ANALYSIS AND MODELING OF SWALLOWING SOUNDS

by

Mohammad Aboofazeli

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 Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba

Copyright © 2006 by Mohammad Aboofazeli

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

ANALYSIS AND MODELING OF SWALLOWING SOUNDS

BY

Mohammad Aboofazeli

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

DOCTOR OF PHILOSOPHY

Mohammad Aboofazeli © 2007

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

ACKNOWLEDGMENT

First, I would like to thank my advisor, Dr. Zahra Moussavi, for her guidance, commitment and constructive criticism. Thanks are also extended to my advisory committee, Dr. Edward Shwedyk, Dr. Gabriel Thomas, and Dr. Madjid Birouk, for their contributions.

Also, I wish to express my sincere appreciation to my family, for their encouragement, understanding, and patience.

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) and Telecommunications Research Laboratories (TRLabs) and I thank them for their support.

ABSTRACT

Swallowing function is a complicated process involving several highly coordinated events. In recent years, acoustical analysis of the swallowing mechanism has received considerable attention due to its diagnostic potentials. This thesis presents novel approaches for analysis and modeling of swallowing sounds in individuals with and without swallowing disorder (dysphagia).

Different techniques based on nonlinear dynamics, recurrence quantification analysis, hidden Markov model (HMM), and multiresolution wavelet analysis were used to detect characteristic features of swallowing sounds for automatic swallowing sound detection, swallowing sound segmentation, as well as classification of normal and dysphagic swallowing sounds. Data from 27 healthy subjects and 11 dysphagic patients were used, in which swallowing and breath sounds were recorded with an accelerometer placed over the trachea (suprasternal notch), while the subjects were fed three different textures with a bolus size of 5 ml. Submental electromyogram signals were also recorded in 12 normal subjects simultaneously with their swallowing sounds. In terms of swallowing sound detection in the swallow and breath sound recordings, the performance of the HMM based method using recurrence plot features was superior to that of the other methods. As for the segmentation of the swallowing sounds, multiscale products of wavelet coefficients along with an HMM yielded the least error in detecting the boundaries between the segments. Lastly, in terms of classification of normal and dysphagic swallowing sounds, nonlinear metric tools, i.e., correlation dimension and time delay, resulted in a high accuracy of 83%. The results of HMM based methods for classification of swallowing sounds between the two groups of healthy and dysphagic subjects using

RMS achieved 85.5% accuracy. The results of the study on the timing of submental muscle contraction in relation to swallowing sounds showed that the movement of the larynx precedes the first audible segment of swallowing sounds. It also showed that the lag between the onset of submental muscle contraction and the beginning of initial discrete sounds is shorter for thin liquid texture.

Overall, the outcomes of this study have paved the way for a better understanding of the swallowing mechanism and improving clinical assessment techniques of swallowing disorders using its acoustic signatures.

TABLE OF CONTENTS

ACKNOWLEDGEMENT II

ABSTRACT III

TABLE OF CONTENTS V

LIST OF FIGURES VII

LIST OF TABLES IX

CHAPTER 1 INTRODUCTION 1

1.1 Motivation 1

1.2 Thesis Scope 3

1.3 Thesis Organization 4

CHAPTER 2 BACKGROUND 6

2.1 Anatomy and Physiology of the Swallowing Mechanism 6

2.2 Swallowing Assessment Techniques 9

2.3 Acoustical Analysis of the Swallowing Mechanism 12

CHAPTER 3 BACKGROUND THEORY 18

3.1 Nonlinear Dynamics 18

3.2 Hidden Markov Models 35

CHAPTER 4 METHODS 39

4.1 Experimental Data 39

4.2 Swallowing Sound Detection 41

4.3 Swallowing Sound Segmentation 49

4.4 Swallowing Sound Characterization 53

4.5 Swallowing Sound Classification 53

4.6 Timing Analysis of Submental Electromyogram 55

CHAPTER 5 RESULTS AND DISCUSSION 58

5.1 Swallowing Sound Detection 58

5.2 Swallowing Sound Segmentation 72

5.3 Swallowing Sound Characterization 76

5.4 Swallowing Sound Classification 77

5.5 Timing Analysis of Submental EMG 80

CHAPTER 6 CONCLUSIONS 83

6.1. Concluding Remarks 83

6.2. Recommendations for Future Studies 87

APPENDIX A THRESHOLD SELECTION USING THE OTSU METHOD..... 89

REFERENCES 91

LIST OF FIGURES

1.1	Block diagram of this study.	4
2.1	Midsagittal section of the swallowing mechanism.	9
2.2	Three segments of a typical swallowing sound signal.	14
2.3	A typical sequence of swallowing and breath sounds and the corresponding spectrogram.	16
3.1	Three state variables for the Lorenz model.	22
3.2	Trajectory for the Lorenz model.	23
3.3	Reconstruction of 2-dimensional trajectory of a signal.	26
4.1	A typical tracheal sound signal decomposed by multiresolution analysis.	43
4.2	Schematic description of the wavelet transform based filtering method.	44
4.3	The waveform of EMG signal in time domain and its RMS.	56
5.1	A typical tracheal sound signal decomposed by the wavelet-based filter, extracted swallowing sound signal, and extracted breath sound signal.	59
5.2	Airflow signal, spectrogram and tracheal sound waveform as a function of time.	60
5.3	False nearest neighbors percentage as a function of embedding dimension.	61
5.4	Average displacement as a function of time delay for different embedding dimensions.	63
5.5	Recurrence plots of breath and swallowing sounds.	64
5.6	A tracheal swallowing sound signal, %REC, %DET, ratio of %DET to %REC, and entropy of the length of the lines parallel to the identity line.	68
5.7	A tracheal swallowing sound signal, the maximum and average length of the lines parallel to the identity line, the average of Poincare recurrence times of the first and second type, and Shannon entropy of the tracheal sound signal.	69
5.8	The tracheal sound signal and the boundaries detected by thresholding of the %REC, ratio, and entropy.	70

5.9 A swallowing sound signal, its WFD, and the most likely hidden states of the swallowing sounds. 73

5.10 A swallowing sound signal in the time-domain and the most likely hidden states when WFD, RMS, MSP1, MSP2, MSP3, average power, and LPC were used to model the swallowing sound signal. 75

5.11 Correlation dimension for IDS and BTS segments in normal and dysphagic subjects. 76

5.12 Classification accuracy for different number of states and features. 79

5.13 Waveform of a swallowing sound signal, an EMG signal, and the boundaries of IDS segment of the swallowing sounds and muscle contraction period of EMG. 81

LIST OF TABLES

5.1	Mean ± standard deviation for recurrence parameters of the swallowing and breath sound signals.	66
5.2	Performance comparison of the swallowing sound detection algorithms.	71
5.3	Mean ± standard deviation for delay in detection of boundaries of swallowing sound phases obtained by HMM in different textures.	74
5.4	Optimum time delay ranges for IDS and BTS segments of swallowing sound signals in normal and dysphagic subjects.	78
5.5	Duration of swallowing sound phases and submental muscle contraction in different textures.	81

CHAPTER 1

INTRODUCTION

1.1 Motivation

Swallowing function is a complicated process involving several highly coordinated events. During a swallow, the bolus passes through the pharynx, which is a shared pathway for ingestion and respiration. Accordingly, coordination of both the swallowing mechanism structures and the swallowing and respiratory processes are essential for a safe feeding to occur. Lack of coordination may result in aspiration, i.e., entry of the bolus into the airway. Aspiration is a rare occurrence in healthy individuals; conversely, it is very likely in patients with swallowing disorders.

Difficulty in moving food from the mouth to the stomach is called dysphagia. Swallowing disorders can occur in all age groups as a result of congenital abnormalities, structural damage, and/or medical conditions such as cerebral palsy, head trauma, spinal cord injury, and damage after treatment for the oral, oropharyngeal and laryngeal cancer [Logeman, 1998]. Approximately 30-40% of the patients who have had cerebrovascular accidents suffer from complete or partial swallowing disorders; although some of them return to normal condition, generally prolonged treatment and rehabilitation are required [Reddy *et al.*, 1990]. Dysphagic patients are at risk of repeated choking, pneumonia, malnutrition, and dehydration [Logeman, 1998]. These health hazards reveal the essential need for assessment and treatment of swallowing disorders. This has motivated many investigators to study normal and abnormal swallowing mechanisms and develop

quantitative and objective swallowing assessment techniques.

Currently, the gold standard for swallowing assessment is videofluoroscopy swallowing study (VFSS), in which patients are given the bolus mixed with a contrast agent. The movement patterns of the oral cavity, pharynx, and esophagus are observed with a continuous X-ray stream [Palmer *et al.*, 1993]. VFSS provides immediate visual inspection of the swallowing mechanism, localization of abnormalities, and detection of aspiration when it occurs. However, VFSS requires radiation exposure; hence, it can only be run for a short duration. Therefore, it is quite possible to miss aspirations during VFSS for patients with mild dysphagia. Accordingly, there is a need to develop non-invasive techniques to assess the swallowing mechanism and its abnormalities.

Auscultation (listening to swallowing sounds with a stethoscope) is the primary tool used by clinicians as a noninvasive technique for the clinical evaluation of the swallowing process. Auscultation is applicable in a wide variety of feeding circumstances and may be used for evaluation of the pharyngeal phase of swallow [Vice *et al.*, 1990]. However, the quality of information obtained by auscultation depends on the perceptual skills of the examiner and utilization of appropriate instrumentation and technique [Hamlet *et al.*, 1994]. In recent years, digital signal processing of swallowing sounds has drawn considerable attention (e.g. [Cichero and Murdoch, 2002] and [Lazareck and Moussavi, 2004]); swallowing sounds are recorded using a microphone and/or accelerometer and analyzed by digital signal processing techniques.

This thesis presents novel approaches for analysis of swallowing sounds based on nonlinear dynamics theory and modeling of swallowing sounds using hidden Markov models. It is hoped that the outcomes of this study will result in a better understanding of

the swallowing mechanism and improve clinical assessment techniques of swallowing disorders using its acoustic signatures.

1.2 Thesis Scope

Currently, in clinical assessment of swallowing sounds, swallowing sound signals are listened and/or recorded by placing a stethoscope, microphone, and/or accelerometer on the subject's neck, either laterally or over the suprasternal notch. Since the recorded sound signals contain both swallowing and breath sounds, the swallowing sounds must first be extracted from the recorded signals. Currently, this is done in a time-consuming manual manner by repeated listening. In this study, swallowing sounds were recorded by placing a stethoscope over the suprasternal notch of the subjects. Then, three novel methods were investigated for automated detection of swallowing sounds in the records of breath and swallowing sounds.

Swallowing sound signals do not have a unique temporal and durational pattern even if recorded from a subject, being fed a particular size and texture of the bolus. Additionally, the phases of swallowing sounds, especially initial discrete sounds and bolus transit sounds, are not easily distinguishable from each other. Therefore, detection of the swallowing sound phases without videofluoroscopy data is a challenging task. This thesis presents a novel technique for swallowing sound segmentation.

Characterization of a signal facilitates signal processing and understanding of the underlying mechanism. This thesis presents the outcomes of the study performed on the characteristics of swallowing sound signals using nonlinear dynamic parameters. In this study, the discriminating ability of nonlinear dynamic parameters for the classification of

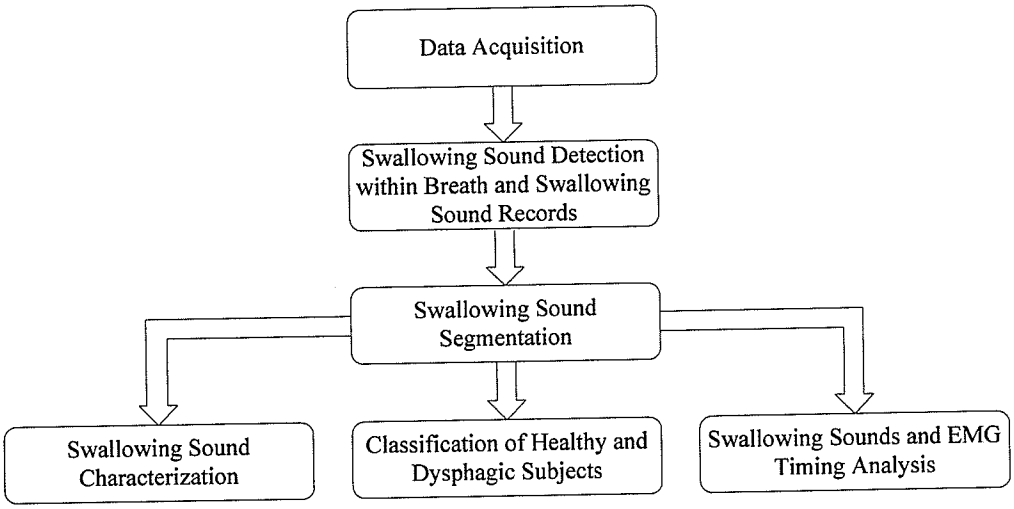


Fig. 1.1. Block diagram of this study.

swallowing sounds of healthy and dysphagic subjects was confirmed. Furthermore, a hidden Markov model (HMM) based method for the classification of swallowing sounds of healthy and dysphagic subjects was proposed. Several features were used for this HMM based classification. A comparison between features was performed in order to determine which feature(s) provided the best performance.

This thesis also reports the outcomes of the study on swallowing mechanism timing, based on the analysis of submental surface electromyogram (EMG) signals, which were simultaneously recorded with the swallowing sounds. The durational and temporal patterns of the submental muscle activities were analyzed using the EMG signal.

1.3 Thesis Organization

This thesis is organized in six chapters. This chapter provides a general introduction and the thesis scope. A background on the anatomy and physiology of the swallowing

mechanism, the techniques currently used for the clinical assessment of the swallowing mechanism, and the characteristics of swallowing sound signals are presented in Chapter 2. Chapter 3 provides the theoretical background of the methods used in this study. The methods used for the automated detection of swallowing sounds are described in Chapter 4. In addition, Chapter 4 presents the methods used for the analysis of swallowing sounds, based on nonlinear dynamics theory and hidden Markov modeling and their applications in swallowing sound segmentation as well as classification of swallowing sound signals of healthy and dysphagic subjects. Chapter 5 provides detailed results of this study, including outcomes of the nonlinear analysis of the signal and the comparison of different methods for detection, segmentation and classification of swallowing sound signals. The concluding remarks, contributions of the thesis, and recommendations for future studies are presented in Chapter 6.

CHAPTER 2

BACKGROUND

This chapter elaborates on the anatomy and physiology of the swallowing mechanism, followed by a discussion of standard and conventional assessment techniques. A literature review of recent studies on acoustical analysis of swallowing sounds, as an alternative method for swallowing assessment, will then be provided.

2.1 Anatomy and Physiology of the Swallowing Mechanism

The anatomic area of the swallowing mechanism includes the oral cavity, pharynx, larynx, and esophagus. The midsagittal section of the swallowing mechanism is shown in Fig. 2.1. Structures in the oral cavity include the lips, teeth, hard palate, soft palate, uvula, mandible, floor of the mouth, tongue, and faucial arches [Logeman, 1998]. The pharynx is a part of the digestive tube and may be divided into nasal, oral, and laryngeal parts. The nasal part lies posterior to the nose and is entirely for respiration, whereas the oral and laryngeal parts provide a shared pathway for food and air intake [Last, 1978]. At the base of the tongue, the pharynx opens to the larynx. The larynx is a valve in the air passage that connects the pharynx with the trachea; it is primarily a valve to prevent food from entering the airway during swallowing [Logeman, 1998]. The esophagus is a canal for ingestion that extends from the pharynx to the stomach [Gray, 1985].

According to [Logeman, 1998], the act of swallowing may be divided into four distinct phases:

1) The oral preparatory phase: After the food is placed in the mouth, it is manipulated so that it is ready to be swallowed. Movement patterns in the oral preparatory phase depend on the viscosity of the material to be swallowed. The larynx and pharynx are at rest in the oral preparatory phase, the airway is open and nasal breathing continues.

2) The oral phase: This phase is initiated with the backward movement of the bolus by the tongue. The pressure of the oral tongue against the palate increases when the food viscosity thickens. The oral phase typically takes less than 1 to 1.5 seconds. Posterior movement of the bolus results in the stimulation of the sensory receptors in the oropharynx and tongue and in sending sensory information to the cortex and brain stem. The sensory information is decoded and the pharyngeal phase is triggered.

3) The pharyngeal phase: After the pharyngeal swallow is triggered, elevation and retraction of the velum close the velopharyngeal port and prevent material from entering the nasal cavity. Contraction of the floor of mouth muscles (including the mylohyoid, geniohyoid, and the anterior belly of digastric) causes elevation and forward movement of the larynx and hyoid bone, which contribute to the closure of the airway entrance and the opening of the upper esophageal sphincter, respectively. In a normal swallow, velopharyngeal closure and movement of the larynx and hyoid occur almost simultaneously. Closure of the larynx begins at the level of vocal folds and progresses upward. Closure of the larynx and epiglottis contribute to the closure of the airway. Opening of the upper esophageal sphincter allows material to pass from the pharynx into the esophagus. In normal swallows, the opening of the upper esophageal sphincter and closure of the airway usually begin simultaneously. Then, ramping of the base of the tongue directs the bolus into the pharynx. It is followed by tongue base retraction and

progressive top to bottom contraction in the pharyngeal constrictors.

4) The esophageal phase: In this phase, esophageal peristalsis carries the bolus through the cervical and thoracic esophagus and into the stomach. The time taken for the bolus to move from the point where it enters the esophagus to the stomach entrance is between 8 to 20 seconds. In the esophageal phase, the epiglottis returns to its rest position and the airway opens while the esophageal peristalsis movement continues. According to our own observation using videofluoroscopy images, the return of the epiglottis takes less than 30 ms.

The duration and characteristics of each of the swallowing phases depend on several factors including the type and volume of the food being swallowed and the voluntary control exerted over it [Kahrilas and Logeman, 1993]. For example, the duration of the airway closure tends to increase as the bolus volume increases. The bolus volume depends on the food viscosity. Small boluses of higher viscosity food pass easily through the pharynx and the upper esophageal sphincter. For thin liquids, the bolus volume ranges from 1 ml (saliva bolus) to 20+ ml (cup drinking), whereas the maximum volume of higher viscosity boluses decreases; for example, swallows of pudding may be 5 to 7 ml on average, and swallows of thicker mashed potatoes may be 3 to 5 ml.

Since the characteristics of the swallowing phases are functions of the viscosity and volume of the food, in most swallowing studies (including this study), the subjects are fed a particular texture at a controlled volume.

Swallowing and respiration are reciprocal functions, i.e., respiration halts during the pharyngeal phase of a swallow. The airway is entirely open during the oral preparatory and oral phases; it is partially open during the esophageal phase. Lack of swallow-

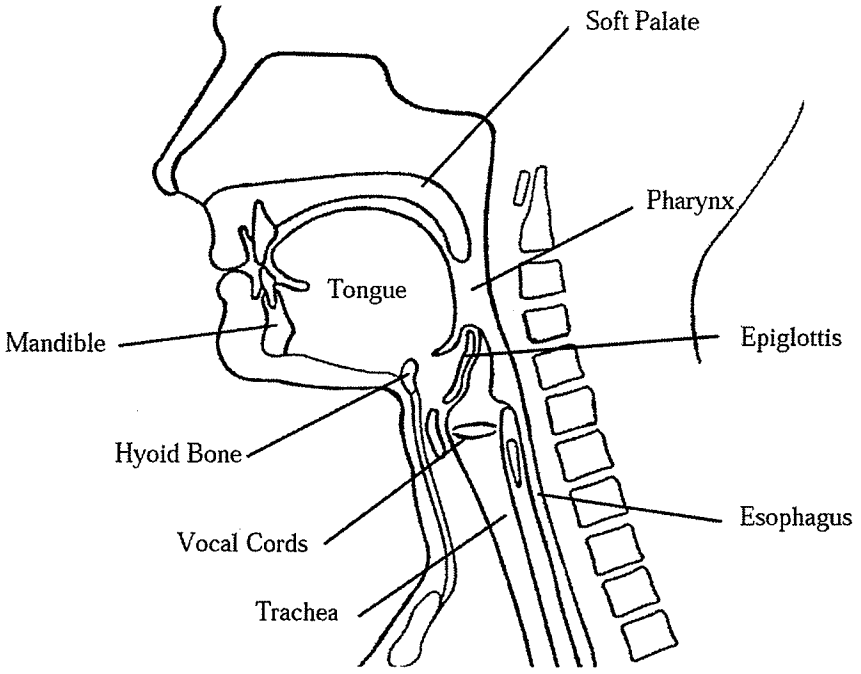


Fig. 2.1. The midsagittal section of the swallowing mechanism.

respiratory coordination may result in aspiration. Aspiration rarely happens in healthy individuals whereas the occurrence of fault in swallowing mechanism is very likely in dysphagic patients [Logeman, 1998].

The normal anatomy of the upper airway in a young infant is not the same as that of an adult [Logeman, 1998]. Accordingly, characteristics of a normal swallow in young infants differ from that of the adult.

2.2 Swallowing Assessment Techniques

Several instrumental techniques have been used for swallowing mechanism assessment, which may be categorized into imaging and non-imaging techniques.

Imaging techniques include: ultrasonography, videoendoscopy, videofluoroscopy, and scintigraphy. Non-imaging techniques include: electromyography, pharyngeal manometry, and auscultation.

Ultrasound analysis of the oral cavity has been used to observe the tongue function, as well as the motion of the hyoid bone and larynx, during swallowing [Shawker *et al.*, 1984]. However, due to the mix of tissue types in the pharynx, ultrasonic scanners cannot image the pharynx. This fact limits the application of this method [Logemann, 1998].

Videoendoscopy has been used to examine the anatomy of the oral cavity and pharynx, in addition to examining the pharynx and larynx before and after swallowing. This method can be performed using a flexible scope inserted into the nose, down to the level of the soft palate. When the pharyngeal swallow triggers, closure of the pharynx blocks the image; this is a major drawback, as many important events occur during this period [Logeman, 1998]. Additionally, the invasive nature of videoendoscopy is a drawback for this technique.

The most frequently used technique in oropharyngeal swallow assessment is videofluoroscopy. It enables visualization of the movement of the oral cavity structures, larynx, hyoid, tongue base, pharyngeal walls and cricopharyngeal region, along with monitoring the bolus position. The food mixed with a contrast agent, e.g. barium, is given to the patient and the movement patterns are monitored on the videofluoroscopy screen. As swallowing occurs very quickly, fluoroscopic studies are recorded on a videotape for further examination by the specialists; this allows them to follow the movements of the swallowing structures and detect plausible disorders. However, VFSS is based on X-ray; since the maximum amount of a patient's exposure to X-ray must be limited, it should be

run for the shortest possible duration.

Scintigraphy is a nuclear medicine test, in which the patient is fed a measured amount of a radioactive substance and the bolus is imaged by a gamma camera. Aspiration can be monitored by this technique, but the physiology of the mouth and pharynx cannot be visualized [Logeman, 1998].

The electromyography technique is used to evaluate the timing and strength of contraction of the particular muscles involved in the swallowing mechanism [Palmer, 1988]. The swallowing mechanism has been studied using surface EMG as in [Klahn *et al.*, 1999] and/or hooked-wire EMG [Perlman, 1989]. In surface EMG, the electrodes are placed on the skin surface above the muscle of interest.

Pharyngeal manometry is a technique in which several pressure transducers are placed on different locations, including the tongue base, cricopharyngeal sphincter, and cervical esophagus, and the pressure changes in these locations are measured [Brasseur and Dodds, 1991]. However, the use of pharyngeal manometry is limited due to its invasive nature.

The primary tool for swallowing assessment is auscultation, i.e., listening to the sounds of swallowing by a stethoscope placed over the neck of the subject. Auscultation is used for clinical evaluation of the pharyngeal swallow. This technique is noninvasive and easy to use. Stethoscopes are basically designed for listening to the heart and respiratory sounds. They are best suited for listening to sounds with a frequency spectrum below 1000 Hz; however, their frequency response is not flat over the frequency band of swallowing sounds (up to 2500 Hz). In addition, frequency responses of different stethoscopes are not the same, which may result in subjective diagnoses. In many

research studies, swallowing sounds have been listened to and/or recorded by microphones or accelerometers, which have frequency responses that are flat over a wide range of frequencies. When listening to the recordings of swallowing sounds recorded by microphones or accelerometers, clinicians have remarked that the recordings do not sound quite like what was heard by a stethoscope [Hamlet *et al.*, 1994]. The limited knowledge about the cause of swallowing sounds, and the differences between the swallowing sounds of normal and abnormal subjects has limited the use of auscultation.

2.3 Acoustical Analysis of the Swallowing Mechanism

There have been some attempts to identify the physiological cause of swallowing sounds. Based on the existence of several valves and pumps in the pharynx, which have analogous functions to the valves and pumps in the heart, it was speculated that an analogy existed between the generation of heart sounds and swallowing sounds [Cichero and Murdoch, 1998]. However, this hypothesis has not yet been proven.

In recent years, swallowing sound analysis has gained more credibility. The procedure and equipment used for auscultation has been a subject of study. The optimal site for recording the swallowing sounds of infants was reported to be at the lateral aspect of the larynx [Vice et al., 1990]. Also, Takahashi et al. [1994] investigated the reproducibility of swallowing sound patterns and the best-suited site for swallowing sound recording. Due to the greatest signal to noise ratio, they selected the lateral border of the trachea immediately inferior to the cricoid cartilage as the optimal site for detection of swallowing sounds. The suitability of this recording site was confirmed in a later investigation [Cichero and Murdoch, 2002]. In order to find suitable equipment for

auscultation of swallowing sounds, the frequency response of six popular stethoscopes were compared [Hamlet et al., 1994]. Although the characteristics of two stethoscope models were found to be more suitable for listening to swallowing sounds, all stethoscopes showed poor frequency response for frequencies higher than 1000 Hz. The performance of a microphone and an accelerometer for recording swallowing sounds was investigated by Cichero and Murdoch [2002]. Their results showed that the microphone provided superior signal to noise ratio.

Swallowing sound signals are non-stationary signals generated by the complicated physiological events in the swallowing mechanism. Although the swallowing mechanism in infants and adults is different, there are some common patterns in the swallowing sounds of both age groups. Vice et al. [1990] recognized these common patterns as initial discrete sounds (IDS), bolus transit sounds (BTS), and final discrete sounds (FDS). IDS precede and FDS follow the BTS. The opening of the cricopharyngeal sphincter, in order to allow the bolus to pass from the pharynx into the esophagus, is speculated to contribute to the most distinct characteristics of the swallowing sounds, i.e., IDS. They are mainly characterized by their time domain features, such as a relatively short duration. IDS may be repeated in a highly similar form. BTS are the gurgle sounds generated by bolus transmission into the esophagus during the pharyngeal phase. The amplitude and pattern of BTS may vary; they are functions of the bolus texture, particularly whether air is mixed with the bolus. FDS immediately precede the respiratory sound, which follows the swallowing sounds. FDS are usually very brief and are not necessarily present in all swallows. A typical swallowing sound signal and its common characteristic segments are shown in Fig. 2.2.

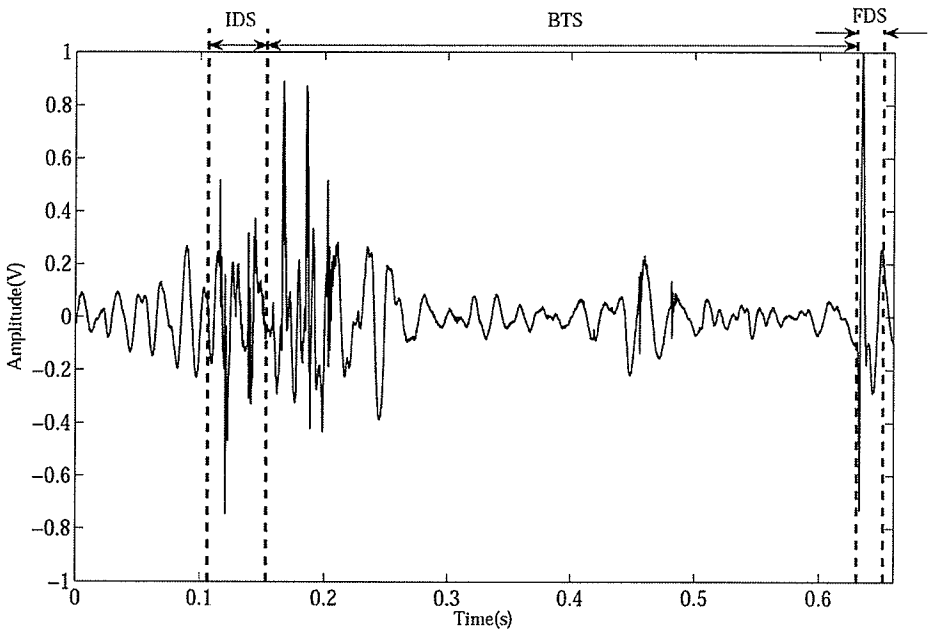


Fig. 2.2. Three segments of a typical swallowing sound signal.

When recording swallowing sounds, the recorded signal will also contain breath sounds. Respiration halts during the pharyngeal phase of the swallow. The pattern of occurrence of the swallow within a breath cycle (i.e., inspiration-swallow-expiration-inspiration or inspiration-expiration-swallow-inspiration) is also of interest, as it can be an indicator of the risk of dysphagia [Rempel & Moussavi, 2005]. In healthy adults, swallows are generally followed by expiration [Selley et al., 1990]. In one study, it was found that 93% of the swallows preceded by expiration and 100% of the respiratory phases following swallow were expiratory [Klahn et al, 1999]. The breath-swallow patterns of healthy individuals and dysphagic patients have been found to be different and are functions of the bolus viscosity. For instance, it was shown that the rate of post-swallow inspiration in young people with cerebral palsy is higher than that of normal

individuals when they were drinking thin liquids; however, this was not the case when they were consuming thicker substances, i.e., milk shake or pudding [Rempel and Moussavi, 2005].

Swallowing and breath sounds are substantially different and have to be analyzed separately. The duration of breath sounds is longer than the duration of IDS and FDS. Breath sounds are relatively stationary (especially when the air flow plateaus), as opposed to swallowing sounds which consist of discrete nonstationary click sounds. A typical waveform and spectrogram for a sequence of swallowing and breath sounds, recorded by an accelerometer placed over the suprasternal notch, are shown in Fig. 2.3. As the swallowing sound signal is basically a nonstationary signal its frequency content varies in different areas. As can be seen in Fig. 2.3, the spectrum of a swallowing sound signal is in the frequency range below 2500 Hz.

In recent years, the use of signal processing techniques in swallowing sound analysis has drawn considerable attention. Most studies are based on the time domain characterization or Fourier analysis (time-frequency domain) of the signal, e.g., duration pattern of swallowing sound signals as a function of the bolus volume and texture has been studied [Boiron *et al.,* 1997]. Durational pattern and frequency contents of swallowing sounds following total laryngectomy has been investigated [Hamlet *et al.,* 1992].

Since swallowing sound signals are generated by a complicated nonlinear mechanism, i.e., the swallowing mechanism, analysis of swallowing sound signals based on the nonlinear dynamics may lead to the improvement in clinical assessment techniques of swallowing disorders using their acoustic signatures. Although most biological

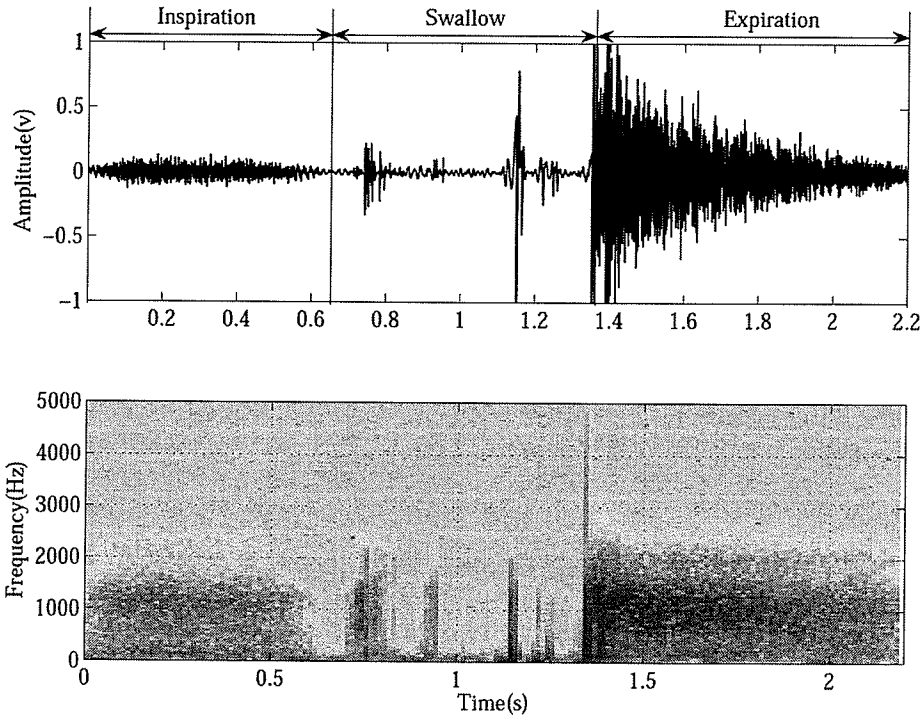


Fig. 2.3. A typical sequence of swallowing and breath sounds (top) and the corresponding spectrogram (bottom).

systems are nonlinear, due to the complexity of nonlinear analysis, biological systems have often been approximated by linear models. Recent developments in the theory of nonlinear dynamics have provided new theoretical and conceptual tools for a better understanding of the complex behavior of real-world systems and signals, offering a better option for analysis of the signals originating from very complicated nonlinear biological systems such as swallowing mechanism. This study is the first attempt at applying nonlinear dynamic parameters to swallowing sound signals.

Swallowing sounds are distinguishable from breath sounds by auditory means; however, finding a feature with high objective discriminating ability between swallowing

and breath sounds is still a challenge to meet; hence, this study investigated several nonlinear dynamic parameters. Recurrence quantification analysis (RQA) is a nonlinear analysis method that develops the parameters with discriminating capability for the detection of the swallowing sounds.

More recently, it has been shown that swallowing sound analysis is a promising diagnostic tool for the assessment of swallowing disorders [Lazareck and Moussavi, 2004]. Using features that have the ability to discriminate between the swallows of normal and dysphagic subjects can lead to an enhancement of the performance of automated classification. In this study, the discriminating ability of nonlinear dynamic parameters was confirmed. In addition, several features were used for the HMM based classification of swallowing sounds and the performances of the features were compared.

The temporal and durational patterns associating respiration and swallowing [Klahn et al., 1999], and timing of velopharyngeal closure [Smith et al., 1990] have also been studied based on swallowing sound analysis. This thesis reports the outcomes of the study on the swallowing mechanism timing based on the analysis of submental surface EMG and swallowing sounds. The durational and temporal pattern of the submental muscle activities were analyzed using the EMG signal.

CHAPTER 3

BACKGROUND THEORY

This chapter provides the theoretical background of the methods used in this study. First, nonlinear dynamics theory will be discussed. A brief introduction to hidden Markov modeling will then be provided.

3.1 Nonlinear Dynamics

Most biological systems are complex processes involving numerous elements; however, their responses to the internal or external stimuli cannot be expressed as a sum of the responses of their elements, i.e., the behavior of biological systems is not subject to the principle of superposition. Accordingly, most biological systems are nonlinear; however, due to the complexity of nonlinear analysis, biological systems have often been approximated by linear models. Recent developments in the theory of nonlinear dynamics have provided new theoretical and conceptual tools for a better understanding of the complex behavior of real-world systems and signals, offering a better option for analysis of the signals originating from very complicated nonlinear biological systems such as swallowing sounds. However, the parameters describing nonlinear processes are complicated and may not have a clear physical interpretation.

One of the goals of this study was to investigate the characteristics of swallowing sounds using nonlinear dynamics theory. The methods employed in this study are based on the analysis of swallowing sounds, considered as a single variable time series

generated by the swallowing process. This observable output of the system was used to reconstruct the multidimensional state space representation of the process to investigate the complexity of the process and the presence of chaos. Furthermore, nonlinear dynamic parameters were used for automatic detection of swallowing sounds in the swallow and breath sound recordings as well as classification of swallowing sounds of healthy and dysphagic subjects.

3.1.1 Nonlinear Dynamical Systems

Any process or set of processes evolving in time, in which the evolution is governed by some set of physical laws, is called a dynamical system. The term is also used to refer to mathematical models that evolve in time. Dynamical systems can range from relatively simple systems, i.e., systems of a few variables governed by a few equations, such as mass-spring systems, to extremely complex systems such as the complete climate system. A dynamical system has a state determined by a collection of numbers that are the coordinates of a geometrical space called state space or phase space. If the evolution rule is fixed in time, the system is called autonomous. A continuous-time autonomous dynamical system of d degrees of freedom may be described by the state equations [Parker and Chua, 1987]:

$$\frac{dX(t)}{dt} = G(X(t)), \tag{3.1}$$

where $X(t) = [x_1(t), x_2(t), ..., x_d(t)] \in R^d$ is the state of the system at time t. $G: R^d \rightarrow R^d$ is called the vector field and is the time independent evolution rule. Equation (3.1) is a set of d ordinary differential equations. The solution to (3.1) with

initial condition $X(t_0) = X_0$ is called a trajectory and is denoted by $\phi_t(X_0)$. The trajectory is the path followed by the system as it evolves with time. If G is linear, the system is a linear continuous-time autonomous dynamical system of d degrees of freedom and may be described by the state equations:

$$\frac{dX(t)}{dt} = AX(t), \tag{3.2}$$

where A is a constant time-independent $d \times d$ matrix. Equation (3.2) is a set of d ordinary linear differential equations. Stability of the system depends on the real part of the eigenvalues of matrix A.

In the discrete-time domain, each state is described at discrete time k:

$$X(k) = X(t_0 + kT_S), \tag{3.3}$$

where T_S is the sampling period. In a discrete-time autonomous dynamical system, the evolution is given by a map $F: R^d \rightarrow R^d$, in which

$$X(k+1) = F(X(k)). \tag{3.4}$$

Continuous-time and discrete-time views of the dynamics evolution can be related by derivative approximation [Abarbanel *et al.*, 1993]:

$$\frac{dX(t)}{dt} \approx \frac{X(t_0 + (k+1)T_S) - X(t_0 + kT_S)}{T_S}, \tag{3.5}$$

or:

$$F(X(k)) = X(k) + T_S G(X(k)). \tag{3.6}$$

In the theory of nonlinear dynamics, the existence, uniqueness and behavior of the solutions to dynamical systems when defined by Equations (3.1) or (3.4) are discussed. These discussions can be found in any text on nonlinear dynamics (e.g. [Strogatz, 2000]).

Behavior of nonlinear dynamical systems differs from that of linear dynamical systems. Specially, chaotic behavior of nonlinear systems does not occur in linear systems. Chaos occurs when trajectories are bounded, but they do not converge to an equilibrium point or to a periodic or quasiperiodic trajectory. Chaotic behavior is not similar to unstable behavior because in the unstable behavior the trajectories are not attracted by a fixed point or limit cycle and may be unbounded. Chaotic behavior differs from stable behavior, in which after some transient period the system settles on a bounded steady state trajectory. An important feature of chaotic systems is sensitivity to the initial conditions, i.e., in a chaotic system, the smallest change in the initial conditions may cause the trajectory to quickly follow a completely different path.

The Lorenz model is an example of nonlinear systems, which is a simplification of the fundamental Navier-Stokes equations for convecting fluids. It is an autonomous dynamical system of three degrees of freedom and described by the following set of equations:

$$\begin{aligned} \frac{dx_1}{dt} &= x_2 x_3 - b x_1 \\ \frac{dx_2}{dt} &= p(x_3 - x_2) \\ \frac{dx_3}{dt} &= -x_1 x_2 + r x_2 - x_3 \end{aligned} \qquad (3.7)$$

In (3.7), $X(t) = [x_1(t), x_2(t), x_3(t)] \in R^3$ is the state of the system at time t, and b, p, and r are three adjustable parameters of the model. Although the Lorenz model equations look very simple, the model exhibits very complex behavior. The behavior of the system depends on the adjustable parameters. With some values of the model parameters, the systems show a periodic behavior, whereas with other values of the parameters, the

behavior is chaotic, etc. The detailed discussion on the behavior of the model for different values of the model parameters can be found in [Hilborn, 2000]. As an example, three state variables x_1, x_2, and x_3 as functions of time for the Lorenz model with $b=3$, $p=10$, $r=28$ and the initial conditions

$$X(0)=\begin{bmatrix} x_1(0) \\ x_2(0) \\ x_3(0) \end{bmatrix}=\begin{bmatrix} 35 \\ -10 \\ -7 \end{bmatrix} \tag{3.8}$$

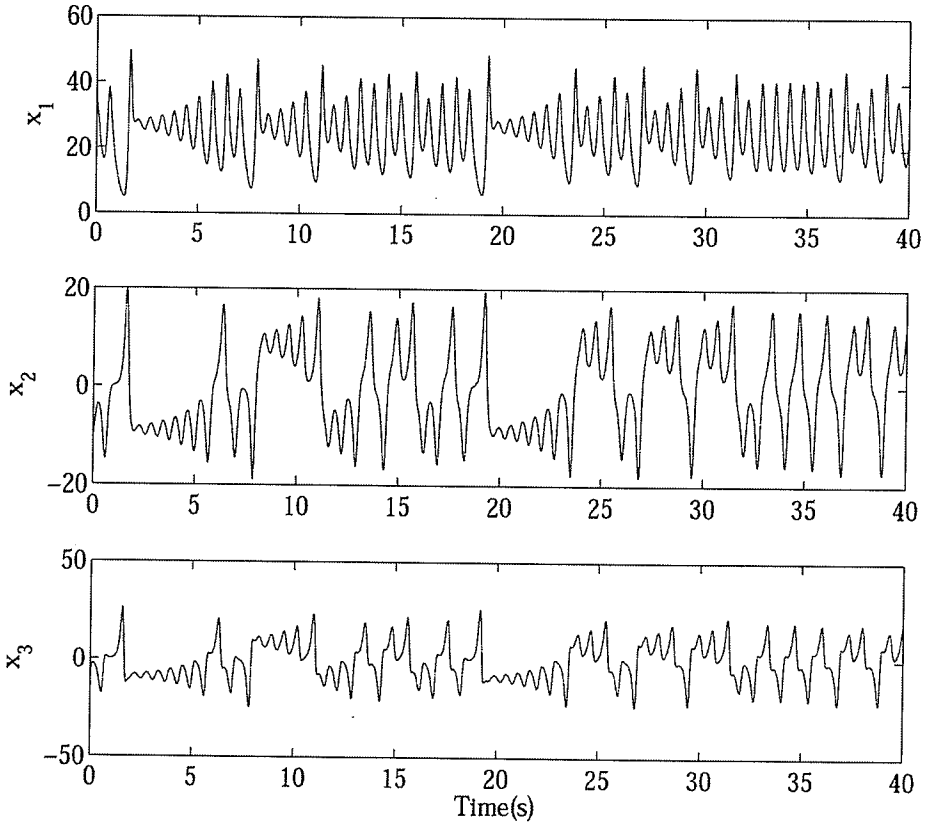


Fig. 3.1. Three state variables x_1, x_2, and x_3 as functions of time for the Lorenz model.

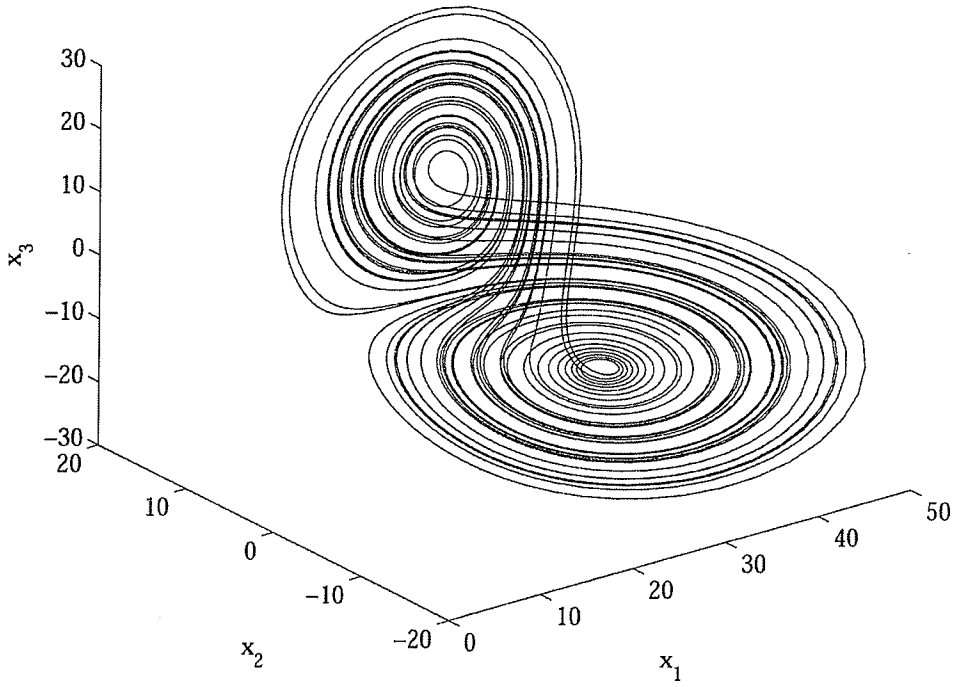


Fig. 3.2. Trajectory for the Lorenz model.

are plotted in Fig 3.1. The Lorenz model with these values of the adjustable parameters shows a chaotic behavior. Although the signals appear to be random, they are the state variables of a deterministic system. Figure 3.2 shows the trajectory of the Lorenz model in a three dimensional state space with the same parameters and initial conditions of Fig. 3.1. As it can be seen in Fig. 3.2, the trajectory does not converge to a periodic trajectory. In a chaotic system, trajectories get close enough to the so called attractor and remain close to the attractor. Attractors of chaotic systems are not simple geometrical objects and have a fractional dimension.

An interesting aspect of nonlinear dynamics is the analysis of signals generated by

nonlinear systems. In this case, the evolution equations of the nonlinear systems are not known and available information is limited to time series data representing signals generated by nonlinear dynamical systems. Although these signals may appear to be random signals, these observable outputs of the system might be used to describe the deterministic structure of the system, i.e., assuming that the system is fundamentally deterministic, the state space representation of the system may be reconstructed based on the information in the time series.

The most common method for extraction of the information in the time series is based on time delay coordinates. The idea of using time delay coordinates to reconstruct the trajectory in the state space was proposed in [Packard, 1980]. Later, Takens proved a theorem that is the firm basis of the method of delays [Takens, 1981]. The basic idea is that we can use the delayed variables to form a coordinate system to reconstruct the structure of trajectories in the state space. Available scalar data must be transformed into multidimensional phase space plots. Available data can be described by the one dimensional time series $x(i)$, where $i = 1,2,...,N$ is the time index. The method of delays reconstructs the trajectory by using delay coordinates to form state space vectors $X(k)$ in a multidimensional phase space. The E-dimensional vectors $X(k)$ can be constructed as:

$$X(k) = [x(k), x(k+\tau), x(k+2\tau),..., x(k+(E-1)\tau)], \tag{3.9}$$

where τ is an appropriate time delay (an integer multiple of the sampling period), $X(k)$ is one point of the trajectory in the phase space at time $k = 1,2,..., N-(E-1)\tau$, and E is the so-called embedding dimension. According to Takens theorem [Takens, 1981], regardless of the time delay one chooses, the minimum embedding dimension E_{min} depends on the dimension of the system d. If the embedding dimension is

larger than $2d$, embedding will result in a state space reconstruction that contains a smooth manifold for the system [Banbrook *et al.*, 1999].

3.1.2 Choosing the Time Delay

Although Takens theorem does not say anything about a proper time delay parameter τ, the quality of the reconstruction using the method of delays depends on τ [Rosenstein *et al.*, 1994]. A value of τ that is too small provides little information gain between successive delay coordinates, and the reconstructed trajectory may become very close to the identity line of the embedding space. On the other hand, if τ is too large, successive delay coordinates may become casually unrelated, and the trajectory will wander all over the embedding space. The effects of a change in time delay parameter are more significant when we work with finite length experimental time series and/or noisy data.

Figure 3.3 shows an example of the effect of time delay on the quality of reconstruction. The two dimensional trajectory of a time series representing one of the state variables of the Lorenz attractor was reconstructed with three values of time delay, i.e., $\tau = 1, 10,$ and 100. With $\tau = 1$, the reconstructed trajectory is very close to the identity line of the embedding space; with $\tau = 100$, the trajectory wanders all over the embedding space; however, with $\tau = 10$ a proper state space representation of the signal has been reconstructed.

A commonly used method for selecting the time delay is based on the autocorrelation function. Autocorrelation is a measure of the similarity between a signal and its delayed version; it is maximum when the delay is zero. The time delay, τ, is chosen as the time

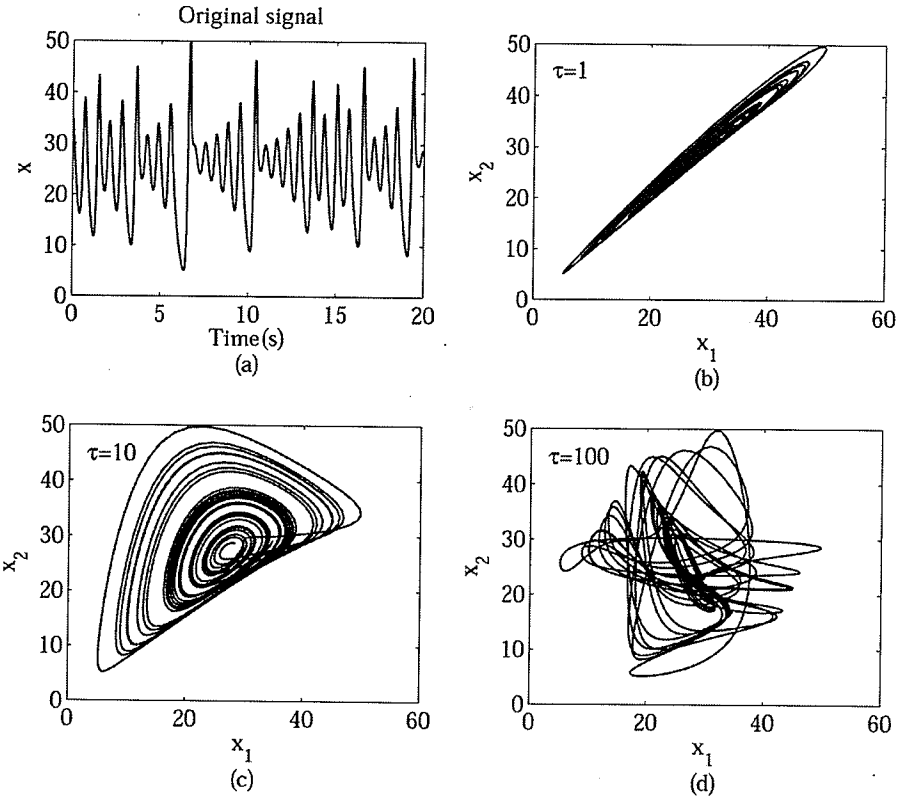


Fig. 3.3. Reconstruction of 2-dimensional trajectory of a signal with different values of time delay

when the autocorrelation function first drops to nearly zero or a certain fraction, e.g. 0.5, of its maximum. Using the fast Fourier transform (FFT)-based algorithm, this method of time delay selection is efficient in terms of computation time; however, according to [Martinerie et al. 1992] and [Fraser et al., 1986] the autocorrelation-based method produces poor results for many dynamical systems including Lorenz and Rossler models.

Another method for time delay selection is based on the mutual information function [Fraser et al., 1986], which selects the delay that produces a local minimum in the mutual information function. The method is computationally complicated. Furthermore, as

suggested by [Martinerie et al. 1992], the mutual information function-based method is inconsistent in identifying the optimal value of the time delay.

In this study, the method proposed in [Rosenstein *et al.*, 1994] was used which is a geometry based method for choosing a proper time delay. This method measures the average displacement, $< S_E(\tau) >$, of the embedding vectors from their original locations on the identity line (when $\tau = 0$), and it is evaluated as a function of τ such that:

$$< S_E(\tau) >= \frac{1}{M} \sum_{k=1}^{M} \| X^{\tau}(k) - X^{0}(k) \|, \tag{3.10}$$

where the superscripts denote the time delay between successive embedding components, and M is the number of vectors for the corresponding dimension E. As the time delay increases, the average displacement increases accordingly and reaches a plateau gradually. With larger values of E, reconstruction expansion reaches a plateau at smaller values of τ. The best time delay is chosen as the point where the slope of the average displacement curve decreased to less than 40% of its initial value [Rosenstein *et al.*, 1994].

3.1.3 Choosing the Minimum Embedding Dimension

The purpose of time delay embedding is to unfold the trajectories in an Euclidean space large enough such that all self–crossings of the orbit can be eliminated. The trajectories will be unfolded if we use the minimum embedding dimension E_{min}, or any $E > E_{min}$.

Finding a proper embedding dimension is an important problem, because in a space

with an embedding dimension less than the E_{min}, some neighbor points may be close to each other due to the projection from some higher dimension down to this lower dimension. On the other hand, calculation in any dimension larger than the E_{min} leads to excessive computation during the evaluation of parameters such as correlation dimension and Lyapunov exponents.

As shown in [Grassberger *et al.*, 1999], dimensionality of various parts of nonstationary signals may be different from each other. On the other hand, as indicated in [Ding *et al.*, 1999], noise will tend to increase the required dimension. Hence, it is important to choose a sufficiently large embedding dimension that contains the relevant dynamics in the presence of noise.

In this study, the method proposed in [Kennel *et al.*, 1992] was used to find a proper value for E_{min}. This method is based on increasing the dimension, i.e., $E = 1,2,3,...$ until no false neighbors remain. In dimension E, each vector, $X(k)$, has a nearest neighbor, $X^{NN}(k)$. Let $R^2{}_E(k)$ be the distance between the vectors $X(k)$ and $X^{NN}(k)$:

$$R^2{}_E(k) = \| X(k) - X^{NN}(k) \| = \sum_{j=0}^{E-1} [x(k+j\tau) - x^{NN}(k+j\tau)]^2 . \qquad (3.11)$$

In dimension $E+1$, the distance between the vectors $X(k)$ and $X^{NN}(k)$ can then be described as

$$R^2{}_{E+1}(k) = R^2{}_E(k) + [x(k+E\tau) - x^{NN}(k+E\tau)]^2 . \qquad (3.12)$$

If $R^2{}_{E+1}(k) >> R^2{}_E(k)$, the closeness $X(k)$ and $X^{NN}(k)$ is due to the projection from some higher dimensional attractor down to dimension E. When the dimension is increased to $E+1$, the distance between these two points will noticeably increase. A

threshold T is required to decide which neighbors are false. The neighbors that fulfill the condition

$$\frac{|x(k+E\tau)-x^{NN}(k+E\tau)|}{R_E(k)}>T, \tag{3.13}$$

at time instant k, are considered as false neighbors.

The criterion stated in (3.13) is not sufficient for proper embedding. When (3.13) is applied to some signals generated by very high dimensional systems, it indicates that the signal can be embedded in a relatively small dimension. This problem happens when we have a finite length experimental time series and can be solved by increasing the length of the experimental time series [Abarbanel *et al.*, 1993]. However, the length of a time series may not be increased arbitrarily in many cases due to hardware limitations. This problem happens because in the embedding of finite length time series, the nearest neighbors are not necessarily close to each other. These neighbors do not move apart when the dimension is increased. Therefore, [Kennel *et al.*, 1992] suggested to consider those nearest neighbors that are not close to each other as false neighbors, i.e., if

$$\frac{R_{E+1}(k)}{R_A}\geq 2, \tag{3.14}$$

the nearest neighbor is false. In (3.14) R_A is defined as

$$R_A=\frac{1}{N}\sum_{k=1}^{N}[X(k)-\overline{X(k)}]^2, \tag{3.15}$$

where $\overline{X(k)}$ is the mean of state space points and defined by

$$\overline{X(k)}=\frac{1}{N}\sum_{k=1}^{N}X(k). \tag{3.16}$$

3.1.4 Correlation Dimension

The correlation dimension is a parameter for evaluation of the complexity of a process, and it is used to measure the dimensionality of the space occupied by the trajectory. In this study, the correlation dimension was calculated using the Grassberger-Procaccia algorithm [Grassberger & Procaccia, 1983], which is based on calculation of the relative number of pairs of points in the phase space that are separated by a distance of less than or equal to r. To compute the correlation dimension, the correlation sum, $C(r)$, is calculated as

$$C(r)=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Theta(r-|X(i)-X(j)|), \tag{3.17}$$

where $X(i)$, $X(j)$ are the points of the trajectory, N is the total number of points of the trajectory, the distance r is a radius around each point $X(i)$, and Θ is the Heavyside function, which excludes values outside the radius r. The slope of the linear region of the plot of $\log(C(r))$ versus $\log(r)$ is considered as the correlation dimension. This evaluation method of the dimensionality may be qualitatively understood by noting the fact that there will be more ways for points to be close to each other for higher dimensional objects; therefore, the number of pairs close to each other will rise more rapidly for higher dimensions.

3.1.5 Lyapunov Exponents

Lyapunov exponents are parameters commonly used to detect the presence of chaos

in dynamical systems. The Lyapunov exponents are measures of the exponential rate of divergence or convergence of nearby state space trajectories. A positive Lyapunov exponent implies divergence of nearby trajectories, whereas a negative one implies convergence of nearby trajectories. In a system with positive Lyapunov exponents, with the smallest change in initial conditions, the trajectory quickly follows a completely different path. If a system is known to be deterministic, the positive largest Lyapunov exponent can be taken as the proof of the sensitivity to initial condition and presence of chaos. If both positive and negative Lyapunov exponents exist, two different initial conditions will diverge exponentially at a rate given by the largest (dominant) Lyapunov exponent [Hilborn, 2000].

Different methods have been proposed to estimate the largest Lyapunov exponent (e.g. [Wolf *et al.,* 1985; Rosenstein *et al.*, 1993]). In this study, the largest Lyapunov exponent was approximated by the estimation procedure proposed in [Rosenstein *et al.*, 1993] for noisy and short time series. After reconstruction of the dynamics, for a point $X(k)$, the nearest Euclidean neighbor $X^{NN}(k)$ is searched for in the phase space. Let's call their distance as $d_k(0)$:

$$d_k(0) = \min_{X^{NN}(k)} \| X(k) - X^{NN}(k) \| . \tag{3.18}$$

This k^{th} pair of the nearest neighbors diverges approximately at a rate given by the largest Lyapunov exponent. The distance $d_k(i)$ between the developed points (after passing i samples) is:

$$d_k(i) \approx d_k(0) \exp(\lambda_1 i T_S) , \tag{3.19}$$

where T_S is the sampling period of the time series and λ_1 is the largest Lyapunov

exponent. Taking the logarithm of both sides of Equation (3.19), we obtain:

$$\ln(d_k(i)) \approx \ln(d_k(0)) + \lambda_1 i T_S . \tag{3.20}$$

Equation (3.20) represents a set of approximately parallel lines, each with a slope roughly proportional to λ_1. In this study, the largest Lyapunov exponent was calculated as the average slope of these parallel lines.

3.1.6 Recurrence Quantification Analysis

Recurrence plots are useful tools for visualization of the recurrence of the states in the state space. Usually, state spaces with dimensions higher than two or three can only be visualized by projection into two or three dimensional sub-spaces. Recurrence plots enable us to investigate higher dimensional state space trajectories through a two-dimensional representation of their recurrences. Recurrence plots were introduced in [Eckmann *et al.*, 1987] as graphical tools for measuring the time consistency of dynamical systems. The effects of a drift in the system parameters were addressed and it was suggested that the length of line segments parallel to the identity line of the recurrence plots were inversely related to the largest positive Lyapunov exponent.

RQA was used in this study to investigate the key structures of swallowing sounds in comparison with breath sounds. As the first step, multidimensional state space representation of the tracheal sound signal was reconstructed based on the Taken method of delays [Taken, 1981]. Let $X(k)$ be one point on the trajectory in the E-dimensional state space at different time instances, $k = 1,2,...,M$, where M is the number of points on the trajectory. Then, the recurrence matrix R is an $M \times M$ square matrix with elements defined as:

$$R_{ij} = \Theta(r - | X(i) - X(j) |), \tag{3.21}$$

where $X(i)$, $X(j)$ are the points of the reconstructed state space plot, r is a radius around each point $X(i)$, and Θ is the Heavyside function which excludes values outside the radius r. In other words, R_{ij} is set to one when the distance between $X(i)$ and $X(j)$ is less than r, otherwise it is set to 0. The radius r should be chosen such that a reasonable number of points of the trajectory are in the sphere of radius r centered at $X(i)$. A value of r that is too small results in quantification of the noise only, whereas a value of r that is too large considers far points as recurrent points. In this study, the radius was chosen to be the same for all points on the recurrence matrix; hence, the recurrence matrix was symmetric with respect to its identity line $i = j$. In order to determine the radius r, the criterion proposed in [Zbilut e*t al.*, 2002a] was adopted, in which the radius was initially set to a very small value. Then, the percentage of recurrence points (%REC) was computed. If in any given window %REC was less than 1%, the radius should be increased.

It is common to change a recurrence matrix to a digital image with $M \times M$ resolution, in which each value of one is converted to a black dot. The obtained image is called a recurrence plot. Therefore, a dot in a recurrence plot indicates that two points of the trajectory pass through the same neighborhood in the state space. As i and j are time coordinates, a recurrence plot implicitly describes time correlation information. When a series of consecutive points of the trajectory are located in the vicinity of another series of consecutive points, a straight line parallel to the main diagonal line (identity line) occurs in the recurrence plot. The maximum length of the lines depends on the largest

Lyapunov exponents [Eckmann *et al.*, 1987].

Various structures of the recurrence plots can be detected and interpreted. Since detection of subtle patterns in the recurrence plots by qualitative visual inspection is difficult and subjective, recurrence quantification parameters [Webber and Zbilut, 1994] have been proposed for better understanding of recurrence plots. The proposed parameters are based on detection and measurement of the diagonal structures distribution, including the percentage of recurrence points (%REC); the percentage of recurrence points forming the lines parallel to the identity line (%DET); the ratio of %DET to %REC; Shannon entropy of distribution of the diagonal lines; and the trend that measures the paling of the recurrence plots away from the identity line. Several parameters based on vertical or horizontal structures have also been introduced [Marwan *et al.*, 2002]; which are laminarity (the percentage of the recurrence points forming vertical lines); trapping time (the average length of the vertical lines); and the length of the longest vertical line.

Recurrence plots may be used to extract features describing system behavior. Quantification of recurrence plots was used for estimating the proper embedding dimension and time delays of scalar time series [Zbilut and Webber, 1992]. Structures of recurrence plots, detection of nonstationarity and state transitions of time series were discussed in [Gao and Cai, 2000] and [Gao, 2001]. Also, utilization of features extracted from recurrence plots in the calculation of Lyapunov exponents of nonstationary signals has been studied in [Zbilut *et al.*, 2002a].

3.2 Hidden Markov Models

Signal modeling has been a tool of choice for a wide range of applications. Signal modeling may be categorized into two classes: deterministic and statistical. Deterministic modeling refers to characterization of the signal with known mathematical equations or physical components. Meanwhile, statistical modeling uses stochastic processes to characterize the events that are too complex to be completely characterized with a few equations or components.

A hidden Markov model (HMM) is a statistical model that characterizes an observation sequence and can infer the state sequence of a hidden random process from values of the observation sequence. Benefits of using an HMM encompass its effectiveness in the characterization of signals and systems and its ability to model unknown signal dynamics statistically. The primary application of HMMs has been in speech processing, such as speech modeling, recognition, and classification of normal and pathological speech [Rabiner, 1989; Rabiner *et al.*, 1993]. In a similar manner, this thesis addresses the application of HMMs in the automated detection of the swallowing sounds in the tracheal swallow and breath sound recordings, the segmentation of swallowing sound signals, and the classification of swallowing sounds of healthy subjects and dysphagic patients.

3.2.1 Theory of Hidden Markov Modeling

A Markov model describes a system that can be completely characterized by a finite number of distinct states. The system may be in one state of a set of N states $S = \{s_1, s_2, ..., s_N\}$. At time $t = 1,2,3,...,T$ the state changes from q_{t-1} to q_t, where

q_t denotes the actual state of the system at time t. In a time-independent first order Markov model, it is assumed that the next state is dependent only upon the current state [Gold and Morgan, 1999]. Transition between the states is governed by the state transition probabilities described by

$$a_{ij} = P[q_t = s_j \mid q_{t-1} = s_i], \qquad 1 \le i, j \le N \tag{3.22}$$

where a_{ij} is an element of the state transition matrix A. The state transition coefficients a_{ij} are non-negative and

$$\sum_{j=1}^{N} a_{ij} = 1. \tag{3.23}$$

The initial state of the system may be one state of N states $S = \{s_1, s_2, ..., s_N\}$. The initial state distribution is denoted by $\Pi = \{\pi_i\}$ where π_i is the probability of the initial state q_1 being in state i

$$\pi_i = P[q_1 = s_i], \qquad 1 \le i \le N. \tag{3.24}$$

An HMM is a Markov model whose states are not observable directly. Instead, the output of the process, also called a symbol, is observable and the observation sequence is generated by the system. Therefore, the states are *hidden* to the outside; hence the name hidden Markov model [Rabiner *et al.*, 1993].

In an HMM, the observations are probabilistic functions of the states. The observations may be chosen from a set of continuous symbols. Otherwise, they may be characterized as discrete symbols chosen from a set of finite symbols, i.e., observations that belong to a set of finite distinct symbols. In the continuous case, a continuous probability distribution function (pdf) describes the probability of each observation in each state. The most

general representation of the pdf is a finite mixture of elliptically symmetric functions (e.g., Gaussian). In this case, the mean and standard deviation of these functions are found in the training phase of the HMM.

A discrete HMM is characterized by the following elements:

1- N, the number of hidden states in the model. N represents the total number of distinct events that sufficiently characterizes the underlying dynamics of the signal. Each hidden state does not necessarily correspond to a physical state.

2- M, the number of distinct observation symbols per state. We denote the individual symbols as $V = \{v_1, v_2, ..., v_M\}$. M represents the size of the symbol set that models the physical output of the system.

3- The state transition matrix $A = \{a_{ij}\}$, where a_{ij} is defined in (3.22).

4- The observation symbol probability distribution in state j, $B = \{b_j(k)\}$, where

$$b_j(k) = P[o_t = v_k \mid q_t = s_j], \quad 1 \le j \le N, \quad 1 \le k \le M. \tag{3.25}$$

In (3.30) o_t denotes the observation at time t. The observation sequence $O = o_1 o_2 ... o_T$ is generated by the system.

5- The initial state distribution $\Pi = \{\pi_i\}$, where π_{ij} is defined in (3.24).

Three probability parameters A, B, and Π are indicated by the compact notation

$$\lambda = \{A, B, \Pi\}. \tag{3.26}$$

In this study, discrete HMMs were used. Hence, there was no need to make any assumption on the distribution of the observations and a non-parametric estimation was used for the computation of the observation symbol probability distribution B.

Three basic problems, which must be solved for the model to be useful for different

applications, are:

1- Evaluation of the probability of an observation sequence given a specific model, i.e., given the observation sequence $O = o_1 o_2 ... o_T$ and the model parameters, how do we calculate the probability of the observation sequence, $P(O \mid \lambda)$? (Classification problem)

2- Determination of the best sequence of the model states to which observations belong, i.e., given the observation sequence and the model parameters, how do we choose a corresponding state sequence $q_1 q_2 ... q_T$ which is optimal in some meaningful sense?

3- Refinement of the model parameters that will result in the maximum likelihood of the observation sequence, i.e., how do we adjust the model parameters $\lambda = \{A, B, \Pi\}$ to maximize $P(O \mid \lambda)$? (Training problem)

Different solutions to these problems can be found in any text on HMM (e.g., [Rabiner, 1989]).

CHAPTER 4

METHODS

4.1 Experimental Data

Data used in this study were in two groups: data of 12 healthy adults recorded for this study; and data of healthy children and cerebral palsy young adults adopted from a previous study [Rempel & Moussavi, 2005]. 12 subjects, who participated in this study, were adults (ages 21-35, 9 males) and in good health without any history of respiratory or swallowing disorder, eating or nutrition problem, lower respiratory tract infection, or neurological impairment. Three recordings were conducted. In the first and second recordings, the subjects were fed pre-packed pudding (semisolid texture), and orange juice (thin liquid) with a bolus size of 5 ml, respectively. In the third recording, the subjects drank orange juice by a straw. In each recording trial, 5-8 swallows were recorded. Swallowing sounds along with tracheal breath sounds were recorded by a Siemens accelerometer (EMT25C) placed over the suprasternal notch by double-sided adhesive tape rings.

In order to find the respiration phases, i.e., inspiration/expiration, another accelerometer was placed on the left or right second intracostal space (whereever the difference between inspiration and expiration was higher) in the midclavicular line to record lung sounds. This lung sound was used only for respiratory phase information using the method introduced in [Moussavi *et al.*, 2000] for manual confirmation and validation of swallowing localization in a breath cycle. Simultaneously, the EMG signal

was recorded with one bipolar channel using Ag-AgCl disposable electrodes placed over the midsagittal line of the submental area. The sounds and EMG were amplified and bandpass filtered in the range of 50-2500 Hz and 15-500 Hz, respectively. The signals were simultaneously digitized at a 10240 Hz sampling rate. The gain of the amplifier was adjusted such that the sound level of the signals was approximately the same for all subjects. Additionally, each recorded signal was normalized to the average power of the signal.

As mentioned earlier, in addition to the data acquired for this study, data were also adopted from a previous study data set [Rempel & Moussavi, 2005], in which tracheal sound recordings of two groups of subjects were collected. The first group consisted of 12 healthy children (ages 3-16 years) and three healthy adults (ages 35, 38, and 54 years). The second group consisted of 11 young adult patients (ages 16-25 years) with swallowing disorder. During the test, participants were fed three textures: pre-packaged pudding (semisolid texture), diluted pudding (thick liquid texture) and fruit juice (thin liquid) in a bolus size of 5 ml throughout the experiment. The method of accelerometer placement was the same as that of this study. The sound signals were amplified, bandpass filtered (30-2500 Hz) and digitized at 10240 Hz. It should be noted that we excluded the swallows of the second group, in which aspiration had occurred. However, since other swallows of the second group belonged to dysphagic patients, we considered them as marginally normal and at risk of dysphagia. In addition, VFS images of two subjects of the first group and three subjects of the second group who had gone through a VFS swallowing study were also available, in which the participants were fed the same three textures mixed with barium. As the subjects were fed barium-mixed boluses, their bolus

movement patterns on the videofluoroscopy screen were recorded by a VCR simultaneously with the audio signals using the same signal conditioning as the rest of the data used in this study. The sound and images were later digitized and transferred to a computer. VFS allows visual assessment of the swallowing mechanism; hence, we used those data for validation of the proposed method's capability in detecting the boundaries of different phases of the swallowing sound signal.

In both data acquisition processes, informed consent was obtained from all subjects or their parents/guardians prior to the study. The study was approved by the Ethics Committee for the Use of Human Subjects in Research of the University of Manitoba.

4.2 Swallowing Sound Detection

Since the recorded sound signals contain both swallowing and breath sounds, the swallowing sounds must be extracted from the recorded signals. In this study, three techniques were used for detection of swallowing sounds in the records of tracheal breath and swallowing sounds. These three methods are described in this section.

4.2.1 Wavelet Transform Based Swallowing Sound Detection

As the first technique, a wavelet transform based filter was used for swallowing sound detection. This filter has been used successfully for extraction of discontinuous adventitious sounds from vesicular sounds [Hadjileontiadis and Panas, 1997], suppressing noise in partial discharge on-line monitoring [Mingyou *et al.*, 2000], and enhancement of bowel sounds [Hadjileontiadis *et al.*, 2000].

The method takes advantage of the fact that swallowing sounds consist of discrete nonstationary click sounds whereas breath sounds are relatively stationary especially when the air flow plateaus. Therefore, swallowing sounds have large components in many wavelet scales whereas wavelet coefficients of breath sounds in higher wavelet scales are small. In the proposed method, swallowing sounds were separated from breath sounds through a multiresolution decomposition-reconstruction process.

Multiresolution analysis of tracheal sound is the basis of the wavelet transform based filter. The signal is decomposed into band limited components, which can be reassembled to reconstruct the original signal without error. Since the bandwidth of the resulting subbands is smaller than that of the original signal, the subbands can be downsampled without any loss of information. Reconstruction of the original signal is accomplished by upsampling, filtering and summing the individual subbands. The decomposition process can be iterated such that one signal is broken into many lower resolution components. A complete description of multiresolution analysis can be found in [Mallat, 1998].

In the filtering method used in this study, iterative sequences of multiresolution decomposition (MRD) and multiresolution reconstruction (MRR) were employed. MRD of a typical tracheal sound signal in six levels using the Daubechies-5 wavelet is shown in Fig. 4.1. As the swallowing sound signal includes discrete nonstationary click sounds, it has large components in different wavelet scales whereas the breath sound is relatively stationary and its wavelet coefficients die out in higher wavelet scales (Fig. 4.1).

A schematic description of the filter is shown in Fig. 4.2. In this study, swallowing sounds are detected in a normalized N-sample tracheal sound signal denoted by $X(n)$, $n = 1,2,..,N$. During the k^{th} iteration, the wavelet coefficients at levels $j = 1,..,6$ are

calculated. The wavelet coefficients at level j are compared to a threshold determined by the Otsu method [Otsu, 1979]; the wavelet coefficients at each scale are divided into big ($\geq T_j^k$) and small ($< T_j^k$) coefficients, WT_S^k and WT_B^k, respectively. Then, MRR is

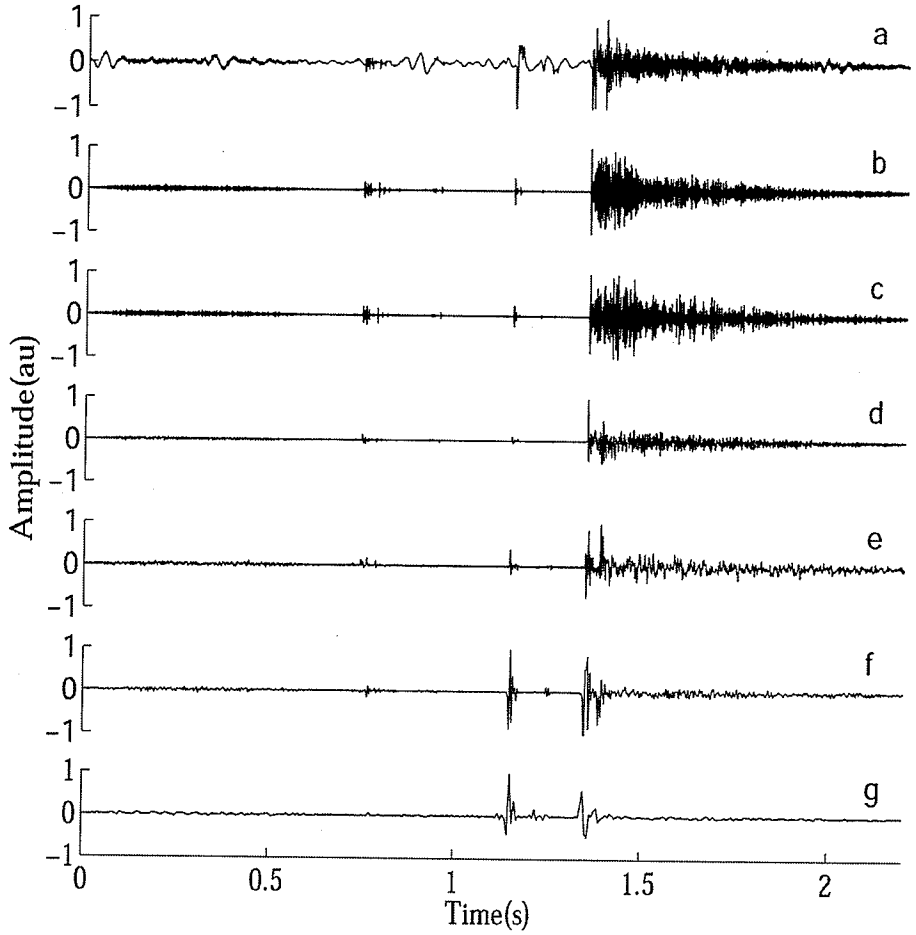


Fig. 4.1. A typical tracheal sound signal decomposed by multiresolution analysis. (a) Original signal; (b-g) the wavelet coefficients of the tracheal signal at levels 1-6, respectively.

applied to WT_S^k and WT_B^k . The iterative procedure stops after a predefined number of iterations or when:

$$STC = \mid E\{X_{B,k-1}^2(n)\} - E\{X_{R,k}^2(n)\} \mid < \varepsilon, \tag{4.1}$$

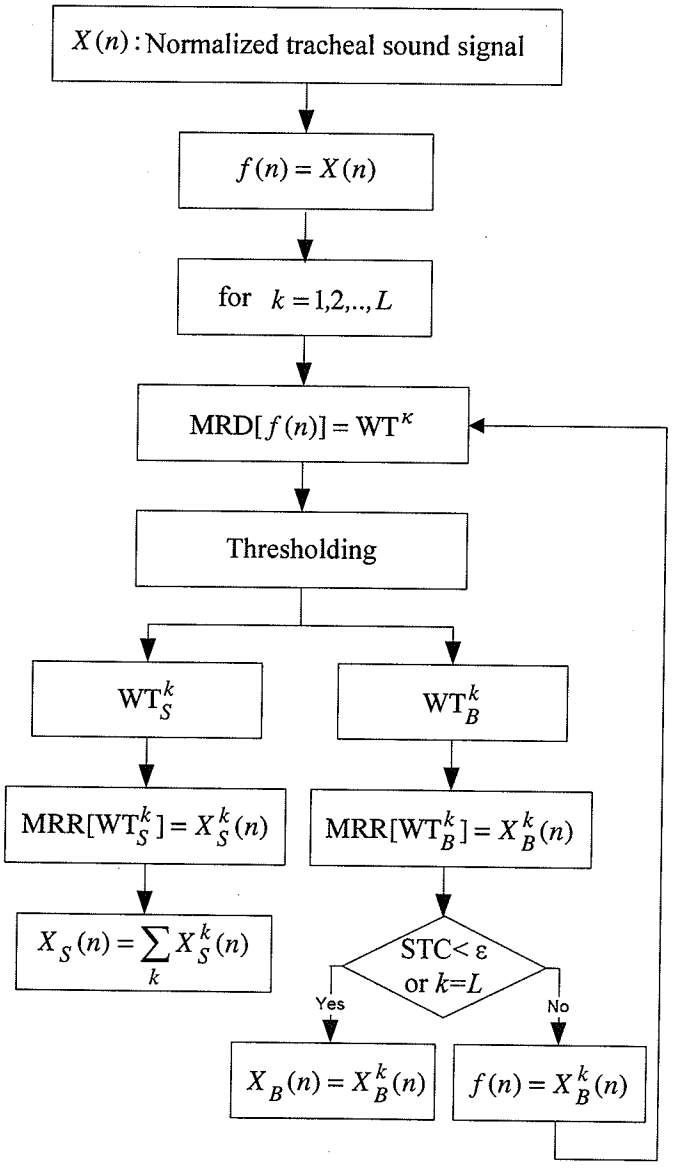


Fig. 4.2. Schematic description of the wavelet transform based filtering method.

where $E\{.\}$ denotes the expected value and ε is a very small positive number by which the desired accuracy in the detection procedure is determined. When convergence is achieved or after iteration L, $X_S(n)$ is determined by combining the coherent parts derived at iterations $k = 1,...,L$, i.e.,

$$X_S(n) = \sum_{k=1}^{L} X_{S,k}(n). \tag{4.2}$$

Although the filter described above has been reported to be successful in separation of discontinuous adventitious sounds from vesicular sounds [Hadjileontiadis and Panas, 1997], suppressing noise in partial discharge on-line monitoring [Mingyou *et al.*, 2000], and enhancement of bowel sounds [Hadjileontiadis *et al.*, 2000], $X_S(n)$ and $X_B(n)$ obtained by this filtering method give only an approximation of the swallowing and breath sound signals. In many cases, $X_B(n)$ contained some parts of swallowing sounds. Therefore, in order to detect boundaries of the swallowing and breath sounds more accurately, the RMS of 128 sample windows of $X_S(n)$ and $X_B(n)$ were compared. If for a segment

$$\mathrm{RMS}(X_S(n)) > \mathrm{RMS}(X_B(n)), \tag{4.3}$$

then, the segment was considered to be part of the swallowing sound signal; otherwise, it was considered to be part of the breath sound signal.

4.2.2 Swallowing Sound Detection Using Recurrence Plot Features

The second method was based on recurrence plot features of tracheal breath and swallowing sounds. As the first step for extraction of the features, multidimensional state

space representation of the tracheal sound signal was reconstructed based on the Taken method of delays. Recurrence plot features were then extracted.

The measures used in this study are based on estimation of the distribution of diagonal structures. The following parameters were calculated for quantitative analysis of recurrence plots:

1- The %REC, which is the percentage of a recurrence plot occupied by recurrence points. The %REC is defined as:

$$\%\text{REC}=\frac{1}{M^2}\sum_{i,j=1}^{M} R_{ij}\times 100\% \tag{4.4}$$

2- The %DET, which is the percentage of recurrence points forming diagonal lines. The %DET is defined as:

$$\%\text{DET}=\frac{\sum_{l=2}^{M} l.P(l)}{\sum_{l=1}^{M} l.P(l)}\times 100\%, \tag{4.5}$$

where l is the length of the diagonal lines and $P(l)$ is the line length distribution function of the diagonal lines. This parameter is lower when recurrence points are scattered on the recurrence plot.

3- The ratio of %DET to %REC, which addresses nonstationarity in the recurrence plots.

4- The Shannon entropy of the length of the lines parallel to the identity line, which addresses the complexity of the recurrence plot; it is defined as:

$$\text{E}=-\sum_{l} P(l).\log(P(l)) \ . \tag{4.6}$$

5- Max Line, which is the length of the longest line parallel to the identity line of the recurrence plot. The reciprocal of Max Line has been used for estimation of the largest Lyapunov exponents and monitoring the divergence in experimental data [Zbilut *et al.*, 2002b].

Along with the above-mentioned parameters, recurrence time statistic parameters as addressed in [Gao, 1999] and [Gao *et al.*, 2003] were calculated. For calculating recurrence time statistics parameters, a reference point is arbitrarily chosen on the trajectory. The points on the reconstructed trajectory that belong to the neighborhood of radius r of the reference point are called the Poincare recurrence points. The set of Poincare recurrence points is denoted by

$$S_1 = \{X(t_1), X(t_2), \cdots, X(t_i), \cdots\}, \tag{4.7}$$

and the Poincare recurrence times are defined as

$$T_1(i) = t_{i+1} - t_i, \quad i = 1,2,\cdots \tag{4.8}$$

$T_1(i)$ are also called the recurrence times of the first type. The points with $T_1(i)=1$ are called sojourn points. The recurrence points that are not sojourn are called recurrence points of the second type. The recurrence times of the second type are denoted by $T_2(i)$. Application of the averages of the Poincare recurrence time of the first and second types for detection of weak transitions in signal dynamics were studied in [Gao *et al.*, 2003].

To calculate the above mentioned measures of the recurrence plots, the multidimensional state space trajectory of each recorded tracheal sound signal containing swallow and breath sounds was reconstructed using the Taken method of delays. The reconstructed trajectories were divided into sequences of 25 ms (256 samples) segments with 50% overlap between successive segments and the recurrence parameters of each

segment were calculated. The segment duration was selected to be 25 ms to compromise between stationarity of each segment and reasonable number of points in the recurrence plots.

Since the preliminary results suggested that some recurrence parameters including %REC, ratio of %DET to %REC, and the entropy of the length of the diagonal lines were appreciably different in swallowing and breath sounds, an automated method for extraction of swallowing sounds in the records of the tracheal breath and swallowing sounds was also investigated. In order to separate swallowing sounds from breath sounds, each of the above-mentioned parameters was individually tested against a pre-determined threshold to separate swallowing sounds from breath sounds. When %REC and entropy were used for swallowing sound detection, the segments with %REC or entropy higher than the threshold were classified as a segment of swallowing sound, whereas when the ratio parameter was used, the segments with ratio higher than the threshold were considered to be part of breath sound. In order to determine the threshold, Otsu's method [Otsu, 1979] was used; it selects the threshold such that the intraclass variance is minimized.

4.2.3 Swallowing Sound Detection Using Hidden Markov Models

As explained in section 4.2.2, three recurrence plot features have shown discriminating ability between swallowing and breath sounds. In section 4.2.2, thresholding of the features was proposed for swallowing sound detection. In the third method for swallowing sound detection, the same three features were modeled by HMMs. Three recurrence plot parameters (the %REC, ratio of %DET to %REC, and

entropy of the length of the diagonal lines of the recurrence plot of a running window across tracheal breath and swallowing sound signal) as described in the previous section were used as observation sequences of HMM. The results of swallowing sound detection using HMM were then compared to those of using thresholding of the original recurrence plot parameters (described in the previous section).

Three HMMs were tested for swallowing sound detection. Each HMM modeled the sequence of one of three recurrence plot features. Each model had two states. One state was assigned to breath sounds and the other state was assigned to the swallowing sounds. The models were ergodic (fully connected) HMMs, in which one state of the model could be reached from the other state of the model; therefore, every a_{ij} was positive. The sequence of each feature and the sequence of the manually detected states were used for training each model.

Once the three HMMs were trained, determination of the best sequence of the states to which observations belong was the next problem to solve. In other words, given the observation sequence and the model parameters, the most likely states must be determined, for which the Viterbi algorithm [Viterbi, 1967] was used. The boundaries of the states assigned by the Viterbi algorithm were considered as the boundaries of swallowing and breath sounds.

4.3 Swallowing Sound Segmentation

Swallowing sound signals do not have a unique temporal and durational pattern. Additionally, the phases of swallowing sounds, especially IDS and BTS, are not easily distinguishable from each other. The next goal of this study was segmentation of

swallowing sound signals into their characteristic segments. Since FDS are not necessarily present in all swallowing sounds, the swallowing sound signals were segmented into three phases namely initial quiet period, IDS and BTS.

The manually extracted swallowing sound signals were divided into sequences of 25 ms (256 samples) segments with 90% overlap between the successive segments. Since the duration of IDS can be very short in many cases (as short as 30 ms), the time resolution required for this study is higher than that for entire swallowing sound detection. Therefore, the overlap between successive segments was chosen to be 90%. Since the recurrence plot features did not show enough sensitivity for the segmentation of swallowing sound signals, each segment was represented by several other features that were hypothetically useful for HMM based swallowing sound segmentation: waveform fractal dimension (WFD); linear prediction coefficients (LPC); average power (P_{ave}) of the signal over different frequency bands; and root-mean-square (RMS) of the time domain signal and multi-scale products (MSP) of the wavelet coefficients. A comparison between features was performed in order to determine which feature(s) provided the best performance.

WFD is a non-integer or fractional dimension of a waveform. This feature is a measure of the degree of complexity of a signal, and is frequently used in biomedical signal processing. There are a few different definitions for WFD. In this study, it was calculated based on the Katz algorithm [Katz, 1988] which is less sensitive to noise compared to other methods such as Higuchi and Petrosian methods [Esteller *et al.*, 2001] or variance fractal dimension [Gnitecki and Moussavi, 2005]. WFD is defined as:

$$D = \frac{\log n}{\log n + \log \frac{d}{L}}, \tag{4.9}$$

where L is the total length of the curve (the sum of distances between successive points), d is the diameter estimated as the distance between the first point of each segment and the point of the segment that provides the farthest distance and n is the number of steps in the curve defined as $n = L/a$, where a is the average step or average distance between successive points [Katz, 1988].

LPC coefficients of 8th order were computed using the Levinson-Durbin recursion method [Proakis, 2006]. P_{ave} over different frequency bands from 200 to 1600 Hz in 200 Hz frames (i.e. 200-400, 400-600, 600-800, 800-1000, 1000-1200, 1200-1400, and 1400-1600 Hz) were calculated from the spectrum of each segment.

MSP of the wavelet coefficients of the signals were used as another characteristic feature. Multiplying the wavelet transform coefficients at adjacent scales amplifies singularities, whereas it reduces noise [Xu *et al.*, 1994], [Stadler and Swami, 1999], [Moussavi *et al.*, 2004]. Therefore, the product of wavelet transform coefficients has higher amplitudes in discrete swallowing sounds, i.e., IDS and FDS. In this study, the wavelet transform coefficients of each segment were computed using the Debauchies-5 wavelet and the product of scales 2 and 4, the product of scales 4 and 8, and the product of scales 2, 4 and 8 were calculated. Each product contained 256 samples. The RMS of each product was used as a characteristic feature of the swallowing sound segment. RMS of the multiscale products of the scales 2 and 4, 4 and 8, and 2, 4 and 8 hereinafter are referred to as MSP1, MSP2 and MSP3, respectively. Overall, each 25 ms segment was represented by seven features (WFD, LPC, P_{ave} , RMS, MSP1, MSP2, MSP3 with 1, 8, 7,

1, 1, 1, 1 coefficients, respectively).

Several HMMs were tested for segmentation of swallowing sounds into three distinct states, i.e., initial quiet period, IDS and BTS. Since durational patterns of the swallowing mechanism are functions of the viscosity of the bolus being swallowed [Rempel and Moussavi, 2005], individual models were created for each of the three bolus textures. Each model used one of the seven features described above. The number of hidden states in all models was set to three, each representing one of the swallowing sound states.

For swallowing sound segmentation, ergodic (fully connected) HMMs are not suitable, because swallowing sound phases occur in order. Hence, left-right models were used, in which the state index increases or stays the same as the time index increases. In other words, no transitions are allowed to states whose indices are lower than the current state. Obviously, the state sequence must begin in state one. Therefore, the initial state probabilities have the property

$$\pi_i = \begin{cases} 1 & i = 1 \\ 0 & i \neq 1 \end{cases}, \tag{4.10}$$

In order to prepare the data set for training and testing HMMs, available videofluoroscopy images were used in manual detection of the three states. Also, the swallowing sound signals without videofluoroscopy images were manually segmented by an expert. Since the number of subjects was not large enough for randomly dividing the swallowing sound signals to training and testing data sets, the jackknife approach was utilized in which the swallows of one subject from the data set were removed and HMMs were trained using the rest of the data. In order to train HMMs, the sequences of sound signal features and sequences of manually detected states were used.

Trained HMMs were used for segmentation of the swallows of the subject whose swallows had been removed, i.e., once the HMMs were trained, the Viterbi algorithm was used to find the most likely states of the trained model and the boundaries of the states assigned by HMMs were compared with those detected manually. Then, the training and testing process was repeated for all subjects one by one.

4.4 Swallowing Sound Characterization

Characteristics of swallowing sound signals were studied using two nonlinear dynamic parameters: correlation dimension and the largest Lyapunov exponents. Correlation dimension is used to measure the dimensionality of the space occupied by a trajectory. The largest Lyapunov exponent is a parameter commonly used to detect the presence of chaos in dynamical systems.

The methods used for evaluation of these parameters were described in the previous chapter. In this study, feasibility of application of these methods to swallowing sound signals for evaluation of the complexity of underlying dynamical system and the presence of chaos were examined. Since swallowing sound signals are basically nonstationary, correlation dimension and the largest Lyapunov exponents were individually calculated for IDS and BTS segments with more stationarity than the entire swallowing sound signals.

4.5 Swallowing Sound Classification

Another goal of this study was classification of swallowing sounds of healthy and

dysphagic subjects. Two methods were used for this purpose.

4.5.1 Classification Using Nonlinear Dynamic Parameters

The first method for classification was based on nonlinear dynamic parameters. Optimum time delay, correlation dimension and the largest Lyapunov exponent were initially considered as characteristic features of swallowing sound signals. However, due to inconsistency in the estimation results of the largest Lyapunov exponents, this feature was excluded from the feature set. A three-nearest neighbor classifier and a two-layer feed-forward neural network were created to classify the swallowing sound signals of normal and dysphagic subjects based on nonlinear dynamic parameters. The optimum time delay and correlation dimension of IDS and BTS segments (four features in total) were used as features for the classifiers.

For the three-nearest neighbor classifier, the swallows of one subject were first removed from a set of labeled swallows of all subjects. Then, the three-nearest neighbor rule classified each swallow of the subject whose swallows had been removed by assigning it the label most frequently represented among the three nearest samples of other subjects' swallows. In other words, a decision was made by examining the labels on the three nearest neighbors and taking a vote. This process was repeated for all subjects.

Since the number of subjects was not large enough for randomly dividing the recordings to training and testing data sets, the jackknife approach was utilized for training and testing the neural network classifier, in which the accuracy was estimated by training the classifier using all data but the data of one subject, and each time testing it on the left out data as mentioned above. The jackknife estimates of the classification were

averaged between the subjects.

4.5.2 HMM-Based Classification

The second method for classification was based on HMMs. Individual models for swallowing sounds of healthy and dysphagic swallows and for each bolus texture were trained. The number of states was set to different values from four to nine states. In this case, a hidden state does not necessarily correspond to a physiological state. Therefore, ergodic (fully connected) models were used, in which every state of the model could be reached from every other state of the model. The iterative expectation maximization (EM) algorithm [Duda *et al.*, 2001] was used to train the models, i.e., to find model parameters that would result in the maximum likelihood of the observation sequences.

Once all of the models were trained, the feature sequences of each of the swallowing sound signals of the testing data set were computed. Then, the likelihood of each observation sequence based on HMM belonging to healthy or dysphagic groups was computed. If the likelihood of a sequence based on the model of healthy subjects was higher than that of the model of dysphagic patients, the swallowing sound signal was considered as a swallowing sound signal of a healthy subject. Otherwise, it was considered as that of a dysphagic subject.

4.6 Timing Analysis of Submental Electromyogram (EMG)

As described in Introduction (chapter 2), during a swallow, the floor of the mouth muscles contract and pull the larynx and hyoid bone. This contraction results in elevation

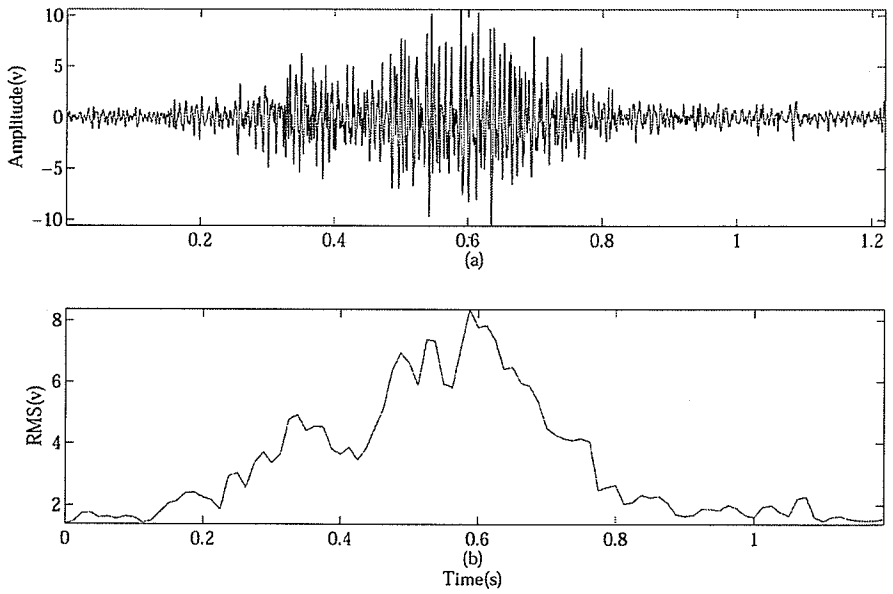


Fig. 4.3. (a) The waveform of EMG signal in time domain and (b) its RMS.

and forward movement of the larynx and hyoid bone. The elevation contributes to closure of the airway entrance and the forward movement contributes to opening of the upper esophageal sphincter. When the submental muscles start contracting, the amplitude of the submental surface EMG increases drastically. Accordingly, the RMS of the EMG activities during submental muscles contraction is much larger than that when the muscles are at rest. In this study, the EMG signal was divided into a sequence of 25 ms (256 samples) segments with 50% overlap between the successive segments. Each segment was then represented by its RMS. A typical example of the swallowing EMG signal in the time domain and its RMS sequence are shown in Fig. 4.3. The RMS of each segment was compared with a threshold (15% of the maximum value of the RMS sequence). This threshold level was empirically selected to ensure that strong noises were

not considered as muscle contraction. When RMS was larger than the threshold, submental muscles were considered to be active.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Swallowing Sound Detection

In this study, three methods based on wavelet transform, thresholding of recurrence plot features and hidden Markov modeling of recurrence plot features were used for swallowing sound detection within breath and swallowing sound records. The results are presented in the following subsections.

5.1.1 Wavelet Transform Based Swallowing Sound Detection

The wavelet based filtering method discussed in Section 4.1, followed by the extra criterion expressed in (4.3), was used to detect swallowing sounds. Figure 5.1.a illustrates a 5-second period of a typical tracheal sound signal, which was processed by the proposed method. Figures 5.1.b and 5.1.c show the refined swallowing and breath sound signals, $X_S(n)$ and $X_B(n)$, respectively. The refined signals, $X_S(n)$ and $X_B(n)$, do not contain most parts of breath and swallowing sounds, respectively; however, some parts of unwanted sounds are still in the refined signals. Therefore, the filtering method alone is not able to separate the two swallowing and breath sounds with accurate boundaries. However, by applying the criterion in (4.3), the boundaries of the swallowing sounds were detected and the results are shown in Fig. 5.2. As it can be seen, during the detected swallowing period, the recorded airflow is zero. The onset of the swallow

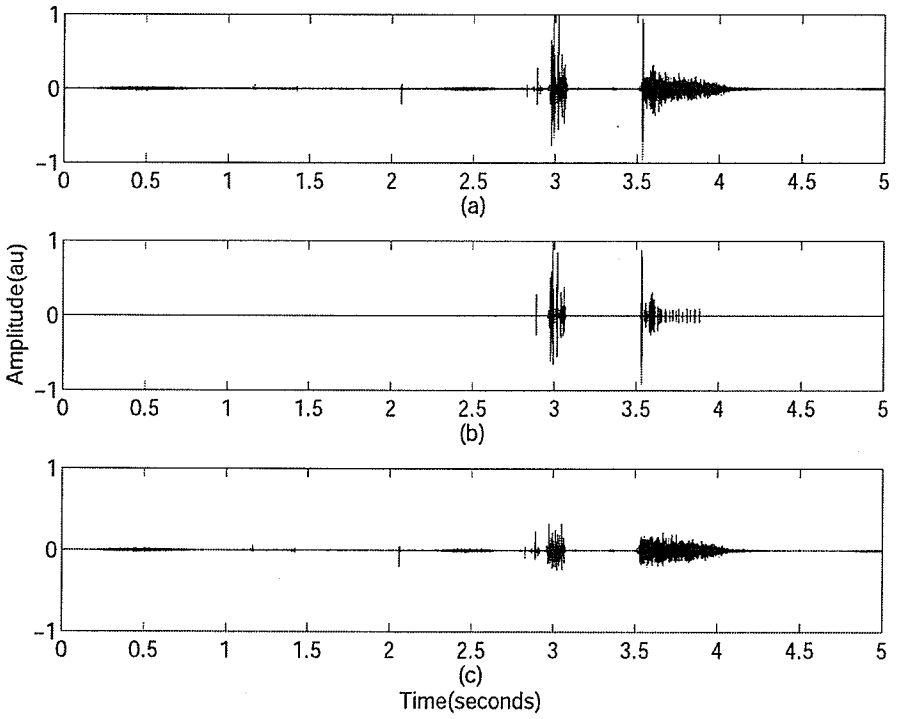


Fig. 5.1. A typical tracheal sound signal decomposed by the wavelet-based filter: (a) original signal; (b) extracted swallowing sound signal; (c) extracted breath sound signal using wavelet transform based filter. Legend: 'au' arbitrary unit.

coincides with the end of the preceding respiratory phase and the end of the swallow is the onset of the respiratory phase following the swallow.

The results of swallowing sound detection were validated manually by visual inspection of the time-domain signal, the simultaneously recorded airflow signal and the spectrogram of the sounds, along with auditory means (listening to the sounds repeatedly). Evaluation of the performance through experimental results on data of healthy subjects and dysphagic patients showed that the method identified 91.3% of the swallowing sound segments correctly, i.e., it could not detect 8.7% of all swallows. In

addition, the method has a false positive error of 13.5%; implying that some breath or extraneous noises such as tongue movement were detected as swallowing sounds by error.

5.1.2 Swallowing Sound Detection Using Recurrence Plot Features

As explained in section 3.1, the first step in the nonlinear analysis of a scalar time

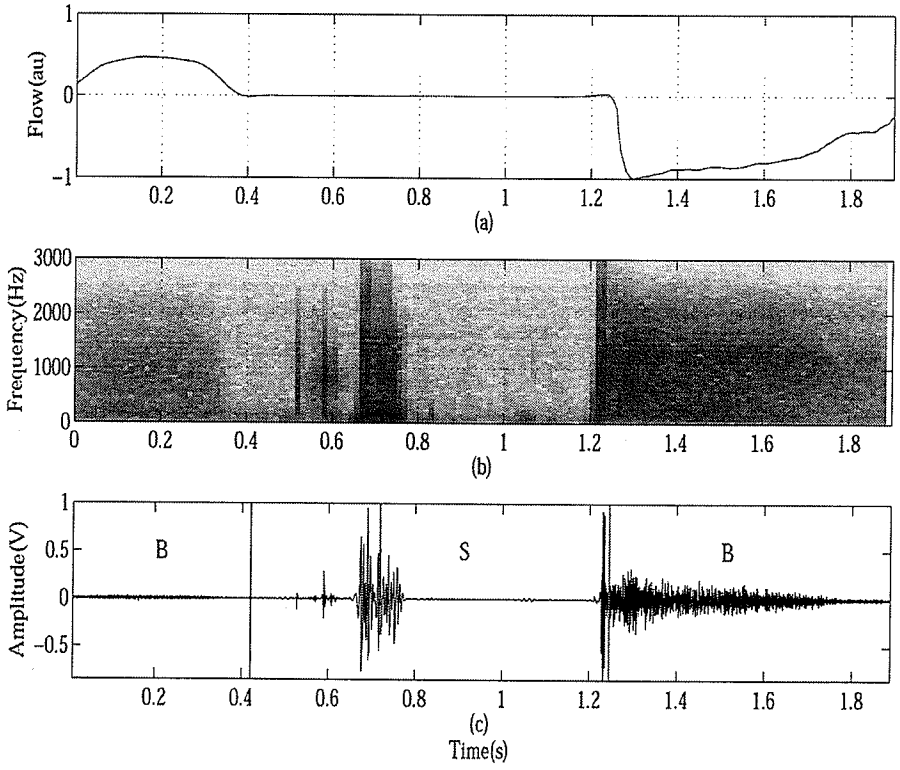


Fig. 5.2. (a) Airflow signal, (b) spectrogram and (c) tracheal sound waveform as a function of time. Legend: B: Breath Sound, S: Swallowing sound, au: arbitrary units for normalized amplitude, red vertical lines show onset of swallows and green vertical lines show the end of swallowing segments.

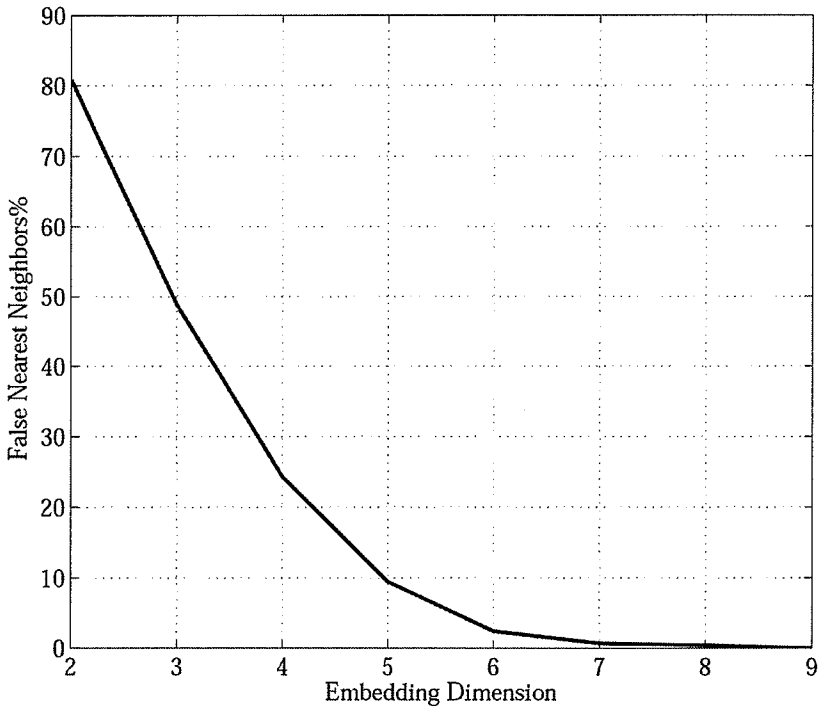


Fig 5.3. False nearest neighbors percentage as a function of embedding dimension.

series is the reconstruction of the multidimensional state space trajectory. In order to reconstruct the trajectory, the proper embedding dimension and time delay must be determined.

In order to determine the proper embedding dimension, the percentage of false nearest neighbors for different values of embedding dimension were computed. Figure 5.3 shows the curve of the percentage of false nearest neighbors for different values of embedding dimension when the state space representation of a swallowing sound signal of a healthy subject was reconstructed. Regardless of the subject and bolus texture, the curves were all similar to that shown in Fig. 5.3. In all cases, for embedding dimension $E = 8$, the false nearest neighbor percentage was less than 1%. In many cases,

even with $E = 7$ the false nearest neighbor percentage was also less than 1%. However; $E = 8$ was considered as the suitable embedding dimension for topologically proper reconstruction of the time series representing the recorded swallowing and breath sounds in the presence of noise. As the signals were nonstationary, a sufficiently high dimension was selected that contained the relevant dynamics of all parts of the signal passing from one dimension to another dimension in the presence of noise. A similar study on nonlinear dynamics of tracheal breath sounds [Ahlstrom *et al.*, 2006] has shown that proper embedding dimension ranged between 6 to 8, which is congruent with the results of present study.

In order to determine the proper time delay, the average displacement of the embedding vectors from their original locations on the line of identity as a function of time delay ($< S_E(\tau) >$) for different embedding dimensions E = 7,8,9 were computed. As an example, the curve of $< S_E(\tau) >$ for embedding of a tracheal sound signal of a healthy subject is shown in Fig. 5.4. The best time lag was chosen as the point, where the slope of the curve decreased to less than 40% of its initial value. The optimum values of τ varied between 3 to 11 samples (0.3 to 1.1 ms) depending on the subject and bolus texture. In the study by Ahlstrom *et al.* [2006] on tracheal breath sounds, the optimum time delay was computed based on the first local minimum of the mutual information function and ranged between 0.4 to 1 ms (between 2 to 5 samples with a sampling frequency of 5000 kHz).

Once the multidimensional state space trajectories of tracheal sound signals were reconstructed, recurrence plots of 25 ms segments of the trajectories were generated. Examples of recurrence plots of 25 ms segments of swallowing and breath sound signals

are shown in Fig. 5.5. The number of black dots in the recurrence plot of swallowing sounds (Fig 5.5.b) is evidently higher than that of breath sounds (Fig 5.5.a), which implies that the %REC of swallowing sounds is higher. Interpretation of the recurrence plots shown in Fig. 5.5 in terms of other recurrence plot features is not straightforward and requires quantitative evaluation.

Figure 5.6 shows a typical tracheal sound signal containing a swallow occurring after an inspiration and followed by an expiration, along with their corresponding airflow and the plot of the recurrence plot parameters (i.e., %REC, %DET, etc.). The boundaries

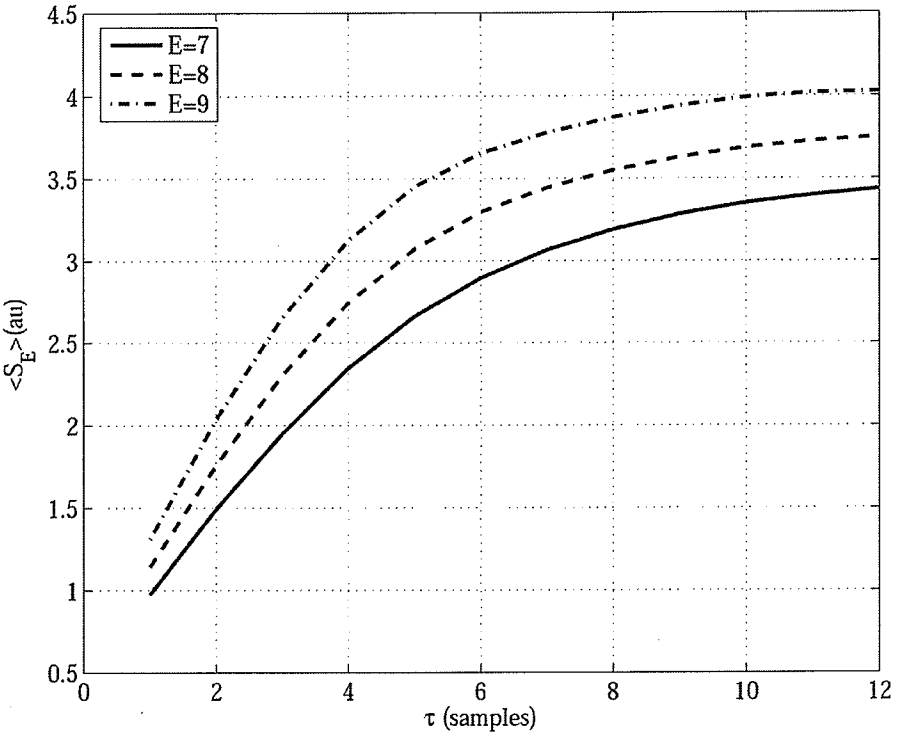


Fig 5.4. Average displacement as a function of time delay for different embedding dimensions. Legend: 'au' is arbitrary unit for normalized amplitude.

between swallowing and breath sounds were manually detected by visual inspection of airflow measurement and spectrogram of the sounds as well as by auditory means, and are shown in Fig. 5.6 by the vertical dashed lines. The first dashed line shows the transition from breath sounds to swallowing sounds and the second dashed line indicates the transition from swallowing sounds to breath sounds. As it is observed from Fig. 5.6, the %REC of the breath sound signal is appreciably lower than that of the swallowing sound signal whereas the %DET of swallowing and breath sounds are very close to each other. As a result, the ratio of %DET to %REC of swallowing sounds is less than that of breath sounds. The ratio is lower in the nonstationary swallowing sound signals. The Shannon entropy parameter (Fig. 5.6.f) also showed discriminating ability for swallowing and breath sounds; this parameter was appreciably lower during breath periods than the swallowing portions, implying a more uniform distribution of the line length in the recurrence plots of the breath sounds.

Although the %REC, ratio and entropy trajectories showed discriminating capability

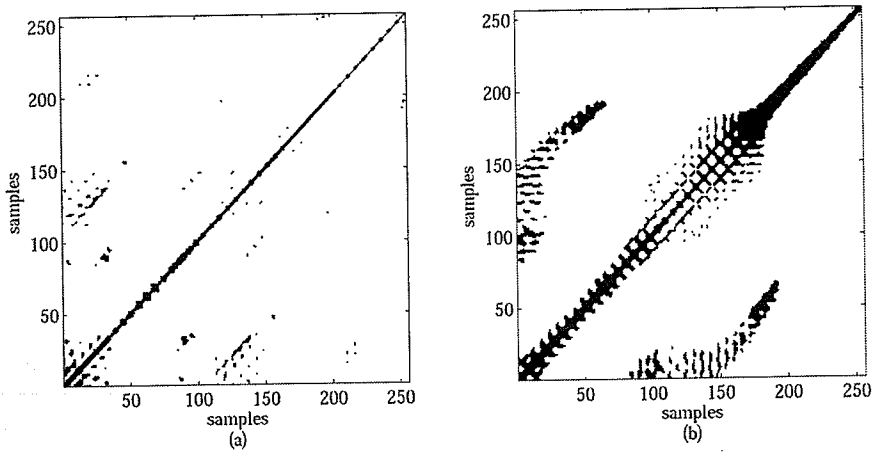


Fig 5.5. Recurrence plots of (a) breath, and (b) swallowing sounds.

between swallowing and breath sounds, the averages of Poincare recurrence time of the first and second types were not much different between swallowing and breath sounds as shown in Fig. 5.7.d and 5.7.e. In search for other features with discriminating capability, the maximum and average length of the parallel lines to the identity line of recurrence plots and Shannon entropy of the original tracheal sound signal were also tested. The maximum and average length of the diagonal lines did not show discriminating ability (Fig. 5.7.b and 5.7.c). Shannon entropy of the original tracheal sound signal has been shown as a robust tool for airflow estimation [Yadollahi and Moussavi, 2006a]. However, as shown in Fig. 5.7.f, this feature cannot be used with a simple thresholding algorithm for detection of swallowing sounds. Entropy of the tracheal sound is not zero while respiration halts during the pharyngeal phase of the swallowing (in which swallowing sounds are generated).

Mean and standard deviation for the recurrence parameters are shown in Table 5.1. As these data suggest, the mean values of the %REC, ratio, and entropy are significantly different between swallowing and breath sounds. There are other parameters that their means are different for swallowing and breath sounds; however, only thresholding the above-mentioned parameters resulted in a reasonable separation of swallowing segments from breath sounds. Therefore, each of these three parameters was individually used for automatic swallowing sound detection in the tracheal sound recordings.

Table 5.1. Mean ± standard deviation for recurrence parameters of the swallowing and breath sound signals

	Swallowing Sounds	Breath Sounds
%REC	13.84±11.54	2.85±4.02
%DET	49.84±2.54	52.28±4.59
Ratio	8.98±8.60	28.38±11.29
Entropy	2.40±0.94	1.31±0.49
Average Length	18.66±15.90	11.55±6.50
Maximum Length	157.57±66.39	89.93±66.21
$\overline{T}_1$	6.09±5.09	12.04±7.04
$\overline{T}_2$	52.39±28.52	69.34±17.98

Figure 5.8 illustrates a typical tracheal sound signal which was processed for swallowing sound detection using the above mentioned parameters. Figure 5.8.b-d shows the boundaries detected by thresholding the %REC, ratio and entropy, respectively. The onset of the swallow coincides with the end of the preceding respiratory phase and the end of the swallow is the onset of the respiratory phase following the swallow. The swallow shown in Fig. 5.8.a is followed by a forceful expiration and yet the thresholding method was able to detect the boundaries of the swallowing sounds correctly. However, the detected boundaries by the thresholding of different features were not exactly the same.

The detection results both in terms of missed or misclassified swallows as well as accuracy in the boundary detection were validated manually by visual inspection of the airflow measurement and spectrogram of the sounds as well as by auditory means. Table 5.2 indicates the performance of thresholding of each of the above mentioned features in terms of false positives, missed swallows and the accuracy of boundary detection. In order to facilitate comparing the results of different detection methods, all of the results are presented in Table 5.2. Among the thresholding based methods, thresholding the %REC showed an overall better performance compared with thresholding the ratio and entropy. Its percentage of false detection was less than those of other features, whereas its percentage of missed swallows was lower than that of thresholding the entropy. In addition, when thresholding of %REC was used, the mean and standard deviation for delay in the detection of swallowing sound was less than the mean and standard deviation of delay when thresholding of the ratio or entropy were used.

In order to decrease the false positive error, two other algorithms were tested, in which a segment was considered as a swallowing sound segment if thresholding of two or three features (i.e., %REC, ratio and entropy) indicated that the segment was a swallowing sound. These two criteria are called V2 and V3, respectively, and the results are shown in Table 5.2. Although the V2 algorithm does not make much difference, V3 decreases the percentage of false positive error and the delay in boundary detection. However, this algorithm increases the percentage of missed swallows. Therefore, depending on the application, one may choose any of these methods individually or a combination of them.

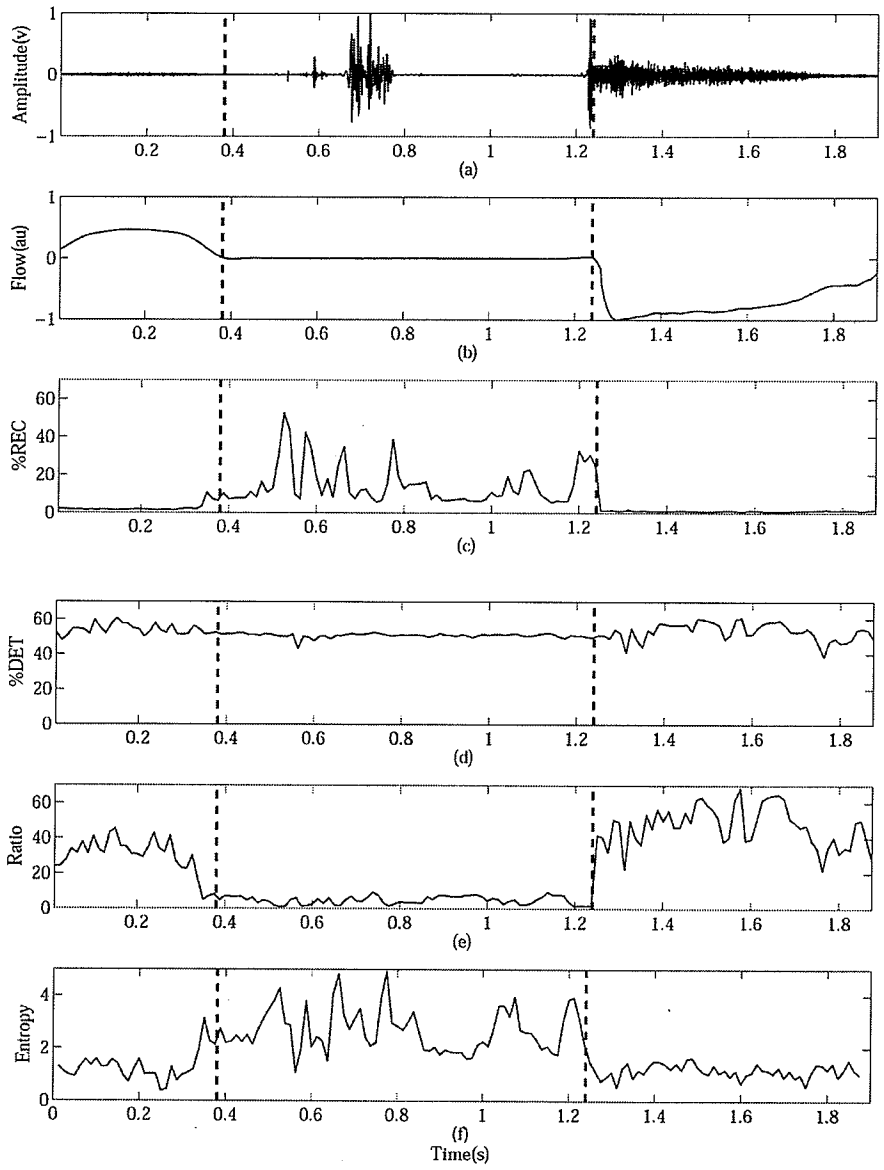


Fig 5.6. (a) A tracheal swallowing sound signal preceded and followed by breath sounds; (b) airflow; (c) %REC; (d) %DET; (e) ratio of %DET to %REC; (f) entropy of the length of the lines parallel to the identity line. The boundaries between swallowing and breath sounds were detected manually and shown with the vertical dashed lines.

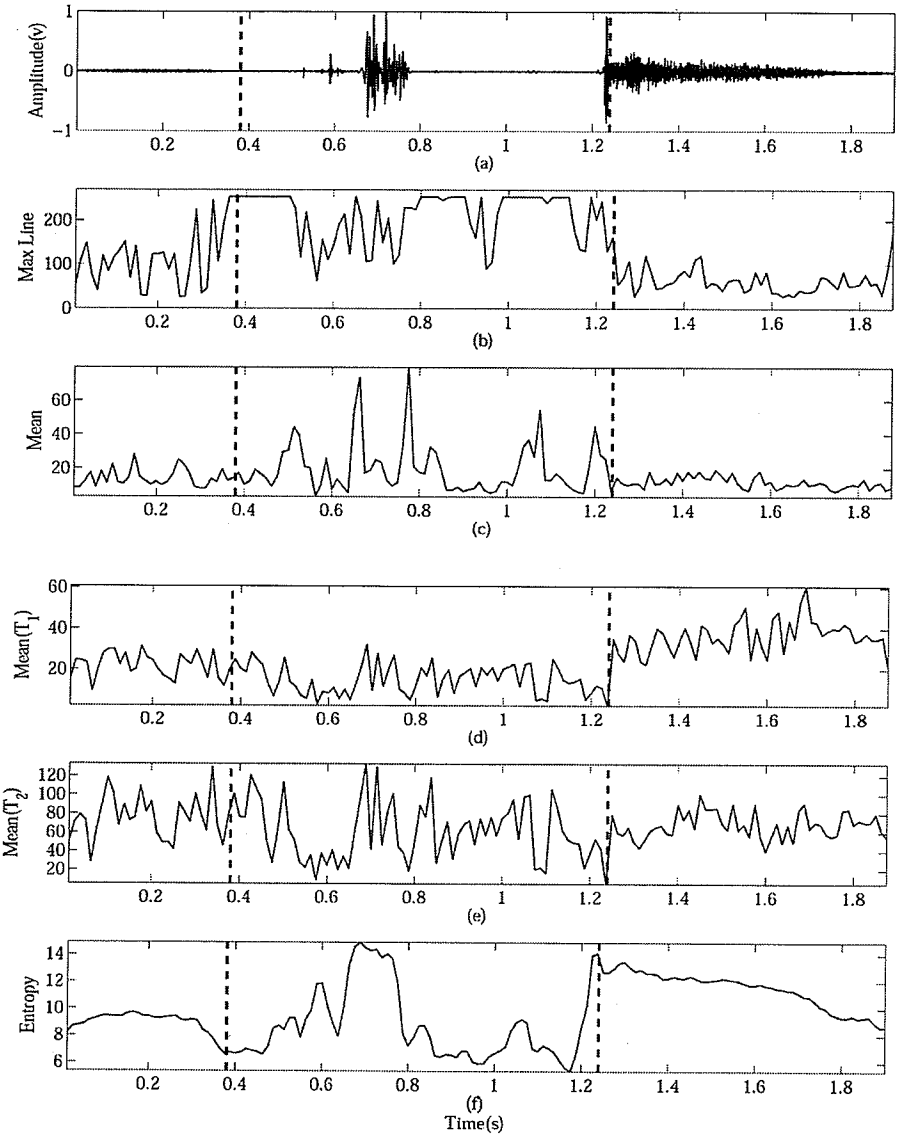


Fig 5.7. (a) A tracheal swallowing sound signal preceded and followed by breath sounds; (b) The maximum length of the lines parallel to the identity line of recurrence plots; (c) The average length of the lines parallel to the identity line of recurrence plots; (d) The average of Poincare recurrence times of the first type; (e) The average of recurrence times of the second type; (f) Shannon entropy of the tracheal sound signal. The boundaries between swallowing and breath sounds were detected manually and shown with the vertical dashed lines.

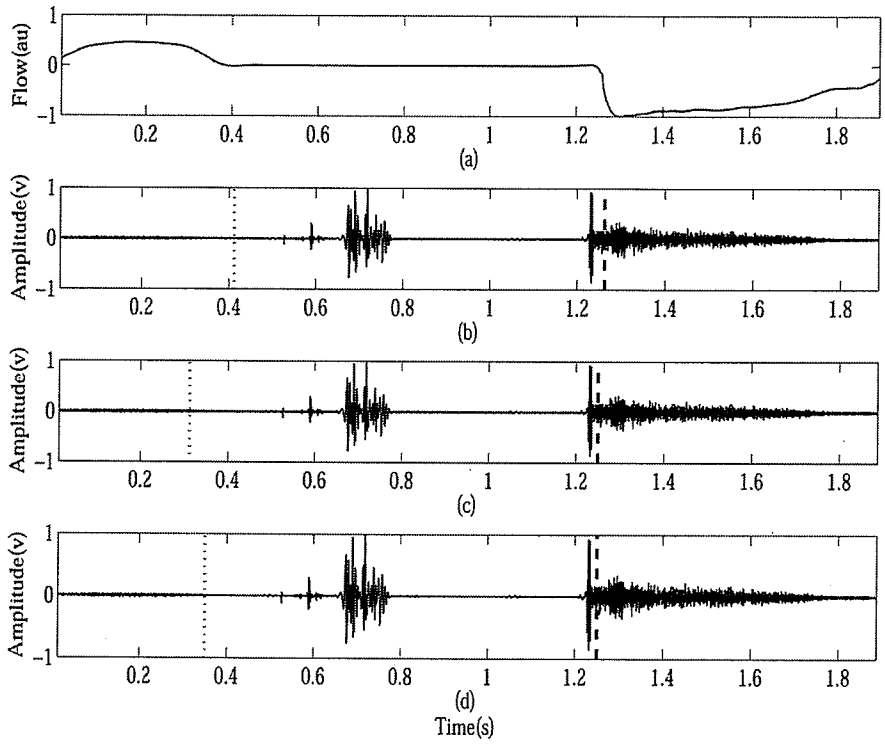


Fig 5.8. (a) Airflow. The tracheal sound signal and the boundaries detected by thresholding of (b) %REC, (c) the ratio, and (d) entropy. The dotted vertical lines show the beginning of the swallowing sounds and the dashed vertical lines show the end of the swallowing sounds.

5.1.3 Swallowing Sound Detection Using Hidden Markov Models

Once the values of the %REC, ratio of %DET to %REC, and entropy of the length of the diagonal lines explained in the previous chapter were extracted, the sequence of each of these three features were modeled by an HMM. The most likely states were found by the Viterbi algorithm. The investigated method was able to detect the boundaries of the swallowing sounds. However, the detected boundaries by modeling of different features were not exactly the same.

Table 5.2. Performance Comparison of the swallowing sound detection algorithms.

Algorithm	False positive%	Missed swallows%	D1	D2
Wavelet-Based Detection	13.5	8.7	101±128	16±232
%REC Thresholding	13.4	4.3	-64±105	88±133
Ratio Thresholding	27.7	2.9	-154±175	142±222
Entropy Thresholding	14.3	7.1	-69±129	63±140
V2	14.2	3.8	-80±127	81±128
V3	10.3	8.9	-36±93	46±123
HMM-%REC	6.7	4.8	-6±74	22±100
HMM-Ratio	15.7	6.4	-166±226	107±223
HMM-Entropy	6.5	8.1	-90±81	32±102

D1: Mean ± SD for delay in the detection of swallowing sound onset (ms)

D2: Mean ± SD for delay in the detection of swallowing sound end (ms)

Table 5.2 shows the performance of the HMMs based on the %REC, ratio and entropy in terms of false positives, missed swallows and the accuracy of boundary detection. The percentages of false positive swallow detection in HMM-based methods are much lower than those of thresholding-based methods. However, percentages of missed swallows in HMM-based methods are slightly higher than those of thresholding-based methods. Additionally, mean and standard deviation for delay in detection of swallowing end when HMM was used is noticeably lower than those of thresholding-based methods. Therefore, it may be concluded that modeling the sequences of the

recurrence plot features has improved the performance of swallowing sound detection. Overall, the performance of HMM-based swallowing detection when %REC was used was better than other methods proposed in this study.

5.2 Swallowing Sound Segmentation

For swallowing sound segmentation, 21 different HMMs (with WFD, RMS, MSP1, MSP2, MSP3, P_{ave}, and LPC as inputs for each of the three textures) were created and trained as explained in the previous chapter. Given the observation sequences and the model parameters, the Viterbi algorithm was used to find the most likely states of each swallowing sound signal.

Figure 5.9 shows a typical swallowing sound signal along with its corresponding WFD and the most likely hidden states of the swallowing sound signal when WFD was used to model the swallowing sound signal. Since the models used for separation were left-to-right models, all sequences of the states assigned by the Viterbi algorithm started in state 1 and ended in state 3. The patterns of the most likely states determined by the Viterbi algorithm were very similar when the input of the model was any of the features. The only difference was the changes observed in the boundaries of the detected states. Figure 5.10 shows the most likely states of a typical swallowing sound signal obtained by the Viterbi algorithm. As it can be observed, the patterns are very much the same for the most likely states when the input of the model is each of the features, with differences in the boundaries of the detected states.

The boundaries of the states determined by the Viterbi algorithm and the sequences of manually detected swallowing sounds were compared. Table 5.3 shows the mean and standard deviation for delays in the detection of boundaries of swallowing sound phases for semisolid, thick and thin liquid textures among all subjects. Negative values in the delays mean that the boundary detected by Viterbi algorithm precedes the manually detected boundary. Delay in detection of the beginning of IDS was less than that of the beginning of BTS (or the end of IDS). That is probably due to the fact that detection of the boundary between the initial quiet period and IDS is easier because the two portions are significantly different in sound intensity. Among the seven features, delays in detection of the beginning of IDS by MSP3-based model were less than those of the other

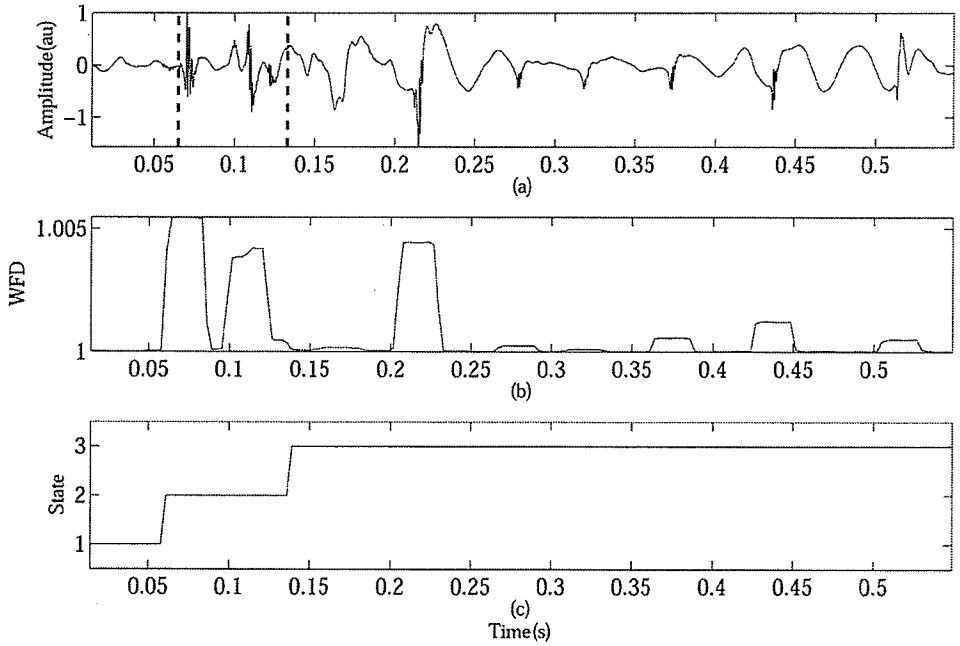


Fig. 5.9. (a) A swallowing sound signal when the subject was fed juice in bolus size of 5 ml in the time-domain; (b) its WFD; (c) the most likely hidden states of the swallowing sounds when WFD was used to model the swallowing sound signal.

Table 5.3. Mean ± standard deviation for delay in detection of boundaries of swallowing sound phases obtained by HMM in different textures (ms).

Feature	Delay in detection of the beginning of IDS	Delay in detection of the beginning of BTS
WFD	-4.2±16.2	24.0±46.5
RMS	-5.2±9.8	35.3±52.1
MSP1	-8.4±22.5	31.4±47.9
MSP2	-10.4±35.2	22.4±49.4
MSP3	2.2±11.2	15.4±43.3
P_{ave}	-5.1±13.7	33.8±52.1
LPC	-21.0±42.9	82.7±107.5

models. This was expected as multiscale product signifies the singular points of a signal, hence resulting in more accurate detection of IDS segments as they are characterized by discrete click sounds. Overall, the LPC-based model showed the weakest performance.

The advantage of using HMM for detection of the boundaries is that when there are several peaks in the sequence of features, HMM can still detect boundaries of swallowing sound signals with reasonable accuracy. In the example shown in Fig. 5.9, there are several peaks in the WFD signal corresponding to IDS and BTS. Therefore, the methods introduced in [Lazareck and Moussavi, 2002b] and [Lazareck and Moussavi, 2004] that were based on thresholding of the WFD [Lazareck and Moussavi, 2004] or variance dimension [Lazareck and Moussavi, 2002b] will most likely fail to detect the swallowing phases boundaries for such a signal, while the proposed method using HMM successfully detected the three phases with reasonably accurate boundaries (Fig. 5.9.c.).

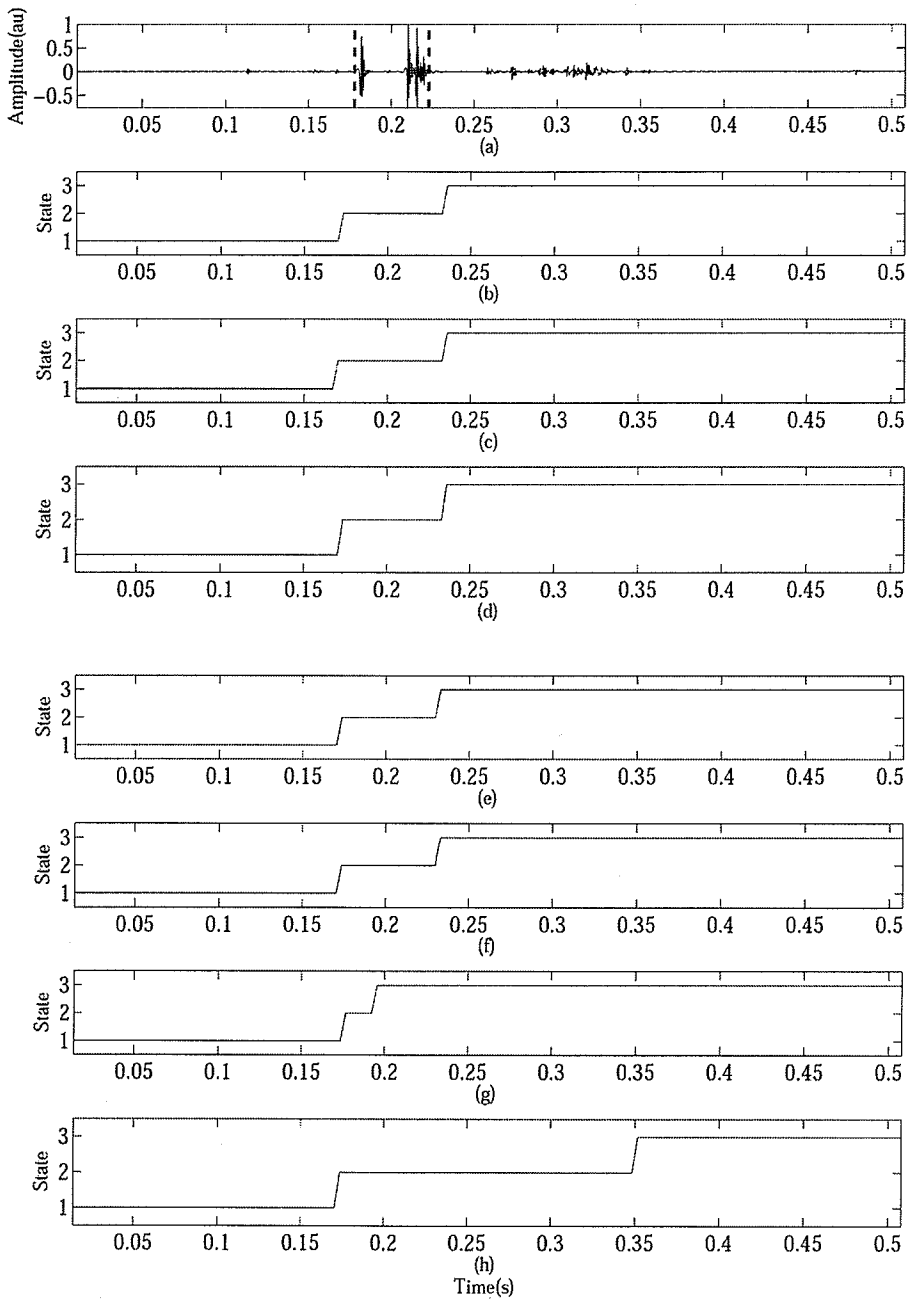


Fig. 5.10. (a) A typical swallowing sound signal in the time-domain. The most likely hidden states of the swallowing sound when (b) WFD, (c) RMS, (d) MSP1, (e) MSP2, (f) MSP3, (g) average power, and (h) LPC were used to model the swallowing sound signal.

5.3 Swallowing Sound Characterization

Using the Grassberger-Procaccia algorithm, correlation dimensions of IDS and BTS segments of the swallowing sounds were evaluated for different subjects and textures. Correlation dimensions did not appear to change appreciably among the swallows of each subject. Correlation dimensions were calculated for embedding dimension 8 and the proper time delay determined by the method discussed in section 3.1. Results are shown in Fig. 5.11. Based on the results, it may be concluded that swallowing sounds are well characterized by a low-order model. Also, the correlation dimensions of swallowing sounds of healthy subjects were generally higher than those of dysphagic patients.

The largest Lyapunov exponent depended on the subject and bolus textures; it varied from -4.9 to 9.7. The inconsistency in the estimated largest Lyapunov exponent even

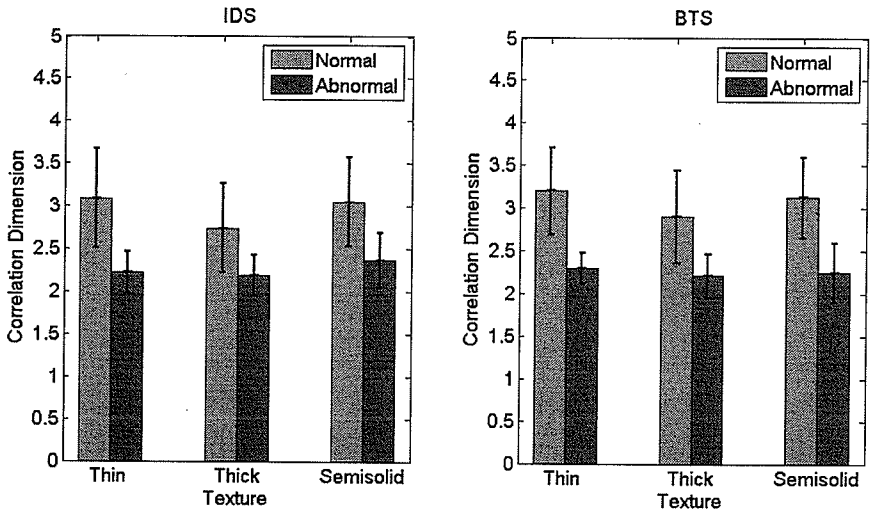


Fig. 5.11. Correlation dimension for IDS and BTS segments in normal and dysphagic subjects.

among different swallows of a particular subject was due to the fact that the recorded swallowing sound signals were short time signals contaminated with noise. As addressed in [Poon and Barahona, 2001], determination of Lyapunov exponents fails to detect the presence of chaos unless the time series is lengthy and free of noise.

5.4 Swallowing Sound Classification

Assessment of swallowing disorders using their acoustical signature was one of the goals of this study. Therefore, two methods based on nonlinear dynamic parameters and hidden Markov modeling were proposed for discrimination between swallowing sounds of normal and abnormal subjects. Classification results using theses two methods are presented in this section.

5.4.1 Classification Using Nonlinear Dynamic Parameters

The first method for classification was based on nonlinear dynamic parameters. Optimum time delay, correlation dimension and the largest Lyapunov exponent were initially considered as characteristic features of swallowing sound signals. These features were extracted for IDS and BTS segments. In order to determine the optimum time delay, the average displacement of the embedding vectors from their original locations on the line of identity as a function of time delay ($< S_E(\tau) >$) were computed and the optimum time delay was chosen as the point, where the slope of the curve decreased to less than 40% of its initial value. Optimum time delay ranges for IDS and BTS segments are shown in Table 5.4.

Table 5.4. Optimum time delay ranges for IDS and BTS segments of swallowing sounds signals in normal and dysphagic subjects.

	Normal			Dysphagic		
	Thin Liquid	Thick liquid	Semisolid	Thin Liquid	Thick liquid	Semisolid
IDS	3-9	3-9	3-9	6-9	5-8	6-8
BTS	4-10	3-10	4-10	7-11	6-10	6-9

Due to inconsistency in the estimated largest Lyapunov exponent, this feature was excluded from the feature set used for classification. Classification of the swallowing sounds was performed using optimum time delay and correlation dimension of IDS and BTS segments. The three-nearest neighbor classifier was found to be superior to the feed-forward neural network classifier. Their accuracies were 83% and 71%, respectively. As a second stage classification, a screening algorithm was used in which if more than 50% of the swallows of a subject were classified as being normal, the subject was considered as normal. Otherwise, the subject was considered at risk of dysphagia. This second stage screening classified 36 subjects of 38 subjects, correctly; one normal subject was misclassified as dysphagic and one dysphagic subject was misclassified as normal.

The proposed method was able to classify normal and marginally normal swallows from each other with high accuracy. Although physical interpretations of nonlinear parameters are usually difficult; however, the results confirm the discriminating ability of the reconstructed phase space features for classification of swallowing sounds of healthy and dysphagic subjects. It should be noted that all the swallows of the dysphagic group were also normal swallows as they were not aspirated. However, since they belonged to the patients at risk of dysphagia, we considered them as marginally normal.

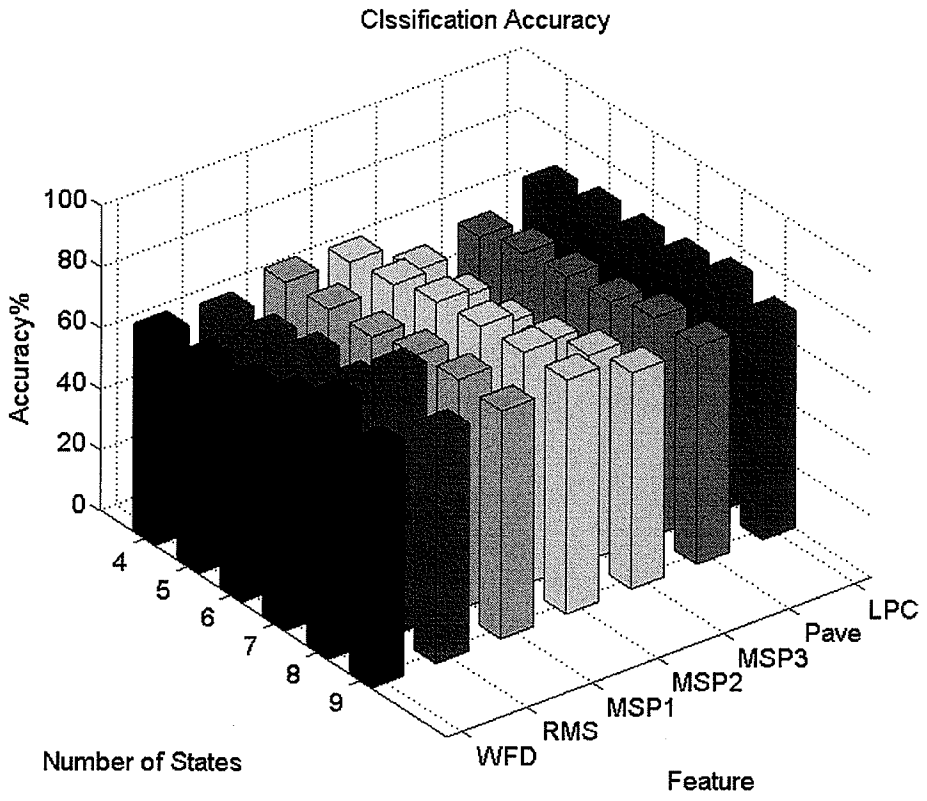


Fig. 5.12. Classification accuracy for different number of states and features extracted from the IDS segment of the swallows

5.4.2 HMM-based Classification

Initial values of HMM parameters were set as explained in the previous section and the iterative EM algorithm was used to find the model parameters resulting in the maximum likelihood of the observation sequences.

When the feature sequences consisted of the features of the entire duration of swallowing sound signals, the accuracy of classification by different HMMs hardly

reached 60%. However, when the sequences of the features of the IDS segments were modeled, the accuracy could reach up to 85.5%. Classification accuracy based on HMMs of the features extracted from the IDS portion of swallowing sound signals are shown in Fig. 5.12. As it can be seen, the accuracy of classification when the HMM was used to classify the sequence of MSP3 or P_{ave} of swallowing sound signals were much lower than that of HMMs with other features. Overall, the classification accuracies when RMS or WFD features were used were better than the classification accuracies when other features were used.

Furthermore, when the number of states was increased, the classification error changed. By increasing the number of states from 4 to 8, the classification error gradually decreased. However, in most cases the error for N=9 was higher than that of N=8.

As a second stage classification, a screening algorithm was used, in which if more than 50% of the swallows of a subject were classified as being normal, the subject was considered as normal. Otherwise, the subject was considered as at risk of dysphagia. This second-stage screening correctly classified all the subjects but one healthy subject when RMS was used as a characteristic feature of the swallowing sounds and number of states was set to N=8. When WFD or MSP1 was used in HMM-based classification, the number of misclassified raised to two.

5.5 Timing Analysis of Submental EMG

The results of detected swallowing sound states by HMM and the corresponding submental muscle activities for a typical swallow when the subject was fed juice are shown in Fig. 5.13. As can be seen the muscle activity proceeds well before the onset of

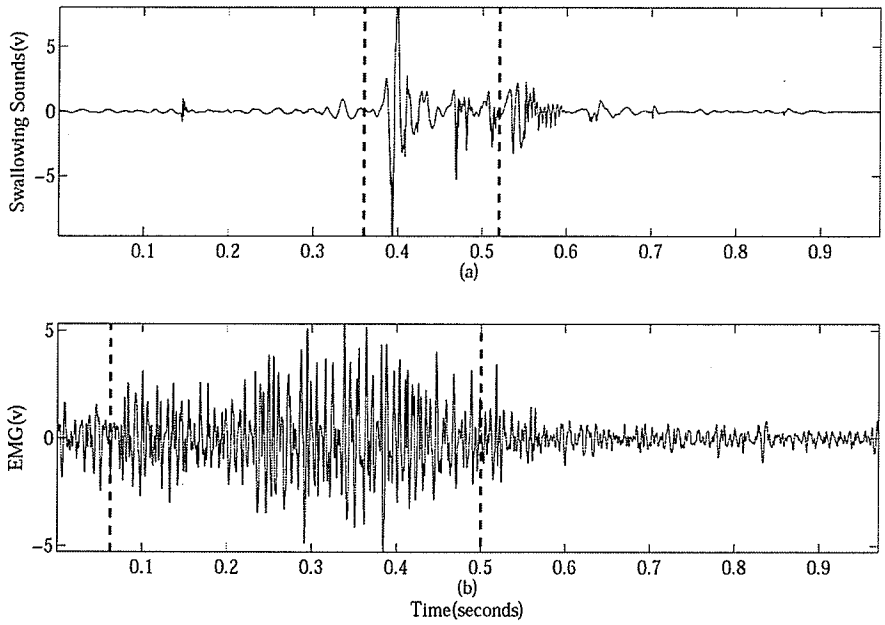


Fig. 5.13. (a) The waveform of a swallowing sound signal; (b) an EMG signal. The dashed vertical lines show the boundaries of IDS segment of the swallowing sounds and muscle contraction period of EMG.

esophageal phase of swallow. The temporal pattern showed in Fig. 5.13 was common in all swallowing sounds and EMG signals.

Table 5.5 shows the mean and standard deviation values of the duration of submental muscle contraction, and time difference between the onset of submental muscle

Table 5.5

Duration of swallowing sound phases and submental muscle contraction in different textures in seconds (Mean ± standard deviation)

	Semisolid	Liquid
Duration of muscle contraction	0.48±0.09	0.58±0.08
Time difference between the onset of muscle contraction and the opening of cricopharynx	0.25±0.12	0.19±0.09

contraction and the opening of cricopharynx when subjects were fed thin liquid and semisolid textures. The values were averaged among all subjects.

The results of this pilot study showed that the movement of the larynx precedes the IDS segment of swallowing sounds. It also showed that the lag between the onset of submental muscle contraction and the beginning of IDS segments of swallowing sounds is shorter for thin liquid texture. Since we did not have EMG data for the dysphagic group, we could not test any difference between the two groups in this topic. However, it can be speculated that the lag between the onset of submental muscle activity and the pharyngeal phase of swallow might be longer in dysphagic patients; this has the potential to be investigated further.

CHAPTER 6

CONCLUSIONS

6.1 Concluding Remarks

This thesis pursued five major goals: automated detection of swallowing sounds in the tracheal swallow and breath sound recordings, segmentation of swallowing sound signals, characterization of swallowing sounds using nonlinear dynamics, classification of swallowing sounds of healthy subjects and dysphagic patients, and analysis of the timing of swallowing sounds and EMG.

6.1.1 Swallowing Sound Detection

Three methods were used for detection of swallowing sounds in the tracheal swallow and breath sound recordings and their results were compared. First, since some of the recurrence plot features were observed to be appreciably different in swallowing and breath sounds, thresholding of three recurrence plot features, i.e., the %REC, ratio of %DET to %REC, and Shannon entropy of the length of diagonal lines was used for swallowing sound detection. Second, hidden Markov modeling was used for detection of swallowing sounds using the above three features. Third, a discrete wavelet transform based filter was used for swallowing sound detection. Multiresolution analysis of tracheal sound was the basis of the wavelet transform based filter. In the filtering method used in

this study, iterative sequences of multiresolution decomposition, thresholding of wavelet coefficients and multiresolution reconstruction were employed.

The results of the above mentioned methods for swallowing sound detection were compared in terms of false positive and false negative errors. The percentages of false positives in HMM-based swallow detection methods were much lower than those of thresholding-based and wavelet-based methods. However, percentages of missed swallows (false negative error) in HMM-based methods were slightly higher than those of thresholding-based methods. In addition, the mean and standard deviation of delay in detection of swallowing end point when HMM was used was noticeably lower than that of thresholding-based and wavelet-based methods. Overall, the performance of the HMM-based method using the recurrence plot features was much better than the other two methods. However, due to computational complexity of hidden Markov modeling of recurrence plot features, one may chooses computational efficient and less accurate wavelet based method.

The performance of the only previously proposed method for swallowing sound detection [Lazareck and Moussavi, 2002a] was reported only in terms of percentage of missed swallows. That method yielded 21% error which is higher than that of methods investigated in this study.

6.1.2 Swallowing Sound Segmentation

The ability of HMM in detection of boundaries of swallowing sound phases was investigated using several different features (waveform fractal dimension (WFD), linear prediction coefficients (LPC), average power (P_{ave}) of the signal over different frequency

bands, RMS of the time domain signal and multiscale products (MSP) of the wavelet coefficients). The time difference between the detected boundaries with the above HMM filters were compared with those determined manually by an expert. Among the features tested with an HMM filter, MSP resulted in the least error for detecting the boundaries of the IDS and BTS segments, while LPC resulted in the largest error. This was expected as multiscale product signifies the singular points of a signal, hence resulting in better results in detecting IDS segments as they are characterized by discrete click sounds. There was no quantative report on the previously proposed methods for swallowing sound segmentation in the literature.

6.1.3 Swallowing Sound Characterization

One of the goals of this thesis was to investigate the application of nonlinear analysis to swallowing sounds as opposed to linear analysis, which had been the only tool applied to swallowing sounds in the past. It was shown that swallowing sounds are well characterized by a small number of dimensions. In addition, the largest Lyapunov exponent was estimated to evaluate the presence of chaos. The estimated largest Lyapunov exponent depended on the subject and bolus textures. The inconsistency in the estimated largest Lyapunov exponent even among different swallows of a particular subject was due to the fact that the recorded swallowing sound signals were short time signals contaminated with noise.

Additionally, the characteristics of tracheal breath and swallowing sounds using recurrence statistics were studied. Comparison of the %REC, ratio of %DET to %REC, and Shannon entropy of the length of diagonal lines of the swallowing and breath sounds

recurrence plots showed that these parameters are appreciably different in swallowing and breath sounds; hence they may be used for automatic swallowing sound detection in the breath and swallow sound recordings.

6.1.4 Swallowing Sound Classification

Classification of swallowing sounds of normal and dysphagic subjects confirmed the discriminatory ability of the reconstructed phase space features for detecting the plausible abnormalities in the swallowing process of patients at risk of dysphagia.

The results of HMM-based methods for classification of swallowing sounds between the two groups of healthy and dysphagic subjects were not as promising as the results of speech recognition reported in the literature; HMM is known to achieve high accuracy for speech recognition. However, it should be noted that swallowing sounds are not vocal sounds, and therefore are substantially different from speech signals.

Classification accuracy of different HMM configurations were investigated. When we increased the number of states of the HMM from 4 to 8, the classification error gradually decreased, although in most cases the error for N=9 was higher than that of N=8. Among 7 features used, RMS and WFD showed the best performance in the HMM-based classification of swallowing sounds.

6.1.5 Timing Analysis of Submental EMG

A pilot study on the timing of submental muscle contraction in relation to swallowing sounds was conducted. The results of this pilot study showed that the movement of the larynx precedes the IDS segment of swallowing sounds. It also showed that the lag

between the onset of submental muscle contraction and the beginning of IDS segment of swallowing sounds is shorter for thin liquid texture. Since we did not have EMG data for the dysphagic group, we could not test any difference between the two groups in this topic. However, it can be speculated that the lag between the onset of submental muscle activity and the pharyngeal phase of swallow might be longer in dysphagic patients; this has the potential to be investigated further.

6.2 Recommendations for Future Works

Although the results of the analysis of swallowing sounds based on nonlinear dynamics theory are very encouraging, they were based on the traditional approach for dynamical system analysis, in which the behavior of the dynamical system is assumed to be deterministic, i.e., uniquely determined by its time evolution rules and initial conditions. However, when the Lipschitz continuity condition for uniqueness of the trajectory is violated, this approach is not valid and the problem becomes a non-deterministic chaos problem. At singular points of a trajectory, the Lipschitz continuity condition for uniqueness of the trajectory is violated and various solutions move away from the singular point. When a trajectory approaches a singular point, a small noise will put the trajectory on a different solution. One of the features of the systems that violate the Lipschitz continuity is the abrupt change of the trajectory from one solution to another. Since a swallowing sound signal has several click sounds that can be characterized as singular points, detection of singularities could be helpful in different applications.

One limitation for the analysis of singularities is that it is customary to select the

sampling rate based on the Nyquist theorem. The Nyquist theorem is based on linear systems theory. Based on the Nyquist theorem, the sampling frequency must be more than twice the highest waveform frequency to ensure that all Fourier components of a periodic waveform are recoverable. Additionally, the signals are filtered before the A/D conversion to avoid aliasing. If a discontinuity were sampled, it is very likely to be missed, because the sampling rate might be insufficient or the anti-aliasing filter may distort it. Therefore, selection of a higher sampling frequency would be useful.

Lastly, studying the temporal pattern of EMG activities of the submental muscles involved in the swallowing process with a much higher sampling rate may show some discriminatory differences between normal and dysphagic swallows. As coordination of swallowing mechanism structures is a key factor in a safe feeding, study on the timing of dysphagic swallow in comparison with that of normal swallow may shed some lights on the cause of an abnormality in the swallowing mechanism. However, the EMG activities are too gross for measurement of fine movement of the structures involved in a swallowing process. Therefore, it is suggested to have VFS data with higher resolution than 30 frames along with EMG activities and sounds recorded with a high sampling rate simultaneously.

APPENDIX A

THRESHOLD SELECTION USING THE OTSU METHOD

Otsu method has been proposed for thresholding of gray level images to separate objects from background; however, it is also applicable in thresholding of signals and has been used for apnea detection by acoustical means [Yadollahi and Moussavi, 2006b]. Suppose the points of a given time series, with a total number of N points, are represented in L quantized levels. The number of points in level i is denoted by n_i. Therefore, the probability of each level i is given by:

$$p_i = \frac{n_i}{N}\ ,\ i = 1,2,...,L. \tag{A.1}$$

With a threshold T, the levels are divided into two distinct classes: class C_1 contains the points at levels $i = 1,2,...,T$, and class C_2 contains the points at levels $i = T+1,...,L$. The probability of the classes C_1 and C_2 are denoted by P_1 and P_2 and given by:

$$P_1 = \frac{N_1}{N}, \tag{A.2}$$

and

$$P_2 = \frac{N_2}{N}, \tag{A.3}$$

where N_1 and N_2 are the number of points in the classes C_1 and C_2, respectively. When the threshold is changed, the variance of one class is increased while the variance of the other class is decreased. The goal of Otsu method is to select the threshold that minimizes

the weighted variance, the so-called within-class variance:

$$\sigma^2_{within} = P_1\sigma_1^2 + P_2\sigma_2^2, \tag{A.4}$$

where σ_1^2 and σ_2^2 are the variance of the classes C_1 and C_2, respectively.

In order to increase computational efficiency, instead of finding the threshold that minimizes σ^2_{within}, it is very common to find the threshold that maximizes the so called between class variance $\sigma^2_{between}$:

$$\sigma^2_{between} = \sigma^2 - \sigma^2_{within}, \tag{A.5}$$

where σ^2 is the variance of the entire data. It can be easily shown that

$$\sigma^2_{between} = P_1(\mu_1 - \mu)^2 + P_2(\mu_2 - \mu)^2, \tag{A.6}$$

where μ, μ_1, and μ_2 are the mean of entire data, class C_1, and class C_2, respectively. Furthermore, Equation (A.6) can be simplified as

$$\sigma^2_{between} = P_1P_2(\mu_1 - \mu_2)^2. \tag{A.7}$$

In summay, in order to find a threshold that minimizes within-class variance, the threshold that maximizes between-class variance must be found.

REFERENCES

H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, "The analysis of observed chaotic data in physical systems," *Reviews of Modern Physics*, vol. 65, no. 4, pp. 1331-1392, 1993.

C. Ahlstrom, A. Johansson, P. Hult, and P. Ask, "Chaotic dynamics of respiratory sounds," *Journal of Chaos, Solitons and Fractals*, vol. 29, pp. 1054-1062, 2006.

M. Boiron, P. Rouleau, and E. H. Metman, "Exploration of pharyngeal swallowing by audiosignal recording," *Dysphagia*, vol. 12, pp.86-92, 1997.

J. G. Brasseur, W. J. Dodds, "Interpretation of intraluminal manometric measurement in terms of swallowing mechanics," *Dysphagia*, vol. 6, pp. 100-119, 1991.

J. A. Y. Cichero, B. E. Murdoch, "The physiologic cause of swallowing sounds: answers from heart sounds and vocal tract acoustics," *Dysphagia*, vol. 13, pp. 39-52, 1998.

J. A. Y. Cichero, B. E. Murdoch, "Detection of swallowing sounds: methodology revisited," *Dysphagia*, vol. 17, pp. 40-49, 2002.

M. Ding, C. Grebogi, E. Ott, T. Sauer, and J. A. Yorke, "Estimating correlation dimension from a chaotic time series: when does plateau onset occur?" *Physica D*, vol. 69, pp. 404-424, 1993.

R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, New York, John Wiley & Sons, 2001.

J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters*, vol. 4, no. 9, pp. 973-977, 1987.

R. Esteller, G. Vachetsevanos, J. Echauz, and B. Litt, "A comparison of waveform fractal dimension algorithms," *IEEE Transactions on Circuits and Systems,* vol. 48, no.2, pp. 177-183, 2001.

A. M. Fraser, H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, pp. 1134-1140, 1986.

J. B. Gao, "Recurrence time statistics for chaotic systems and their applications," *Physical Review Letters*, vol. 83, no. 16, pp. 3178-3181, 1999.

J. Gao and H. Cai, "On the structures and quantification of recurrence plots," *Physics Letters A*, vol. 270, pp. 75-87, 2000.

J. B. Gao, "Detecting nonstationarity and state transitions in a time series," *Physical Review E,* vol. 63, pp.1- 8, 2001.

J. B. Gao, Y. Cao, L. Gu, J. G. Harris, and J. C. Principe, "Detection of weak transitions in signal dynamics using recurrence time statistics," *Physics Letters A*, vol. 317, pp. 64-72, 2003.

J. Gnitecki, Z. Moussavi, "The fractality of lung sounds: A comparison of three waveform fractal dimension algorithms," *Journal of Chaos, Solitons and Fractals*, vol. 26, no. 4, pp. 1065-1072, 2005.

B. Gold and N. Morgan, *Speech and Audio Signal Processing*, John Wiley & Sons, New York, NY, 1999.

H. Gray, *Anatomy of Human Body*, Edited by C.D. Clemente, Philadelphia, PA, Lea & Febiger, 1985.

P. Grassberger, and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D*, vol. 9, pp. 189-208, 1983.

P. Grassberger, T. Schreiber, and C. Schaffrath, " Nonlinear time sequence analysis," *Inernational Journal of bifurcation chaos*, vol.1, no. 3, pp. 521-547, 1991.

L. J. Hadjileontiadis, and S. V. Panas, "Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter," *IEEE Transactions on Biomedical Engineering*, vol. 44, pp. 1269-1283, 1997.

L. J. Hadjileontiadis, C. N. Liastos, C. C. Mavrogiannis, T. A. Rokkas, and S. M. Panas, "Enhancement of bowel sounds by wavelet-based filter," *IEEE Transactions on Biomedical Engineering*, vol. 47, pp. 876-886, 2000.

S. L. Hamlet, R. L. Patterson, S. M. Fleming, and L. A. Jones, "Sounds of swallowing following total laryngectomy," *Dysphagia*, vol. 7, no. 1, pp. 160-165, 1992.

S. Hamlet, D. Penney, and J. Formolo, "Stethoscope acoustics and cervical auscultation of swallowing," *Dysphagia*, vol. 9, no. 1, pp. 63-68, 1994.

R. C. Hilborn, *Chaos and Nonlinear Dynamics*, 2nd edition, Oxford university press, 2000.

P. J. Kahrilas and J. A. Logeman, "Volume accommodations during swallowing," *Dysphagia*, vol. 8, pp. 259-265, 1993.

M. J. Katz, "Fractals and the analysis of the waveforms," *Computers in Biology and Medicine*, vol.18, no.3, pp.145-156, 1988.

M. B. Kennel, R. Brown, H. D. I. Abarbanel, "Determining embedding dimensions for phase space reconstruction using a geometric construction," *Physical Review A*, vol. 45, no. 6, pp. 3403-3411, 1992.

M. S. Klahn and A. L. Perlman, "Temporal and durational patterns associating respiration and swallowing," *Dysphagia*, vol. 13, pp. 131-138, 1999.

R. J. Last, *Anatomy, Regional and applied*, 6th Ed. Edinburgh, UK, Churchill Livingstone, 1978.

L. J. Lazareck and Z. K. Moussavi, "Automated algorithm for swallowing sound detection," *Proceeding of Canadian Medical and Biological Engineering Conference,* 2002a.

L. Lazareck and Z. Moussavi, "Adaptive swallowing sound segmentation by variance dimension," *Proceedings of European Medical and Biological Engineering Conference (EMBEC),* 2002b.

L. Lazareck and Z. Moussavi, "Classification of Normal and Dysphagic Swallows by Acoustical Means," *IEEE Transactions on Biomedical Engineering,* vol. 51, pp. 2103-2112, 2004.

J. A. Logemann, *Evaluation and Treatment of Swallowing Disorders*, Austin, TX, PRO-

ED, 1998.

S. G. Mallat, *A Wavelet Tour of Signal Processing.* San Diego, CA, Academic Press, 1998.

J. M. Martinerie, A. M. Albano, A. I. Mees, and P. E. Rapp, "Mutual information, strange attractors, and the optimal estimation of dimension," Physical *Review A*, vol. 45, pp. 7058-7064, 1992.

N. Marwan, N. Wessel, U. Meyerfeldt, A. schirdewan, and J. Kurths, "Recurrence-plot based measures of complexity and their application to heart rate variability data," *Physical Review E*, vol. 66, pp. 1-8, 2002.

H. Mingyou, H. Xie, T. B. Tiong, and X. Wu, "Study on a spatially selective noise filtration technique for suppressing noises in partial discharge on-line monitoring," *Proceeding of the Conference on properties and Application of Dielectric Materials*, pp. 689-692, 2000.

Z. K. Moussavi, M. T. Leopando, H. Pasterkamp, and G. Rempel, "Computerized acoustical respiratory phase detection without airflow measurement," *Journal of Medical & Biological Engineering and Computing,* vol. 38, no. 2, pp. 198-203, 2000.

Z. Moussavi, D. Flores, and G. Thomas, "Heart sound cancellation based on multiscale

products and linear prediction," *Proceedings of IEEE Engineering in Medicine and Biology Conference*, pp. 3840-3843, 2004.

N. Otsu, "A Threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 9, No. 1, pp. 62-66, 1979.

N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical Review Letters*, vol. 45, pp.712-716, 1980.

J. B. Palmer, "Electromyography of the muscles of oropharyngeal swallowing: Basic concepts," *Dysphagia*, vol. 3, pp.192-198, 1988.

J. B. Palmer, K. V. Kuhlemeier, D. C. Tippett, and C. Lynch, "A protocol for the videofluorographic swallowing study," *Dysphagia*, vol. 8, no. 3, pp. 209-214, 1993.

T. S. Parker and L. O. Chua, "Chaos: a tutorial for engineers," *Proceedings of IEEE*, vol. 75, no. 8, pp. 982-1008, 1987.

A. L. Perlman, E. S. Luschei, C. E. DuMond, "Electrical activity from the superior pharyngeal constrictor during reflexive and non-reflexive tasks," *Journal of Speech and Hearing Research* , vol. 32, pp. 749-754, 1989.

J. G. Proakis, D.G. Manolakis, *Digital Signal Processing*, Upper saddle River, NJ, Prentice-Hall, 2006.

L. R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-285, February 1989.

L. R. Rabiner, and H. H. Huang, *Fundamental of speech processing*, prentice-Hall, Englewood Cliffs, NJ, 1993.

N. P. Reddy, B. R. Costarella, R. C. Grotz, and E. P. Canilang, "Biomedical measurement to characterize the oral phase of dysphagia," *IEEE Transactions on Biomedical Engineering*, pp. 392-397, vol. 37, no. 4 , 1990.

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.

M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica D*, vol. 73, pp. 82-98, 1994.

M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest Lyapunov exponents," *Physica D*, vol. 65, pp. 117-134, 1993.

T. Shawker, B. Sonies, T. Hall, and G. Baum, "Ultrasound analysis of tongue, hyoid and larynx activity during swallowing," *Investigative Radiology*, vol. 19, pp. 82-86, 1984.

W. G. Selley, F. C. Flack, and R. E. Ellis, "The Exeter dysphagia assessment technique," *Dysphagia,* vol. 4, pp. 227-235, 1990.

D. Smith, S. Hamlet, and L. Jones, "Acoustic Technique for determining timing of velopharyngeal closure in swallowing," *Dysphagia,* vol. 5, pp. 142-146, 1990.

B. M. Stadler and A. Swami, "Analysis of multiscale products for step detection and estimation," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 1043-1051, 1999.

S. H. strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*, Westview Press, 2000.

K. Takahasi, M. E. Groher, and K. Michi, "Symmetry and reproducibility of swallowing sounds," *Dysphagia*, vol. 9, pp.168-173, 1994.

F. Takens, "Detecting strange attractors in turbulence," in: D. A. Rand and L. S. Young, (Eds.), *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, Springer, vol. 898, pp. 366 - 381, 1981.

F. L. Vice, J. M. Heinz, G. Giuriati, M. Hood, and J. F. Bosma, "Cervical auscultation of suckle feeding in newborn infants," *Developmental Medicine and Child Neurology*, vol. 32, pp. 760-768, 1990.

A. J. Viterbi, "Error bounds for deconvolutional codes and an asymptotically optimal decoding algorithm," *IEEE Transactions on Information Theory*, vol. IT-13, pp. 260-269, 1967.

C. L. Webber Jr. and J. P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *Journal of Applied Physiology*, vol. 76, no. 2, pp. 965–973, 1994.

A. Wolf, J. B. Swift, H. L. Swinney and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D*, vol. 16, pp. 285- 317, 1985.

Y. Xu, J. B. Weaver, D. M. Hearly Jr, and J. Lu, "Wavelet transform domain filters: A spatially selective noise filtration technique," *IEEE Transactions on Image Processing*, vol. 3, no. 6, pp. 747-758, 1994.

A. Yadollahi, Z. Moussavi, "A robust method for estimating respiratory flow using tracheal sounds entropy," *IEEE Transaction on Biomedical Engineering*, vol. 53 , no. 4, pp. 662-668, 2006a.

A. Yadollahi and Z. Moussavi, "Apnea detection by acoustical means," *Proceedings of IEEE Engineering in Medicine and Biology Conference,* 2006b.

J. P. Zbilut and C. L. Webber Jr., "Embeddings and delays as derived from quantification of recurrence plots," *Physics Letters A*, vol. 171, no. 3–4, pp. 199–203, 1992.

J. P. Zbilut, J. Zaldivar-Comenges, and F. Strozzi, "Recurrence quantification based Lyapunov exponents for monitoring divergence in experimental data," *Physics Letters A*, vol. 297, pp. 173-181, 2002a.

J. P. Zbilut, D. D. Dixon, and M. Zak, " Detecting singularities of piecewise deterministic (terminal) dynamics in experimental data, " *Physics Letters A*, vol. 304, pp. 95-101, 2002b.