] and shown in (A-8).

$$\begin{aligned} \text{var } \hat{B}_{av}(j,k) &= E\left|\hat{B}_{av}(j,k)\right|^2 - \left|\hat{B}_{av}(j,k)\right|^2 = \\ &NM^{-4}Q_{j,k}S_x\left(\frac{(2j-1)M}{2N}\right)S_x\left(\frac{(2k-1)M}{2N}\right)S_x\left(\frac{(2j-1)M}{2N} + \frac{(2k-1)M}{2N}\right) + O(M/N) \end{aligned} \quad (\text{A-8})$$

where and $Q_{j,k}$ is the number of p, q in the square that are in D but not on the boundaries, plus twice the number on the boundaries as explained before.

The two random variables $\text{Re}(\mathbf{Z})$ and $\text{Im}(\mathbf{Z})$ are independent and normally distributed $(N(\mu_i, 1))$. Thus, the variable $|\mathbf{Z}(j,k)|^2 = \text{Re}(\mathbf{Z})^2 + \text{Im}(\mathbf{Z})^2$ is distributed according to the chi-square distribution denoted by $\chi_2^2(\lambda)$ with two degrees of freedom [62] and the non-centrality parameter $\lambda(j,k) = \mu_1^2 + \mu_2^2$ [62] that can be written as [59]:

$$\begin{aligned} \lambda(j,k) &= |B(j,k)|^2 (NM^{-4}Q_{i,j})^{-1} S_{xx}^{-1}\left(\frac{(2j-1)f}{2}\right) \\ &S_{xx}^{-1}\left(\frac{(2k-1)f}{2}\right) S_{xx}^{-1}\left(\frac{(2j-1)f}{2} + \frac{(2k-1)f}{2}\right). \end{aligned} \quad (\text{A-9})$$

Therefore, the statistic of $Y(j,k) = \sum_{j,k} 2|\mathbf{Z}(j,k)|^2$ has a chi-square distribution with $2n$ degrees of freedom with the non-centrality parameter as the summation of $\lambda(j,k)$.

A.1.2.1 Test of Gaussianity

The test for gaussianity is designed based on the statistics introduced above. In

case of a Gaussian signal, $B(j,k)$ is equal to zero, which results in the central chi-square distribution as the non-centrality parameter becomes zero. To investigate the gaussianity of a signal statistically, we need two null and alternative hypotheses. The null hypothesis (H_0) is considered such that Y has approximately chi-square distribution $\chi_{2n}^2(0)$, and the alternative hypothesis (H_1) would be that Y has the noncentral chi-square distribution. In all instances a p value less than 0.05 is considered as the significance level. Rejecting the null hypothesis is equivalent to the rejection of the gaussianity assumption. If the signal is a nongaussian one, then the linearity test can be performed. Otherwise, the test of gaussianity does not convey any information about the linearity of the system.

A.1.2.2 Test of linearity

Given that a stationary signal has a linear representative, then the spectrum and bispectrum are obtained according to Eqs. (A-10), and (A-11) respectively:

$$S_{xx}(\omega) = \sigma_\varepsilon^2 |H(\omega)|^2, \quad (\text{A-10})$$

$$B(\omega_i, \omega_j) = \mu_3 H(\omega_i) H(\omega_j) H^*(\omega_i + \omega_j), \quad (\text{A-11})$$

where $E\{\varepsilon(t)\} = 0$, $\sigma_\varepsilon^2 = E\{\varepsilon^2(t)\}$, $\mu_3 = E\{\varepsilon^3(t)\}$ and $H(\omega)$ is the Fourier transform of the filter coefficients. If $\mu_\varepsilon \neq 0$, then the linear process is non-Gaussian.

The linearity is investigated by analyzing the characteristics of the noncentrality parameter of $\chi_2^2(\lambda)$ denoted by $\lambda(j,k)$ which equals to λ_0 (A-12) if the signal is generated by a linear process.

$$\lambda(j,k) = \lambda_0 = C\mu_3^2 / \sigma_\varepsilon^6, \quad (\text{A-12})$$

where $C = NM^{-4}Q$.

The statistic $2|\mathbf{Z}(j,k)|^2$ has a chi-square distribution ($\chi_2^2(\lambda)$). It is assumed that $\lambda(j,k)$ s are generated by a random variable that has a degenerate distribution equivalent to $\chi_2^2(\lambda_0)$ when $\lambda(j,k) = \lambda_0$. The statistical test performs based on the distribution of the random variable that generates λ . The null hypothesis assumes a linear process generates the signal, which results in a chi-square distribution with a constant noncentrality parameter $\lambda_0 = C\mu_3^2 / \sigma_\varepsilon^6$. The alternative hypothesis assumes that a nonlinear system generates the signal. Hence, the random variable generating λ has a non-degenerate chi-square distribution.

The noncentral chi-square distribution of $\chi_p^2(\lambda)$ can be expressed as the Poisson-weighted mixture of central chi-square distributions [62] as indicated in Eq. (16). Therefore, $\chi_p^2(\lambda)$ has a thicker tail under the null hypothesis than under the alternative hypothesis. This can be measured by calculating the interquartile range of the distribution. The result of statistical is obtained based on the comparison between the interquartile ranges of $\chi_p^2(\lambda)$ under each of the two hypotheses. The null hypothesis is true if the interquartile value of $\chi_p^2(\lambda)$ is greater than the interquartile value of $\chi_p^2(\lambda_0)$.

$$P(\chi_p^2(\lambda) < x) = \sum_{m=0}^{\infty} \frac{e^{-\lambda/2} (\lambda/2)^m}{m!} P(\chi_{p+2m}^2(0) < x). \quad (\text{A-13})$$

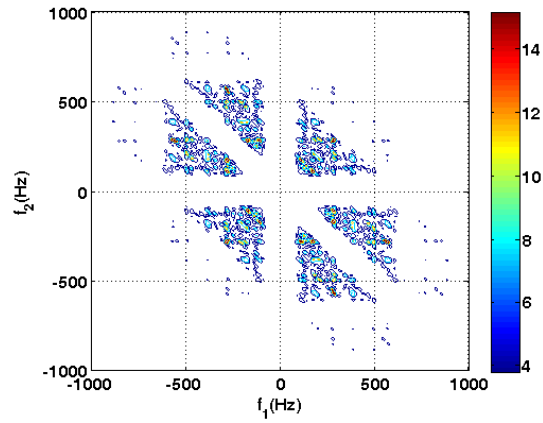
A.2 Results

The bispectra of typical swallowing sounds of a control and a dysphagic subject are depicted in Fig. A-1. The amplitude of the counter plot of these samples of the bispectrum confirms the nongaussian characteristics of the swallowing sound signal. The swallowing sound signals were tested for the nonlinearity.

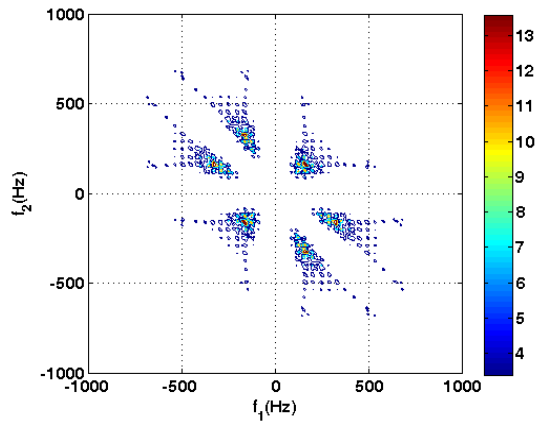
Interestingly, all the statistical significance p values, were zero. Thus, the null hypothesis of gaussianity was rejected.

Next, we performed the linearity test since the signals were categorized as nongaussians. The results show significant differences between the theoretically calculated value of the interquartile range and the estimated one for every swallow of each individual. The mean and standard error, averaged among the swallows in each group of data, are shown in Table A-1. Thus, it can be concluded that swallowing sound is generated by a nonlinear system, and should be analyzed by the nonlinear techniques. Moreover, no trend was found to be characterizing each group (Fig. A-2). Also, the results of both tests did not exhibit any obvious difference between the two groups of data.

It was shown that the bispectrum plots of both groups showed more and less the same pattern. However, further studies may find some characteristic differences between the bispectra of the two groups as this was beyond the scope of this part of the study. The outcome would be considered in future studies on modeling the swallowing mechanism.



(a)



(b)

Fig. A-1. The bispectrum of the swallowing sound of (a) a control and (b) a dysphagic subject.

Table A-1. The difference between the theoretical and estimated values of the interquartile range averaged for each group.

Data	$R_{\text{theory}} - R_{\text{estimated}}$
Control Group	730.9 ± 21.6
Dysphagic Group	1408.0 ± 66.2

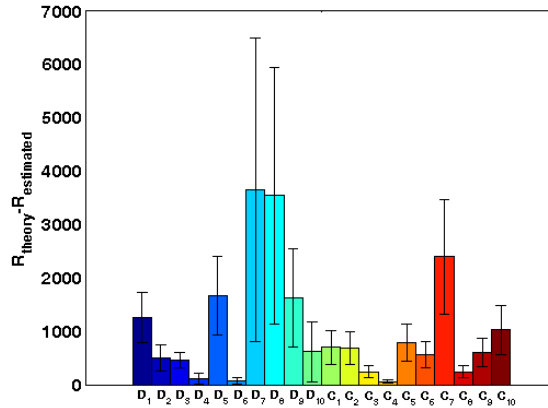


Fig. A-2. The test of linearity results: the difference between the estimated and the theoretical interquartile range for all the 10 control and 10 dysphagic data (averaged over all the swallows of each individual).

A.3 Discussion

The statistical analysis of the swallowing sound reveals that it has a nonlinear generation model. Recently, the nonlinear techniques were used to automatically detect the swallowing sound segments from the breath sounds, as well as classifying normal and dysphagic swallowing sounds. For example, nonlinear dynamic analysis, recurrence quantification analysis (RQA), (HMM), and multiresolution wavelet analysis were among the nonlinear techniques applied to detect characteristic features of swallowing sounds [32]. However, none of the few studies done on the swallowing sound modeling assumed nonlinear models for the swallowing sound generation.

In the model that was suggested in [36] and further studied in [9], swallowing sound was assumed to be produced by exciting the pharyngeal wall structure and tissue with an impulse train coming from the pharynx. The swallowing sound was thought to be the output of an LTI system representing the pharyngeal muscle and tissue responses to the neural activities that trigger the swallow, and are represented by an impulse train. That model benefits from the simple known relationship between the generated sounds during the swallow, and the

physiological events occurred as the result of the neural excitation. These were the first attempts toward modeling the swallowing mechanism.

Although the linear model can shed light on some aspect of the swallowing sound generation model, it may lead to an oversimplification of the actual system dynamics. This study investigated the validity of the linear assumption. The outcome of this study would be considered in future studies on modeling the swallowing mechanism.

Appendix B Characteristics of the Swallowing Sounds Recorded in the Ear, Nose and on Trachea

The various malfunctions and difficulties of the swallowing mechanism necessitate various diagnostic techniques to address those problems. Swallowing sounds recorded from the trachea have been suggested as a non-invasive method of swallowing assessment. However, acquiring signals from the trachea can be difficult for those with loose skin. The objective of this part of our study was to explore the viability of using the ear and nose as alternative recording locations for recording swallowing sounds. One major limitation with acoustical diagnostic techniques for dysphagia is the difficulty of recording a good quality sound signal if the patient has loose skin over the trachea. Therefore, in this study we explore the viability of using the ear and nose as alternative recording locations for recording breathing and swallowing sounds for the purpose of dysphagia and aspiration detection. Although the main motivation for this study comes from the application of the technique in older population, in this pilot study we tested the concept of ear and nose recordings in comparison to tracheal recording only in a few young people. We believe under normal conditions, the relationship between the sounds collected at ear, nose and trachea would not change significantly by age.

The swallowing sounds are commonly recorded over the trachea [63],[30]. Few

studies have been devoted to finding the ideal location of sensor placement on the neck area to acquire adequate signals for analysis [64]. However, it is not always possible to record swallowing acoustics from the neck area; thus, the discovery of alternative recording areas is vital for patient diagnosis. Fortunately, the neck is not the only anatomical area where swallowing sounds can be detected [65]. The ear and nose are two under-researched anatomical areas that show great promise for use in the signal-collection of swallowing sounds. Very appropriately these organs in combination with the structures concerned in the swallowing mechanism form an entire field of study in medicine, otorhinolaryngology (the study of the ear, nose and throat). The objective of this pilot study was to make an initial exploratory foray into investigating the possibility of using the ear or nose as regions from which to record swallowing sounds for the purpose of identify aspiration and potential dysphagia.

B.1 Method

B.1.1 Experimental Data

5 young healthy subjects (20.7 ± 2.3 y, 2 females) participated in this study and gave written consent. The study was approved by the Biomedical Ethics Board of the University of Manitoba. The subjects were prepared in the following manner after being seated in an acoustically isolated room:

A. The neck of each subject was restrained using an Ambuff Perfit ACE extrication collar to limit any potential noise contributions due to neck movement.

B. The first of three Sony ECM-77B electret microphones (40 Hz-20 KHz Bandwidth) was applied over the suprasternal notch using double-sided tape.

C. After piercing a hole through the center of a foam earplug (PharmaSystems Quiet Foam uHear Ear Plugs); the second microphone was

inserted into the earplug, similar to what was employed in previous studies aimed at detecting respiratory sounds from the ear [66], [67]. The earplug and microphone were then inserted into the subject's ear and adjusted until a satisfactory signal quality was achieved.

D. The third microphone was prepared and inserted into the subject's nostril such that the microphone remained securely in place during swallowing, and did not fall out. For each subject the nose microphone was prepared by enveloping it in both plastic wrap followed by a fresh 3.5×9.5 cm sheet of 2-ply nose tissue to isolate the microphone from mucus and nasal fluids. The plastic wrap was placed such that it did not occlude the microphone head and the nose tissue was placed such that that a bubble of air remained between the nose tissue and microphone head. The left nostril was used as a default for the test, however if the signal's quality was found to be lacking, the alternate nostril was attempted.

The signals were amplified and filtered (5 Hz-5 kHz) using Biopac DA100C, and digitized by NI-DAQ (NI cRIO-9215) at 10240 Hz sampling rate. After recording, the signals were filtered through a MATLAB Butterworth band pass filter (100-3000 Hz) to eliminate high frequency ambient noise, and low frequency interferences such as heart sounds and muscle artifacts.

Figure B-1 shows a diagram of the experimental setup described above. Each subject was handed a disposable drinking cup of water, and asked to use a plastic tablespoon to consume the water with spoon at their own pace but allow only one swallow within one breath cycle. Five to seven swallows were recorded. The bolus size of the water was limited to 15 ml (i.e. one full standard US tablespoon).

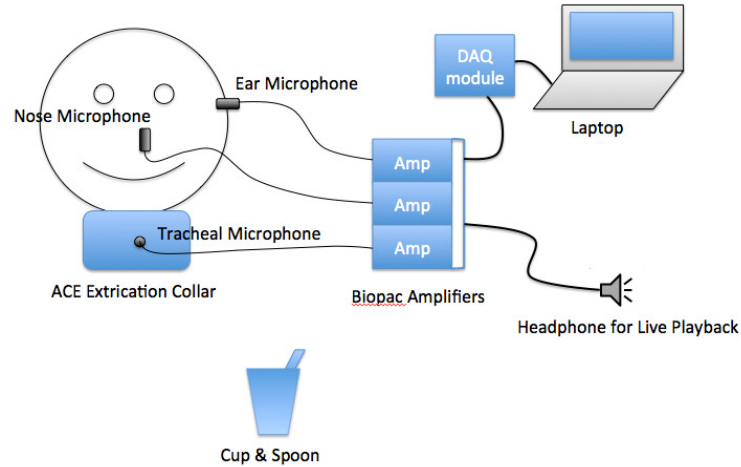


Fig. B-1 Setup for Swallowing Experiment

B.1.2 Signal Analysis

The swallowing segmentation into IDS and BTS was done by aural and visual examination using the tracheal recording as a reference similar to those in [9]. Figure B-2 shows the swallowing and the breath sound signals recorded at the three locations, trachea, nose and ear in time domain.

The recorded sounds were segmented into 3 sections: IDS, BTS and the post-swallow breath, which was an expiratory phase for all subjects in this study. Each signal segment was normalized to its variance (energy). Then, we calculated the Power Spectrum Density (PSD) of each of these sections using Welch's Method with 50% overlapped Hanning windows of 50 ms in length. Figure B-3 shows the PSD of the IDS segment of each recording. The three signals were normalized to their variance. The tracheal sound has the lowest magnitude due to the normalization. The tracheal graph and the ear PSDs would interchange, if the signals were not normalized.

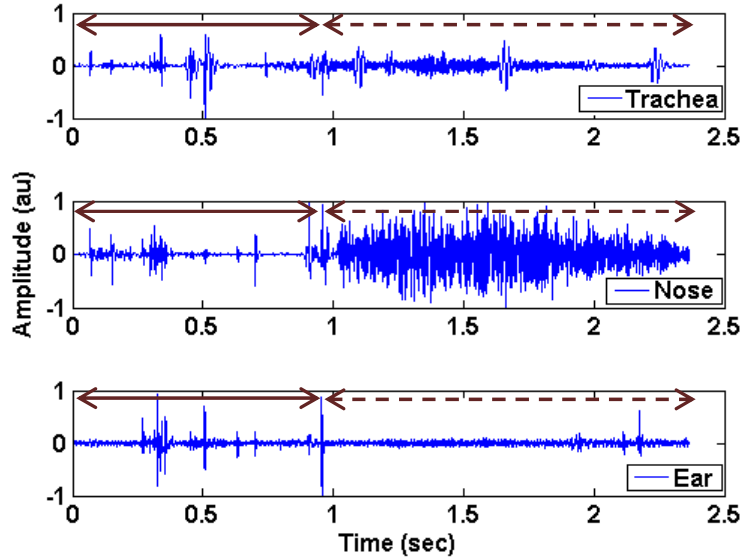


Fig. B-2 A typical normalized swallowing and breath sounds signal as marked by the solid arrow followed by breath sounds as indicated by the dashed arrow; the signals are shown in time domain, and recorded simultaneously from trachea, nose, and ear. Au= arbitrary unit.

We extracted the following features from the PSDs: a) the peak frequency (f_{peak}) as the frequency at which the peak magnitude occurs, b) the frequency at which the signal had lost 90% of its power, called f_{max} , c) the average power of the PSD over the octave bands: 150-300, 300-600, 600-1200, and 1200-2400 Hz as were used in a previous study seeking to detect respiratory sounds at the external ear [67].

Lastly, we calculated the approximation wavelet coefficients at the 2nd and 3rd levels of decomposition using Symlet basis function of order 8. The energy of those wavelet coefficients were shown to distinguish between the two groups of dysphagic and control data [9]. Therefore, we were interested to investigate the quality of the nose and ear signals in comparison to tracheal sound with respect to the same characteristic features.

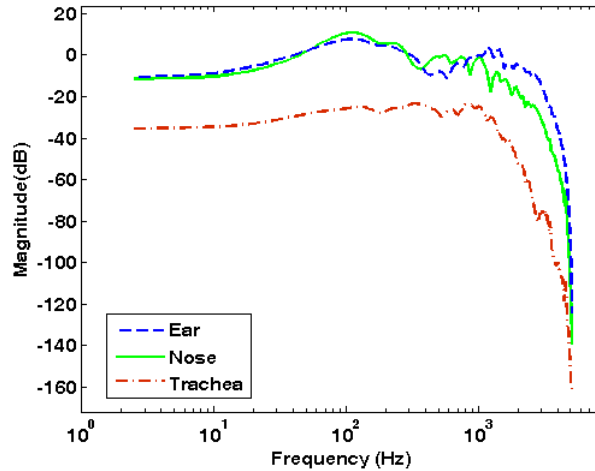


Fig. B-3 Typical spectra of the IDS segment of a swallowing sound recorded at the ear, nose and trachea of one subject. Each segment was normalized to its total energy before spectral estimation.

B.2 Results and discussion

B.2.1 Qualitative Observations of Swallow Signal Quality

Overall, the signals recorded at nose, ear and tracheal, all had a high level with respect to background noise. Compared to signals recorded over trachea, the signal-to-noise (SNR) of the nose breath sounds was higher and SNR of those in ear was lower. The quality of the swallowing sounds for both the ear and nose were very comparable to that recorded over the trachea. Though, these differences, while noticeable, were slight. An interesting observation from the time domain signals (Fig. B-2) was that the final discrete sound (FDS) could be clearly heard in the ear recording. FDS is a short duration click sound at the end of swallow and opening of the airway. It has been speculated to be due to the airway opening. However, based on our experience, FDS is not always present. Comparing all the three signals in the time domain, we found that although the signal recorded in the ear is not as strong as the tracheal or nasal ones, its FDS segment (if present) could be picked up by the ear microphone, which confirms our assumption about the origin of the FDS segment.

B.2.2 Analysis of the Peak (f_{peak}) and Maximum Frequency (f_{max})

Figure B-4 shows the values calculated as f_{peak} for the water swallows of all the subjects' three recorded signals, and for each section (IDS and BTS) of the swallowing signal. The results showed lack of consistency between subjects and between features as to an exact peak and maximum frequency. A larger more in depth study may reveal more about the effects of recording location on the peak and maximum frequencies; however, due to the limited sample size in this study, we refrain from drawing conclusions based on the apparent inconsistency of the data.

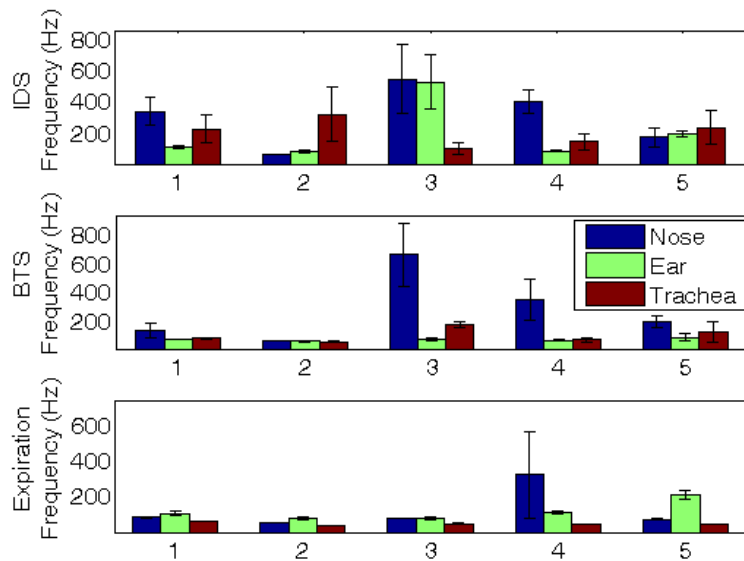


Fig. B-4. f_{peak} of the IDS, BTS and the expiration segments for all locations averaged among subject's data.

B.2.3 Analysis of Average PSD Magnitude over Octave Frequency Bands

Figure B-5 shows the average power calculated in the 4 octave frequency bands, averaged among the subjects, for different recording site. As can be seen, the signals of the three recording sites have a consistent pattern in terms of power over different frequency bands. The ear appears to have the lowest downward

sloping trend (3.8 dB/octave frequency step decay for the breath sound) whereas the trachea has the greatest (7.7 dB/octave frequency step decay for the breath sound). The average power values of the ear falling at a slower rate than the nose and the trachea average power at a faster rate than the nose.

The average power calculated for the tracheal recording falls off at higher frequencies at a greater degree than the nose and ear is consistent with the fact that skin acts as a low-pass filter, with a varying degree of strength dependent on the skin thickness for frequencies from about 500 – 8000 Hz [68]. As the ear and nose signals are not recorded through the skin, they do not suffer this effect.

It should also be noted that the ear signal appeared to have a higher noise floor than those recorded at nostril and trachea. This might have contributed to a lower signal drop off in the higher frequencies (as ambient random noise is constant over all frequencies). Our ear recording results agree with those published in [67]. In that study of breath sounds, recorded at the external ear, a gradual loss of 10 to 20 dB in signal strength between the 150-300 Hz and 1200-2400 octave bands was noted, which may indicate that the noise floor is of less concern than initially thought due to the noisy nature of the signals recorded in the ear.

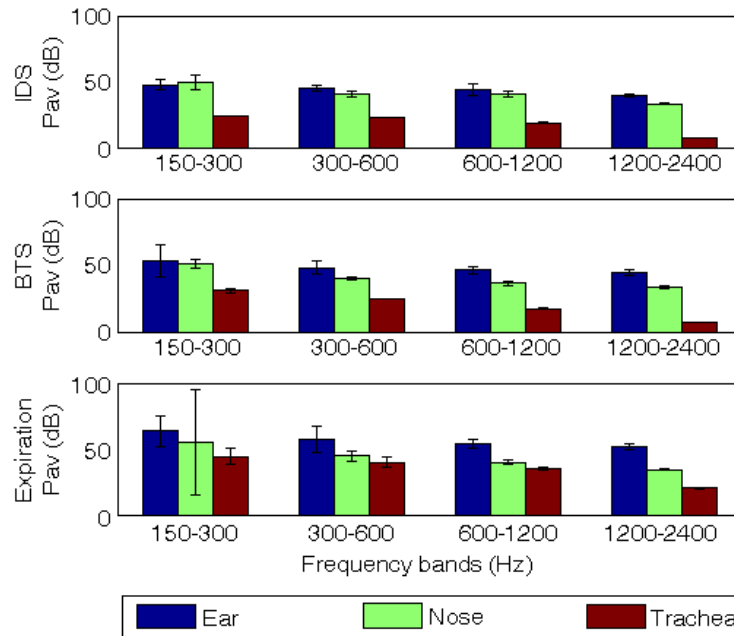


Fig. B-5. The average power of the IDS, BTS and the expiration segments calculated over the octave frequency bands. The values are averaged among subject's data.

As we are interested in the low frequency components of the breath sounds (below 300 Hz) for aspiration detection, it is important that the average power of the signals does not change in the low frequencies; the higher frequencies are of less importance for aspiration detection. Since the PSDs of the signals of the three recording sites remain similar to each other (with less than 20 dB variation) and consistent in the low frequencies, it may be concluded that the ear and nose may hold promise for use in detecting aspiration.

B.2.4 Analysis of Wavelet Coefficients

Figure B-6 presents the calculated 3rd order wavelet coefficients for the IDS of water swallows averaged for each subject. It can be seen that there is no consistent pattern between the recording locations for the swallows in either the 2nd or the 3rd order decomposition. The wavelet coefficients and thus the fundamental waveforms are incongruent between the ear, nose and trachea. Thus,

Appendix B- Characteristics of the Swallowing Sounds Recorded in the Ear, Nose and on Trachea

recordings from these locations should not be arbitrarily interchanged. This implies that during an acoustical swallowing assessment, if the goal is to diagnose dysphagia in general, all recordings must be taken from either the trachea, or the nose or the ear but not from a mixture of the recording sites.

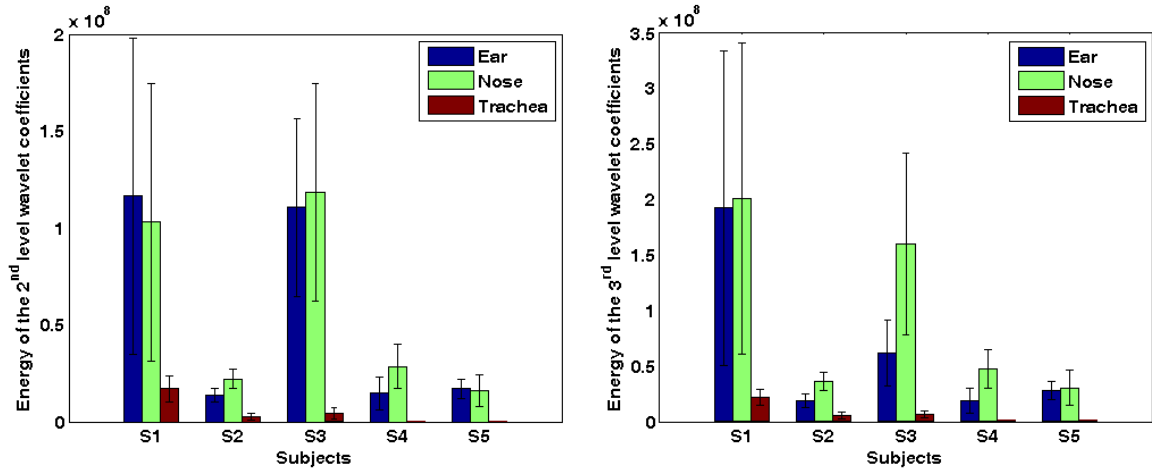


Fig. B-6. The mean and standard error of the energy of wavelet coefficients averaged for the IDS segments of all subjects a) 2nd order decomposition, b) 3rd order decomposition. The value of the standard error shows the variation within each subject.

B.2.5 Study Limitation

There are certain limitations to the results discussed as well as issues discovered during our study that we suggest be considered in subsequent experiments. We found that noise recorded in the ear was strongly dependent on the placement of the microphone in the subject's ear. This is likely due to the variances in ear canal shape and the limitations of using a cylindrical earplug and rigid microphone. It is also important to note that a more thorough study should also have considered normalizations involved with inter subject variance in the physical characteristics of the ear and nose to account for differences in ear canal shape, nose length, and various other factors. These factors were not considered in this pilot study. We also wish to stress that due to the small sample size used and the nature of this being a pilot study we chose to observe strong visual trends in the data as opposed to calculating precise numerical values whose accuracy and statistical significance could not be guaranteed.

B.3 Conclusion

In accordance with the objective of this study, we found that recording swallowing and breath sounds at the ear or nose may be used as alternative recording site as to trachea depending on the goal of acoustical swallowing assessment. If the goal is identifying people with dysphagia in general, the recording site cannot be used interchangeably between the subjects. On the other hand, if the goal is only to detect the swallows with aspiration within a dysphagic patient, the ear and nose sites may be used as an alternative recording site to trachea in case of a patient having loose skin over the neck. In summary, direct comparisons of swallowing sounds recorded at different sites is not recommended. However, recording swallowing sounds at the ear or nose in cases where tracheal

Appendix B- Characteristics of the Swallowing Sounds Recorded in the Ear, Nose
and on Trachea

recordings cannot be used is certainly viable for low frequency breath sound
analysis for use of aspiration detection for a dysphagic patient.

Appendix C Comparison Study of Other Time Frequency Representations

C.1 Introduction

Short-time Fourier transform, or spectrogram, is known as the standard method for the study of signals. The spectrogram uses the quasi-stationary assumption; the signal is assumed to be stationary within a specific time interval by multiplying with a window in time domain. Choosing a long time window leads to a good frequency resolution, and inevitably, poor time resolution. On the other hand, a short time window results in a good time resolution at the cost of poor frequency resolution. Also, according to the uncertainty principle, the size of the window cannot be selected arbitrarily small and, thus, it not possible to achieve a perfect time-frequency resolution. Therefore, there is always a tradeoff between the time and frequency resolutions when calculating the spectrogram.

C.1.1 Cohen's Class of Time-Frequency Representation

To improve the resolution issue of the spectrogram, other time-frequency methods have been introduced based on the generalized bilinear time-frequency representation (TFR). The TFR's can be defined as the Fourier transform of a weighted, symmetric, nonstationary correlation of the signal in the lag variable τ (C-1).

$$P(t, f) = \iint \phi(t-u) s^* \left(u - \frac{1}{2} \tau\right) s \left(u + \frac{1}{2} \tau\right) e^{-j2\pi f \tau} du d\tau, \quad (\text{C-1})$$

where $s(t)$ is the signal, and $\phi(t, \tau)$ is the kernel which is independent of the signal. The TFRs can also be defined in terms of the Fourier transform of the signal (C-2).

$$P(t, f) = \iint \Phi(\eta, f - f') S^* \left(f' - \frac{1}{2} \eta\right) S \left(f' + \frac{1}{2} \eta\right) df' d\eta, \quad (\text{C-2})$$

where $\Phi(\eta, f)$ is the 2-D Fourier transform of the kernel, and $S(f)$ is the Fourier transform of the signal $s(t)$.

The basic goal of these alternative methods is find a joint time-frequency distribution that describes the energy or the intensity of a signal simultaneously in time and frequency [69]. Thus, $P(t, f)$ shows the intensity at time t and frequency f ; or equivalently, $P(t, f) \Delta t \Delta f$ indicates the fractional energy of the signal in a time-frequency interval of $\Delta t \Delta f$ at t, f . The time and frequency marginals are defined as (C-3) and (C-4), respectively. The time marginal is called the instantaneous power and the frequency marginal is known as the energy density spectrum.

$$\int P(t, f) df = |s(t)|^2, \quad (\text{C-3})$$

$$\int P(t, f) dt = |S(\omega)|^2. \quad (\text{C-4})$$

The total energy (E), as shown in (C-5) will be equal to the total energy of the signal if the marginal are satisfied.

$$E = \iint P(t, f) dt df. \quad (\text{C-5})$$

C.1.2 Desired Properties Related to the Kernel

The properties of the TFRs are determined based on the kernel; they are reflected as the constraints imposed on the kernel [70]. Here, we explain some of the desired properties and their relation to the kernel constraints.

- Time shift invariance can be achieved if $\varphi(\eta, \tau)$, which is the 1-D Fourier transform of $\phi(t, \tau)$ in t , is independent of time.
- Frequency shift invariance can be guaranteed if $\varphi(\eta, \tau)$ is independent of frequency.
- The instantaneous power (the time marginal) is obtained from the frequency integration of $P(t, f)$ if $\varphi(\eta, 0) = 1, \forall \eta$.
- The energy spectral density (the frequency marginal) can be gained from the time integration of $P(t, f)$ if $\varphi(0, \tau) = 1, \forall \tau$.
- Bilinear distributions are real if they satisfy the necessary and sufficient condition of $\varphi(\eta, \tau) = \varphi^*(-\eta, -\tau)$.
- The TFR is limited to the same time interval of a time-limited signal (finite support in time) if $\varphi(t, \tau) \equiv \int \varphi(\eta, \tau) e^{j2\pi\eta t} d\eta = 0, |t| > 1/2|\tau|$.
- The TFR is limited to the same frequency range of a band-limited signal (finite support in frequency) if $\varphi(\eta, f) \equiv \int \varphi(\eta, \tau) e^{j2\pi f \tau} d\tau = 0, |f| > 1/2|\eta|$.

C.1.3 Interference Properties

The bilinear structure of the TFRs generates some inherent spurious values in the distribution, which is referred to as the cross terms. These terms are more prevalent in the multicomponent signals and cause difficulty in interpreting the distribution. For example, assume a stationary analytic signal such as (C-6):

$$s(t) = s_1(t) + s_2(t) = e^{j2\pi f_1 t} + e^{j2\pi f_2 t}. \quad (\text{C-6})$$

The TFR representation can be shown as (C-7):

$$P(t, f) = \Phi(0, f - f_1) + \Phi(0, f - f_2) + P_{s_1 s_2}(t, f) + P_{s_2 s_1}(t, f), \quad (\text{C-7})$$

where $P_{s_1 s_2}(t, f)$ can be defined as:

$$\begin{aligned} P_{s_1 s_2}(t, f) &= \iint \phi(t-u, \tau) e^{j2\pi f_1(u+\tau/2)} e^{-j2\pi f_2(u-\tau/2)} e^{-j2\pi f \tau} du d\tau = \\ &= \iint \phi(t-u, \tau) e^{-j2\pi(f_2-f_1)u} e^{-j2\pi[f-(f_1+f_2)/2]\tau} du d\tau = \\ &= \Phi(f_1 - f_2, f - (f_1 + f_2)/2) e^{j2\pi(f_1-f_2)t}. \end{aligned} \quad (\text{C-8})$$

Similarly, we have:

$$P_{s_2 s_1}(t, f) = \Phi(f_2 - f_1, f - (f_2 + f_1)/2) e^{j2\pi(f_2-f_1)t}. \quad (\text{C-9})$$

For the real TFRs, we can express the cross term component as (C-10) and define the cross term envelope according to (C-11) [70].

$$P_{cross}(t, f) = P_{s_1 s_2}(t, f) + P_{s_2 s_1}(t, f) = 2 \text{Re}[\Phi(f_2 - f_1, f - (f_2 + f_1)/2) e^{j2\pi(f_2-f_1)t}]. \quad (\text{C-10})$$

$$K(f; \Phi) \equiv \Phi(f_2 - f_1, f - (f_2 + f_1)/2). \quad (\text{C-11})$$

From (C-6)-(C-9), it can be seen that the cross terms appear as the 2-D Fourier transform of the kernel $\phi(t, \tau)$ calculated along the $\eta = f_2 - f_1$ axis and shifted in frequency by $f = (f_1 + f_2)/2$.

Various time frequency distributions have been proposed to compromise the cross term and the TFR's desired properties [71], [72]. For example, for the Wigner distribution with the kernel function is $\Phi(\eta, f) = \delta(f)$, the cross terms are not attenuated. They are located at the frequency average of the signal frequencies; the cross term envelope is shown to be $K(f; \Phi_{Wigner}) \equiv \delta(f - (f_1 + f_2)/2)$.

For the Born-Jordan TFR, the kernel is defined as $\Phi(\eta, f) = 1/|\eta| \text{rect}(f/\eta)$; hence, the cross term envelop can be obtained as (C-12):

$$K(f; \Phi_{\text{Born-Jordan}}) \equiv (2/(f_2 - f_1)) \text{rect}[(f - (f_1 + f_2)/2)/(f_2 - f_1)]. \quad (\text{C-12})$$

Assuming that $f_1 < f_2$, we can write (C-12) as (C-13).

$$K(f; \Phi_{\text{Born-Jordan}}) \equiv \begin{cases} (2/(f_2 - f_1)) & f_1 \leq f \leq f_2 \\ 0 & \text{otherwise} \end{cases}. \quad (\text{C-13})$$

Using (C-13), we can see that the cross terms are attenuated and spread over the signal frequencies. Thus, the Born-Jordan TFR is known as a member of the class of reduced interference distributions (RID), which has the property of reducing cross terms.

The spectrogram can be also considered as a member of the TFRs whose kernel is defined as $\Phi(\eta, f) = H(f + \eta/2)H^*(f - \eta/2)$, where $H(f)$ is the Fourier transform of the window multiplied by the signal. The cross term envelope can be obtained as

$$K(f; \Phi_{\text{Spectrogram}}) \equiv H(f - f_1)H^*(f - f_2). \quad (\text{C-14})$$

Interestingly, the cross terms are placed at the actual signal frequencies in the spectrogram. Although the cross terms attenuation is not an issue in the spectrogram, it was shown that it is not possible to satisfy the time marginal while attenuating the cross terms and localizing them to the actual frequency components [70]. Although the spectrogram does not satisfy the marginals, its non-negative distribution makes it the desirable TFR when it comes to interpreting the energy distribution in the time-frequency plane. Table C-1 shows a summary of the properties of the three TFRs we have discussed.

Table C-1. The properties of some TFR kernels

TFR	Kernel $\phi(t, \tau)$	Marginal		Finite support		Cross term	
		Time	Freq.	Time	Freq.	Attenuation	Localization
Spectrogram	$h(t + \tau/2)h^*(t - \tau/2)$					✓	✓
Wigner	$\delta(t)$	✓	✓	✓	✓		
Born-Jordan	$1/ \tau \text{rec}(t/\tau)$	✓	✓	✓	✓	✓	

To further explain the TFRs mentioned above, we will show the implementation results of the Wigner, Born-Jordan and the spectrogram on a multicomponent test signal, which consists of 2 finite duration frequencies components at 12Hz and 20 Hz. The test signal can be expressed as (C-15), which is shown in Fig. C-1(a).

$$x(t) = \begin{cases} \sin(2\pi f_1 t) & 0 < t < 1 \\ \sin(2\pi f_1 t) + .8\sin(2\pi f_2 t) & 1 \leq t \leq 2.5 \\ 0 & 2.5 < t < 3 \\ \sin(2\pi f_1 t) + .8\sin(2\pi f_2 t) & 3 \leq t \leq 4, \end{cases} \quad (\text{C-15})$$

where $f_1 = 12\text{Hz}$ and $f_2 = 20\text{Hz}$. As can be seen in Fig. C-1(b), the spectrogram yields the most satisfying results in terms of the interpretation. This is mainly due to spectrogram's cross term properties in localizing the cross terms around the signal frequencies. It should be noted that the spectrogram does not preserve the marginals nor does it maintain the finite support properties.

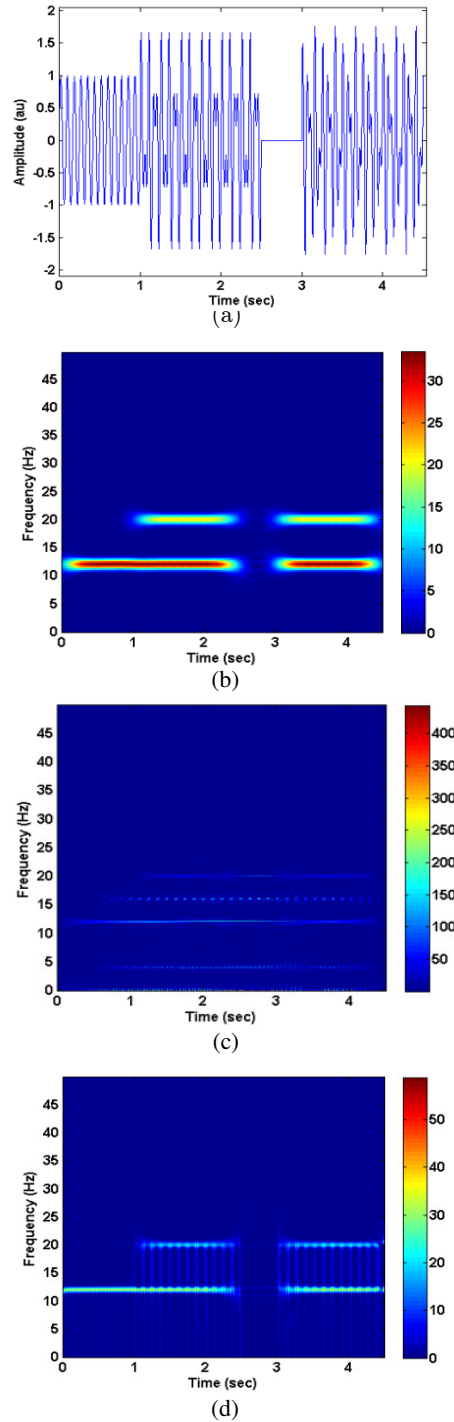


Fig. C-1 (a) The test signal consisting of two frequencies, 12 Hz and 20 Hz in time domain, (b) spectrogram of the signal, (c) the Wigner representation, and (d) Born-Jordan time-frequency representation of the signal.

On the other hand, both the Wigner and Born-Jordan kernels produce cross terms. These cross terms are more significant in Wigner distribution and are

mainly visible at the frequency 4 Hz and 16 Hz in the Wigner results (Fig. C-1(c)). As shown in Fig. C-1(d), the cross terms in the Born-Jordan distribution spread between the signal frequencies.

C.2 Method

To investigate the application of other time-frequency methods in aspiration detection, we calculated the Born-Jordan distribution of the breath sounds of a dysphagic individual of the aspirated group of patients. The goal was to investigate if the features obtained from the Born-Jordan distribution can improve the aspiration detection results. At the first step, we looked into two breath sounds of a dysphagic individual; one breath was related to a swallow without aspiration and the other one followed an aspiration event.

To compare the results with the spectrogram, and analyze the frequency contents of the distribution, we calculated the frequency marginal (C-4), which shows the energy density spectrum of the signal. We seek to gain insight into the characteristics of the Born-Jordan energy distribution of the signal; whether they can be used to distinguish between the aspirated and non-aspirated swallows. Then, we calculated the average power features at the frequency bands below 300 Hz, which are similar to those obtained from the PSD of the signal. Finally, to compare the classification results of the two time-frequency features in detecting aspiration, we applied the fuzzy c-means clustering algorithm to the Born-Jordan average power features.

C.3 Results

Figure C-2 shows the spectrogram and the Born-Jordan distribution of the breath sound after a swallow without aspiration. The same plot for the breath sound related to an aspiration event is shown in

Fig. C-3. The PSD and the Born-Jordan frequency marginal of the two breath sounds were shown in Fig. C-4 and Fig. C-5, respectively. The classification results in terms of the TP, TN, FP, and FN are shown in Table C-2.

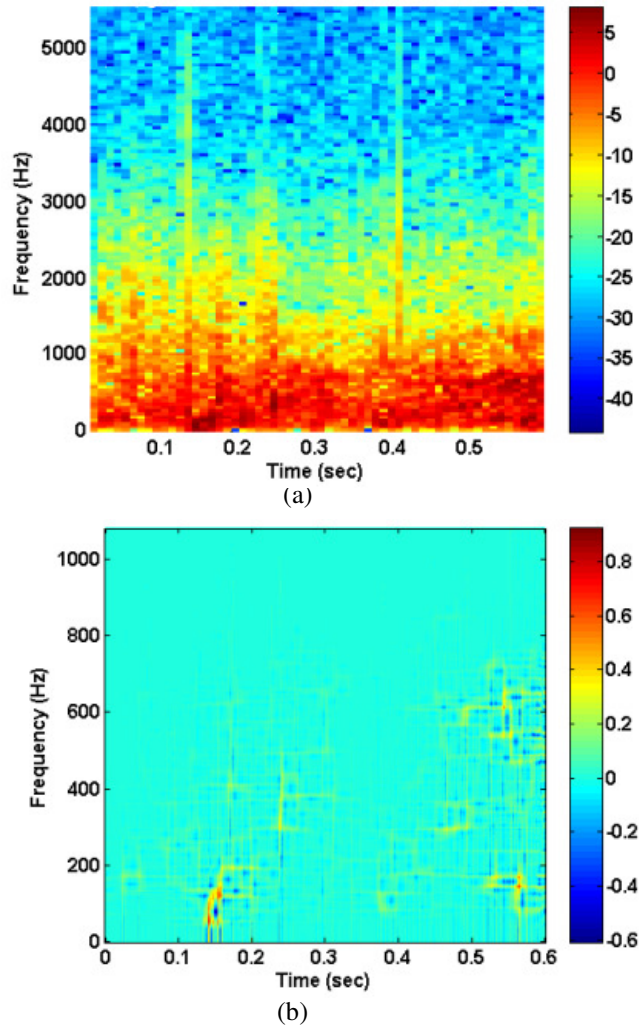


Fig. C-2 (a) The spectrogram of a breath sound after a non-aspirated swallow, (b) the Born-Jordan distribution of the same signal.

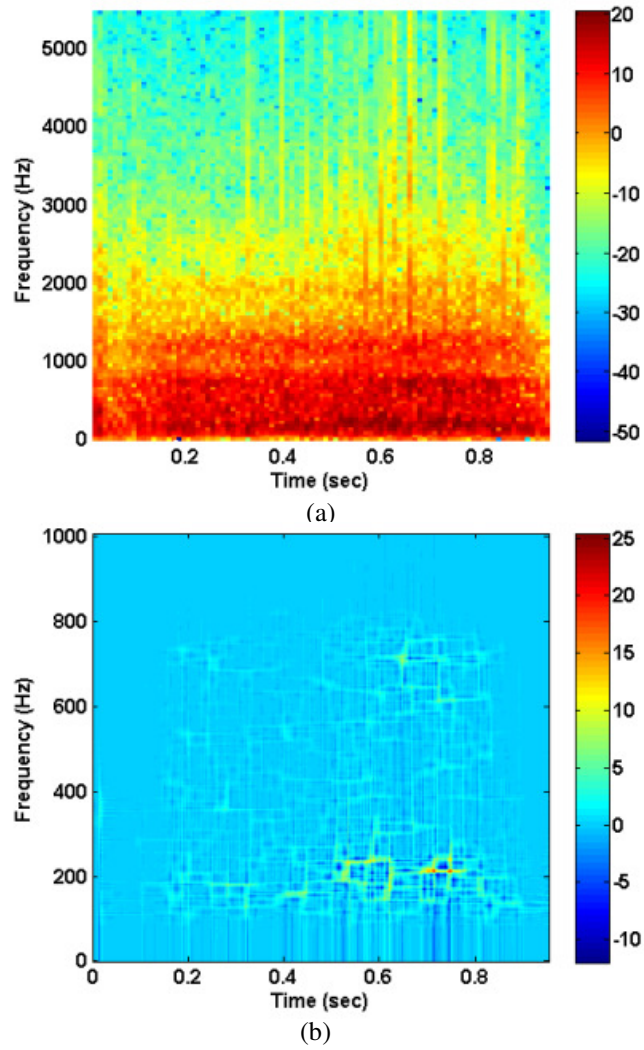
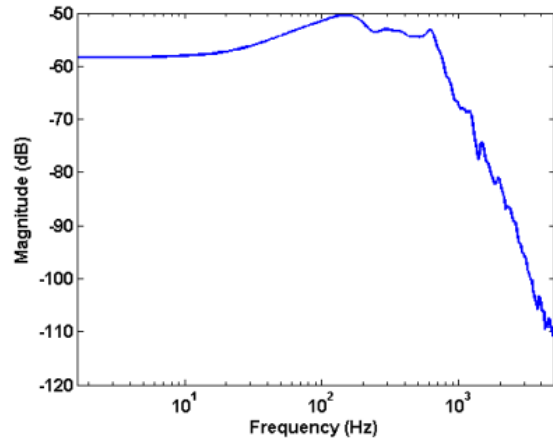
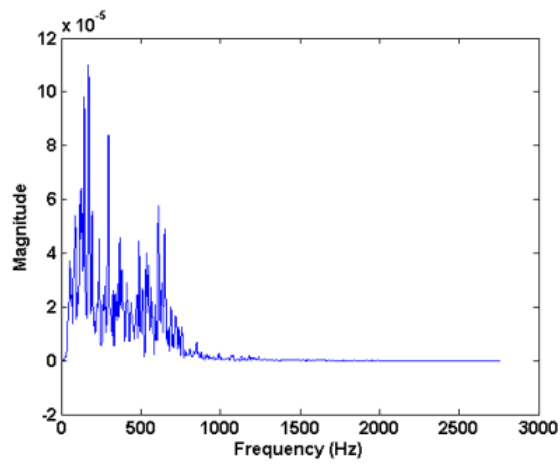


Fig. C-3 (a) The spectrogram of a breath sound following an aspiration, (b) the Born-Jordan distribution of the same signal.



(a)



(b)

Fig. C-4 (a) The PSD of a breath sound after a non-aspirated swallow, (b) the frequency marginal of Born-Jordan distribution of the same signal.

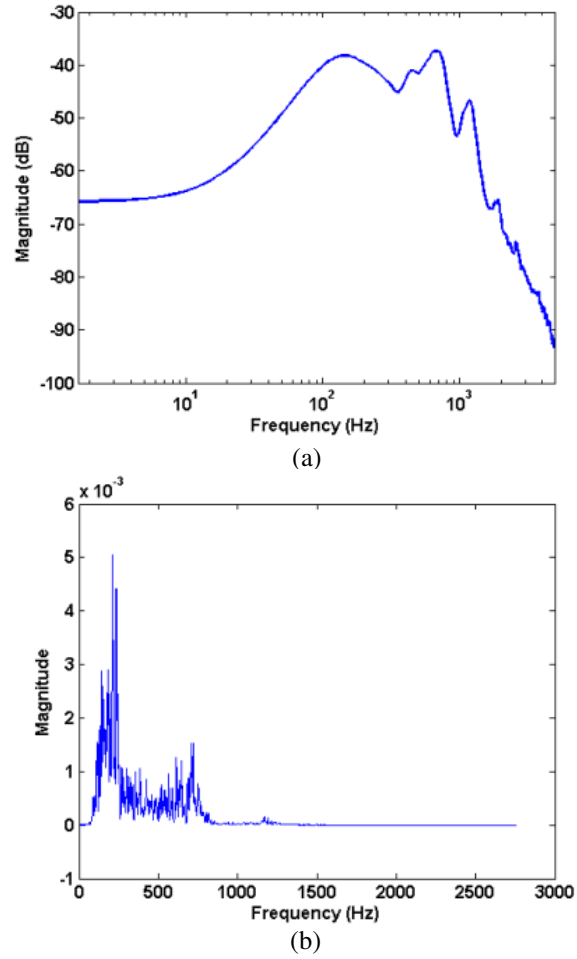


Fig. C-5 (a) The PSD of a breath sound following an aspirated swallow, (b) the frequency marginal of Born-Jordan distribution of the same signal.

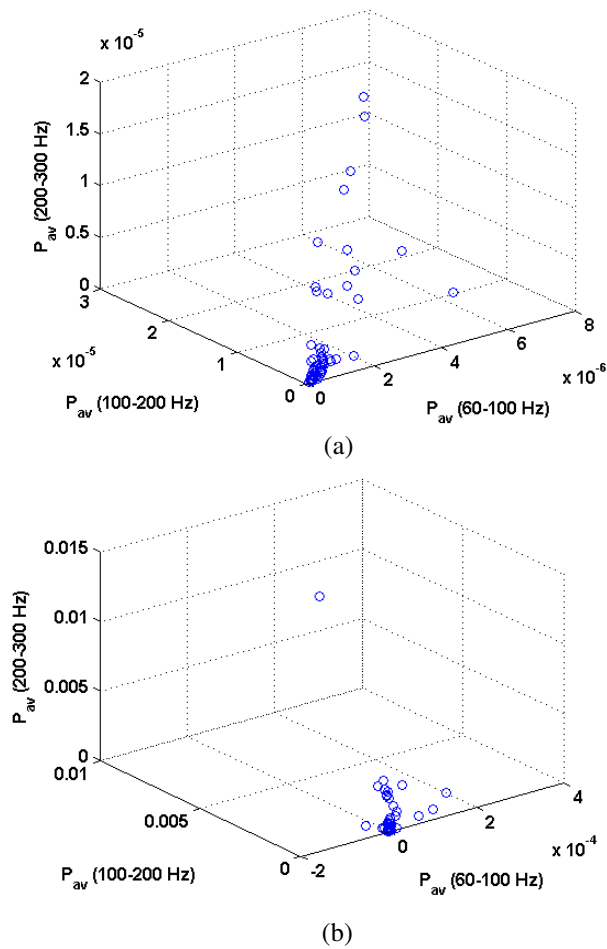


Fig. C-6. Scatter plot of the three average power features of the breath sounds after swallows of a patient having aspiration by using (a) PSD (b) the frequency marginal of Born-Jordan distribution of the same signal.

Table C-2. The classification results of aspiration detection by using the features obtained from two time-frequency representation methods: spectrogram and Born-Jordan.

Method	Total no. of swallows	No. of swallows with aspiration	TP	FP	TN	FN
Spectrogram	19	10	8	2	7	2
Born-Jordan	19	10	2	0	8	9

C.4 Discussion

The relative performance of the various time-frequency distributions is mainly goal/signal dependent. Although the bilinear time-frequency representations (excluding spectrogram) satisfy the marginal properties, they are not preferred when the visual interpretation is desired. The appropriate cross terms attenuation and localization are the important properties required for the visual interpretation of the time-frequency content of the time-varying signal[73]. Also, the computational time should be considered when comparing different time-frequency methods. The spectrogram provides the best visual interpretation characteristics in spite of its poor time-frequency resolution. On the other hand, the Born-Jordan TFR does not suffer from the resolution. However, it lacks the visual interpretation that helps the initial selection of features.

The classification results by using the spectrogram features outperform the Born-Jordan. However, it does not mean that the Born-Jordan representation conveys less information compared to the spectrogram. It might be due to the features that were extracted from the Born-Jordan distribution. The Born-Jordan satisfies the marginal properties as in (C-3) and (C-4). Therefore, calculating the integral over the time yields the frequency distribution of the signal. If we display the x-axis of Fig. C-4(b) and Fig. C-5(b) in logarithmic scale for and show the values of the y-axis (frequency marginal) in dB unit, we obtain Fig. C-7, which resembles the periodogram. Periodogram is not a consistent estimate of the true power density spectrum. Thus, the smoothing techniques such as the Welch power spectrum estimate, which offers trade-off between the bias and the variance of the estimation, existed. Since we calculated the PSD by using the Welch method, the features. In other words, selecting other sets of features can improve the classification results. This can be investigated in future studies.

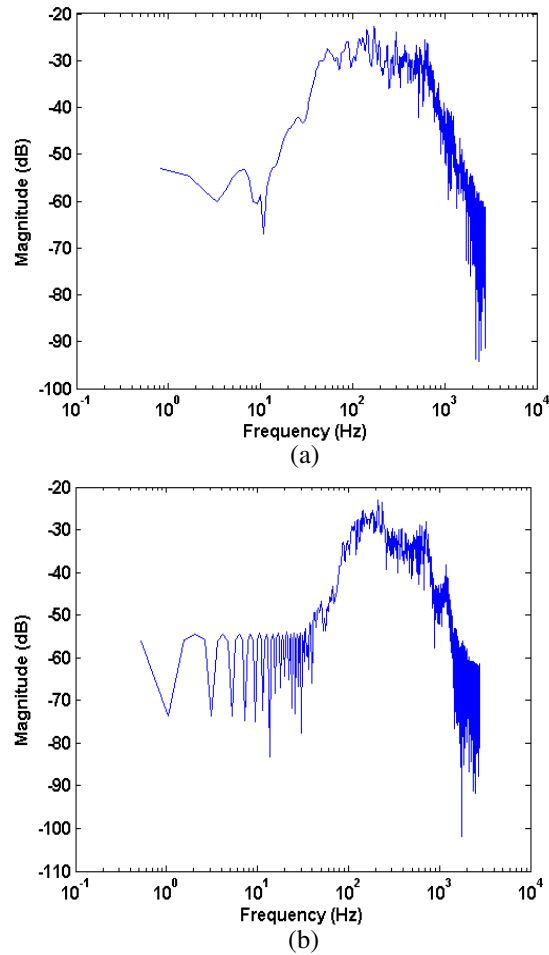


Fig. C-7 The frequency marginal of Born-Jordan distribution of the breath sound following (a) an non-aspirated swallow as shown in Fig. C-4 and (b) an aspirated swallow as shown in Fig. C-5.

Appendix D Frequency Response of the Microphone

D.1 Introduction

In this thesis, the swallowing and breathing sounds were recorded by using a Sony ECM-77B microphone, which is an omni-directional electret condenser microphone. The objective of this appendix is to ensure that the frequency response of the microphone does not affect the acoustic characteristic of the swallowing and breathing sound.

There are different methods for the measurement of the frequency response of a microphone, which can be found in IEC (International Electrotechnical Commission) standards such as 61094-2, 61094-3 and 1094-4 [74]. These standards describe the conditions under which the measurements are performed. The IEC 60268-4 standard specifies that all the measurements should be performed in the free field while the sound source generates the plane or spherical waves. To have the plane or spherical waveforms, the sound source should be small compared to the sound wavelength at the specific test frequency and has to be far away from the microphone. However, this is not always practical. Therefore, we defined our measurement conditions, which are described in the following section.

D.2 Measurement Procedure

The near field effect of the microphone was tested in an anechoic chamber. The microphone was placed in front of the speaker (Cyber Acoustics desktop speaker system, CA-2100) in two different locations: 1- directly connected to the speaker, and 2- hanging at the distance of 25 cm from the speaker. The microphone was connected to the digital audio recorder (Edirol R-44), which recorded the sounds at 44100 Hz sampling frequency. The audio amplifier filters the signal between 20-22000 Hz.

We generated the sinusoidal wave signals of pure tones at different frequencies (Table D-1). For each of the frequencies listed in Table D-1, the pure tone sound was generated at 44100 Hz, and was held for 4 seconds. Each of the 4 second pure tone sine waves was then followed by 1 second of silence; thus, we could easily discriminate between different frequencies.

The sound was generated using Matlab, and was played on an Arduino platform connected to the speaker. The sound signal was recorded on an SD card, which was mounted on the Arduino board. It should be noted that a laptop could not be used to play the sound as laptops' sound output are considerably distorted for our purpose due to embedded filtering of the sound output of the laptop. This was tested in our measurement.

Table D-1. The frequencies used to test the microphone.

Frequency interval	Incremental step
50 Hz-1000 Hz	50 Hz
1000 Hz-10000 Hz	1000 Hz
10000 Hz-20000 Hz	2000 Hz

The spectrogram representation of the sound played through the speaker is shown in Fig. D-1. The blue vertical segments represent the 1 second silence segments between the pure tone sound signals.

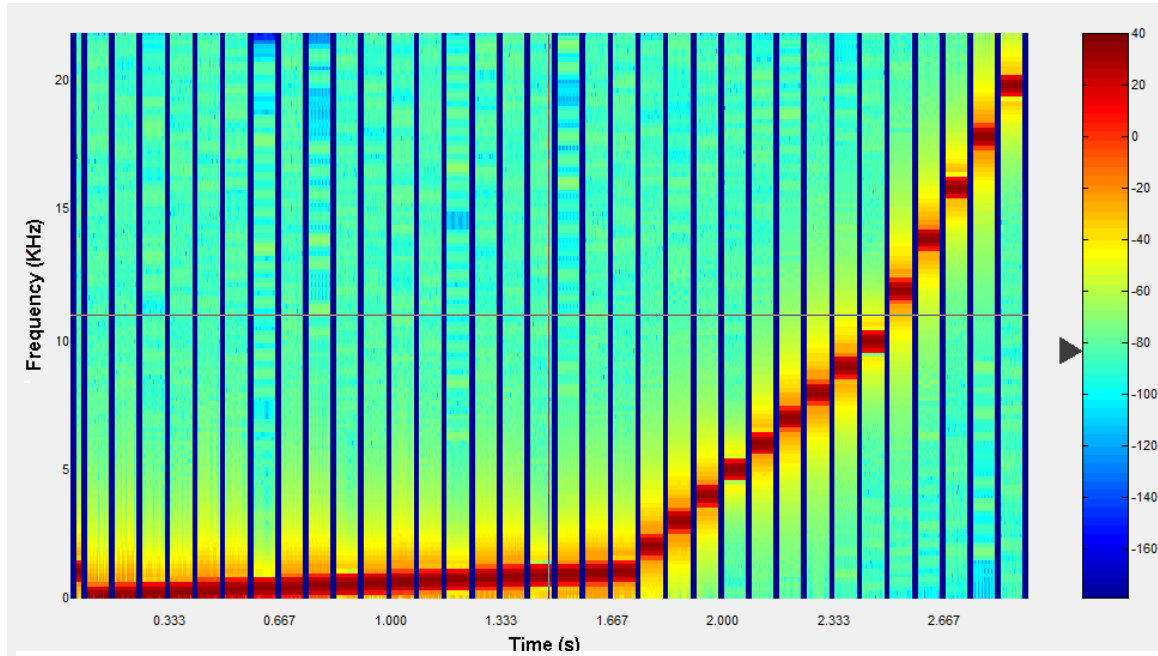


Fig. D-1. Spectrogram of the sound signals played through the speaker.

D.3 Results

D.3.1 Direct Connection of the Microphone to the Speaker

The spectrogram of the sound signal recorded by the Sony ECM 77-B microphone placed on the speaker is shown in Fig.D-2.

Comparison of Fig.D-2 with Fig. D-1, shows that the distortion occurred at high frequencies. As can be seen in Fig.D-2, the distortion appears at 10 KHz, which is in agreement with the manufacturer's frequency response.

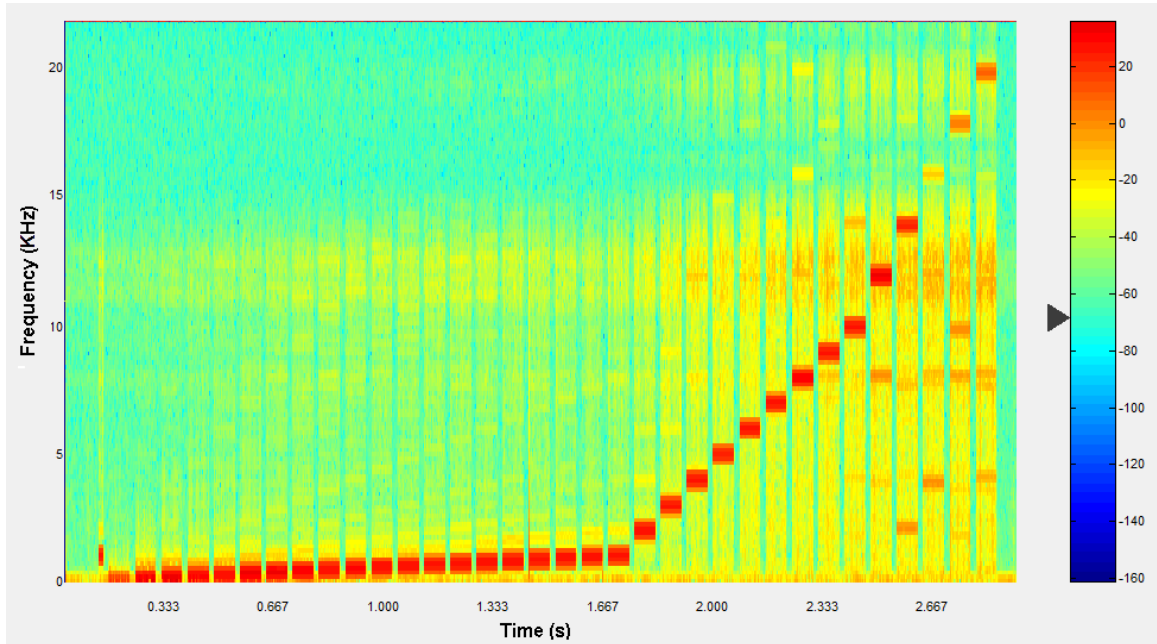


Fig.D-2. Spectrogram of the sound signal, recorded by the Sony ECM 77-B microphone, placed on the speaker.

D.3.2 Microphone in front of the Speaker at 25cm Distance

We repeated the test by changing the recording configuration. For this part, we hung the microphone in the air, 25 cm away from the speaker. Fig. D-3 shows the spectrogram of the sound signal recorded by the microphone.

D.3.3 Frequency Response

To visualize the frequency response of the sensors, we plotted the amplitude of the signal recorded by each sensor at the different frequencies listed in Table 1. Also, to check the consistency, we repeated each measurement three times. Fig. D-4 shows the frequency response of the Sony ECM-77B microphone in three recordings. The magnitude of the responses (y -axis) is normalized to the maximum value. The variation of the frequency response of the Sony ECM 77-B

microphone occurred within 10 dB over the frequency range between 100Hz to 10KHz.

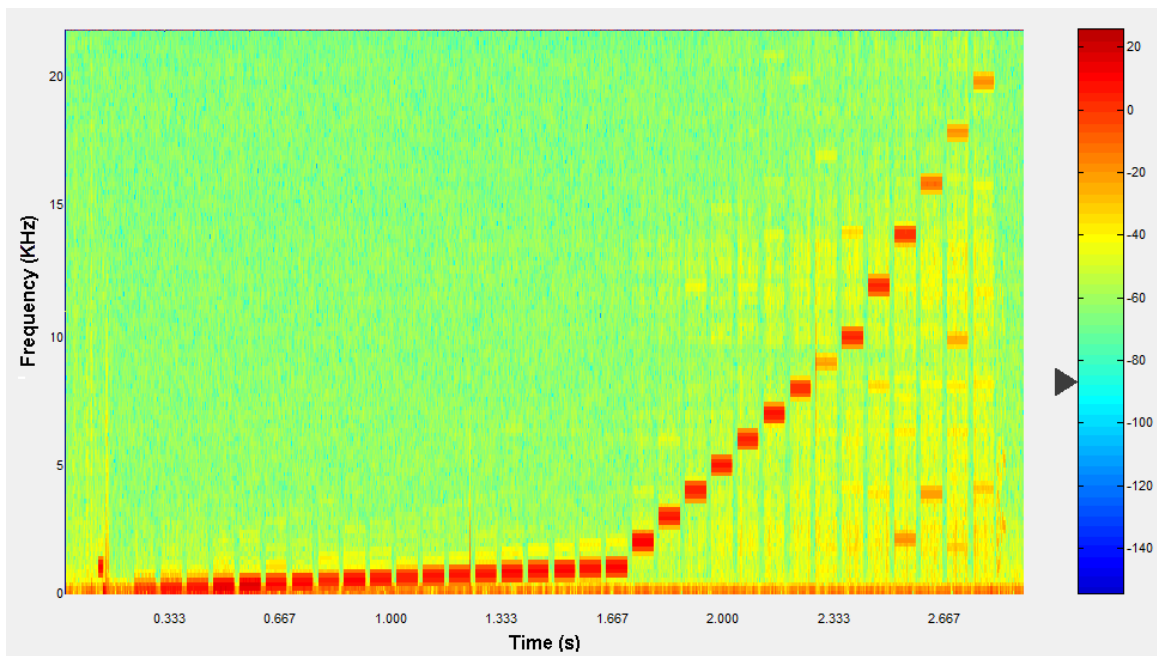


Fig. D-3. Spectrogram of the sound signal, recorded by the Sony ECM 77-B microphone, hanging in front of the speaker at the distance of 25 cm.

Appendix D- Frequency Response of the Microphone

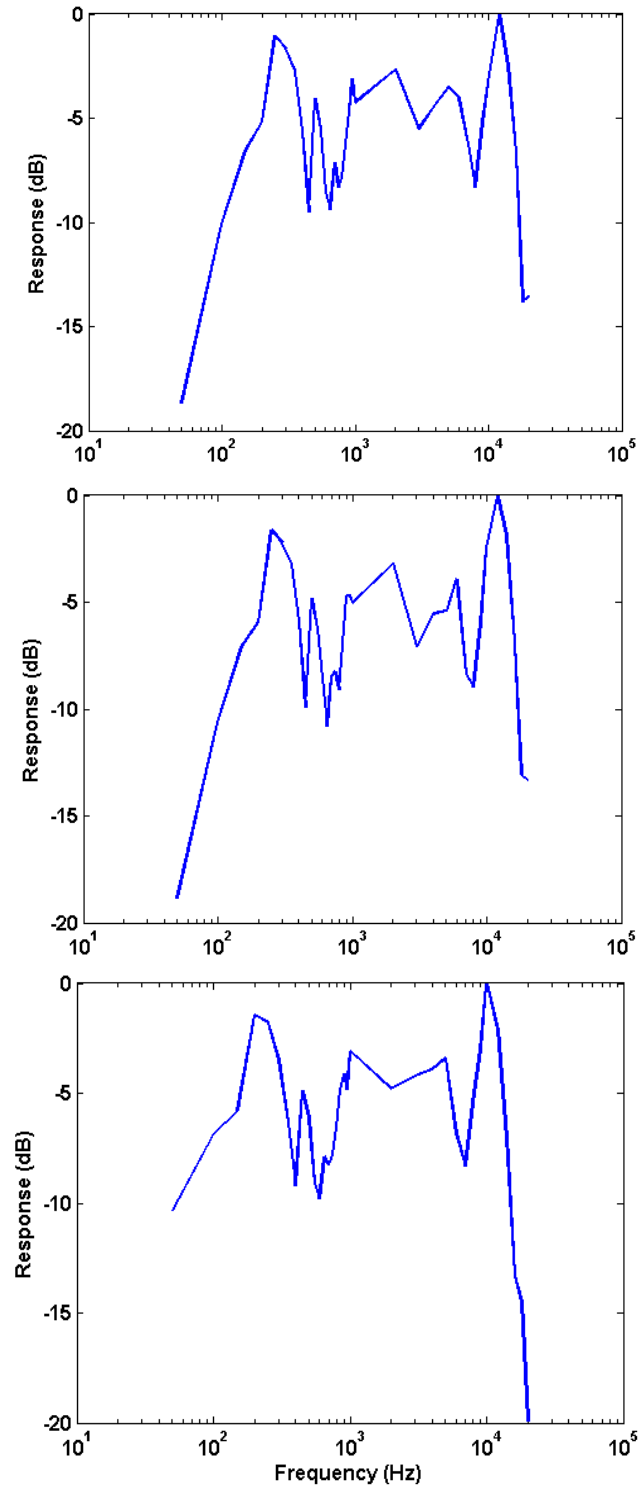


Fig. D-4. Three measurements of the frequency response of the Sony ECM 77-B microphone hanging in front of the speaker at the distance of 25 cm.

D.4 Conclusion

The frequency response of the microphone by which all the swallowing and breath sound data were recorded in this thesis has been studied. The results show that the frequency response matches the specification provided by the manufacturer in the data sheet.