DEVELOPMENT OF AN ACOUSTIC RESPIRATORY MONITOR

by

Bryce Ensign Hill

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

The University of Utah

August 2011

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Bryce Ensign Hill**

has been approved by the following supervisory committee members:

Dwayne R. Westenskow	, Chair	**05/26/11** Date Approved
Douglas A. Christensen	, Co-Chair	**05/31/11** Date Approved
Joseph A. Orr	, Member	**05/26/11** Date Approved
Neal Patwari	, Member	**05/31/11** Date Approved
Neil E. Cotter	, Member	**05/31/11** Date Approved

and by **Gianluca Lazzi** , Chair of

the Department of **Electrical and Computer Engineering**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

Patients sometimes suffer apnea during sedation procedures or after general anesthesia. Apnea presents itself in two forms: respiratory depression (RD) and respiratory obstruction (RO). During RD the patients' airway is open but they lose the drive to breathe. During RO the patients' airway is occluded while they try to breathe. Patients' respiration is rarely monitored directly, but in a few cases is monitored with a capnometer. This dissertation explores the feasibility of monitoring respiration indirectly using an acoustic sensor. In addition to detecting apnea in general, this technique has the possibility of differentiating between RD and RO. Data were recorded on 24 subjects as they underwent sedation. During the sedation, subjects experienced RD or RO.

The first part of this dissertation involved detecting periods of apnea from the recorded acoustic data. A method using a parameter estimation algorithm to determine the variance of the noise of the audio signal was developed, and the envelope of the audio data was used to determine when the subject had stopped breathing. Periods of apnea detected by the acoustic method were compared to the periods of apnea detected by the direct flow measurement. This succeeded with 91.8% sensitivity and 92.8% specificity in the training set and 100% sensitivity and 98% specificity in the testing set.

The second part of this dissertation used the periods during which apnea was detected to determine if the subject was experiencing RD or RO. The classifications determined from the acoustic signal were compared to the classifications based on the

flow measurement in conjunction with the chest and abdomen movements. This did not succeed with a 86.9% sensitivity and 52.6% specificity in the training set, and 100% sensitivity and 0% specificity in the testing set.

The third part of this project developed a method to reduce the background sounds that were commonly recorded on the microphone. Additive noise was created to simulate noise generated in typical settings and the noise was removed via an adaptive filter. This succeeded in improving or maintaining apnea detection given the different types of sounds added to the breathing data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Without the help of a number of people this work would have been impossible. First and foremost I would like to acknowledge and thank my wife Havilah. She has supported me throughout my college career and continues to support me in my endeavors. During the past four years she has given birth to three children and has put up with my service in our community and extended hours spent in the lab.

I express my gratitude to the members of my graduate committee who have helped me a great deal in understanding what it means to earn the Doctorate of Philosophy in Electrical Engineering. Dr. Westenskow has acquired funding for my research for the past four years and been the foundation for my understanding of human physiology. Dr. Orr has been the most available to listen to my ideas and work with me on understanding ideas. He was there for feedback of ideas and helping to inspire new ideas. Dr. Christensen has taught me how to do science, how to write technically, and how to be critical of my own work. Dr. Patwari has not only taught me during several courses, but has also taught me individually on his own time. My entire understanding of statistical analysis is based on his teaching. Dr. Cotter has been the voice of reason and the needed critic of this work. Without his help this work would not be as valuable.

I acknowledge my grandfather, Armin John Hill. He has served as an inspiration throughout my life. In 1932 he received a bachelors degree from the Montana State College in Bozeman Montana, now called Montana State University where I graduated

with my bachelors degree. He later returned to Montana State College and received his masters degree in electrical engineering 1938. In June 1949 he received a second masters degree in optics from CalTech and finally received his Ph.D in spectroscopy in 1950. He has served as the Dean in the College of Science and Engineering and the Dean of Engineering and Technology at BYU. He has served as an example of what I can accomplish throughout my life.

I thank my father, Walter Ensign Hill. He received his bachelors degree from BYU in 1961. He later received his master's degree from the University of Wisconsin in 1964 and then his PhD in biophysics in June 1967. For the past 40 years he has served as a professor at the University of Montana in biochemistry. He has always told me that I am smart enough to get a PhD. He has mentored me throughout my primary education and college career and has given me hope through the harder parts of my research and schooling. He always encouraged my creativity and curiosity even when it cost him tools that were lost in the backyard.

My mother, Annette Smith Hill, told me that I was special from a very early age. She made it a point to make sure that I felt like I was important and smart. For thirty years she packed lunches for all of her children without fail. Every morning she would provide breakfast and every evening she would have dinner ready. She was always there when we got home from school and gave me the encouragement that I needed. Now she is not only a mother to myself and my siblings but a grandmother to 30 grandchildren. She continues to be a loving encouragement to her posterity.

For the past four years I have been supported financially by the Rocky Mountain NASA Space Grant Consortium. I express my appreciation to those who maintain this

funding and have made it possible to receive my education.

The study performed from which the data was collected was funded by Scott Laboratories Inc. from Lubbock Texas. I thank them for the use of the data collected during this study.

Lastly I acknowledge my friends and associates in the Anesthesia Bioengineering Laboratory at the University of Utah. There are too many to name individually for fear that I might miss one. Many of them have helped me develop ideas and have served as valuable references. In most cases they have helped me keep my sanity during the more difficult times. Their friendship and support have helped to shape my research and my personality. Some have gone above and beyond the call of duty helping with personal projects that have taken many days and hours of time. Thank you to all of you who have been there.

CHAPTER 1

INTRODUCTION

1.1 Patient Safety in Sedated Care

Patients often undergo painful procedures that require sedation. When such procedures are performed in the physician's office, adverse events occur at a rate of 66 per 100,000 procedures as compared to 5.3 per 100,000 procedures in an ambulatory surgery center. Death occurs at a rate of 9.2 per 100,000 procedures in the physician's office as compared to 0.78 per 100,000 procedures in an ambulatory surgery center [1]. Mortality rate at an ambulatory surgery center is comparable to mortality rate at an inpatient surgery center [2]. A significant source of the increase in adverse events and deaths is the significant difference in patient monitoring. In an operating room and ambulatory or inpatient surgery center, patients are intubated, allowing breath rate and volume to be monitored and controlled directly. Another significant difference is the absence of an anesthesiologist in the physician's office. Guidelines on what monitoring equipment and personnel are needed vary from state to state, but are still subject to ASA and OBA guidelines [3]. More specifically, Domino et al. [4] state that a significant number of adverse events and deaths could be prevented by improved monitoring in the physician's office.

The objective of this project is to develop a new respiratory monitor that will detect the respiratory function from the sound recorded at skin over the trachea. Sounds

recorded at the skin over the trachea will be processed to detect periods of apnea. In addition, an adaptive filter will be tested to remove any ambient sounds not related to breathing so that apnea can be detected reliably.

1.2 The Need for Better Respiratory Monitoring

1.2.1 Current respiratory monitors

Respiratory monitors used in the physician's office include pulse-oximetry, thoracic-impedance plethysmography, and capnometry. Direct respiratory flow measurements are not typically monitored because intubation or a tight-fitting facemask are not present in a procedure at the physician's office [5].

1.2.2 Pulse-oximetry

Pulse-oximetry measures the blood-oxygen saturation through a transcutaneous light measurement on an appendage. Pulse-oximetry is a very common and useful monitor because of its ability to continuously measure the blood-oxygen saturation and measure the patient's heart rate. Although it does measure the blood-oxygen saturation it cannot replace direct respiratory monitoring. A healthy well oxygenated patient, receiving oxygen via a facemask or nasal cannula, can stop breathing for 131 seconds before the oxygen saturation falls below 92% and the alarm of the pulse-oximeter sounds. This amount of time can be increased up to 215 seconds in cases where hypothermia and vasoconstriction are present [6]. This means that several minutes can pass from the time that the patient has stopped breathing to the time that the pulse-oximeter would show a noticeable change in his/her blood-oxygen saturation. During periods of apnea the oxygen saturation falls, and the CO_2 concentration in the blood increases. Excessive concentra-

tions of CO_2 can cause metabolic acidosis of the blood, which sets the patient at risk for cardiac arrhythmias [7-9].

1.2.3 Plethysmography

Thoracic-impedance plethysmography uses ECG electrodes on the chest to measure the change in thoracic impedance. Less common is respiratory inductance plethysmography (RIP), which is measured by finding the change of inductance in wires sewn into elastic bands that are placed on the chest and abdomen [10]. Both forms of plethysmography can measure the relative change in chest and abdomen volume. Plethysmography directly measures respiratory effort but can be corrupted by body movement of the patient. Plethysmography is a relative measurement. The plethysmography measurement can be calibrated to the change in lung volume measured by a respiratory flow meter if a respiratory flow measurement is present. Using this calibration, change in chest and abdomen volume can be estimated to be the change in lung volume [11]. Although this is a useful measurement, the calibration cannot be performed unless the patient is wearing a tight-fitting facemask, or is intubated.

If a direct airway flow measurement is available in addition to the plethysmography signal, together they can be used to determine whether periods of apnea are due to respiratory obstruction (RO) or respiratory depression (RD). In the case of RO the plethysmography signal would show chest and abdomen movement with no airway flow in the direct respiratory measurement. In the case of RD the plethysmography signal would show no significant change in volume while the direct respiratory measurement would show no significant flow. In the absence of a direct flow measurement, if a patient's airway becomes obstructed, but the patient continues to try to breathe, the meas-

ured change in the calibrated plethysmography signal might be mistakenly counted as a normal breath. This can show that the patient is breathing normally but no valid gas exchange occurs [12].

1.2.4 Capnometry

Capnometry measures the concentration of CO_2 in expired air. The gas sample is obtained through a nasal cannula placed in the nose. Capnometry is important because of its ability to give a relative respiratory rate noninvasively. Capnometry cannot reliably estimate breath volume based on the length of the expiratory period and CO_2 concentration of each breath. Capnometry respiratory rate estimation also relies on the assumption that the nasal passage is not occluded. Nasal cannula are also uncomfortable and dislodge easily. If a patient stops breathing for a short period of time, the CO_2 concentration within the lungs increases as the CO_2 concentration in the venous blood increases. The capnometer can be used to detect hypopnea because the $PaCO_2$ increases with decreased ventilation. Hypopnea is defined as a period of shallow breathing or a period of low respiratory rate.

1.3 Acoustic Respiratory Monitoring

1.3.1 Current use of respiratory sounds by physicians

Since the development of the stethoscope in the 19th century, anesthesiologists have relied on directly listening to the patient breathing through an esophageal stethoscope or precordial stethoscope [13]. Today most anesthesiologists use the pulse-oximeter or capnometer rather than the esophageal or precordial stethoscope to continuously monitor the respiratory activity of a patient [14]. Anesthesiologists listen to breath

sounds to detect esophageal intubation, endobronchial intubation, and airway obstruction. Listening to patients breathe can be laborious and is also not convenient for the anesthesiologist to be tethered to the patient during a procedure. For this reason breath sounds have been analyzed and modeled electronically using computer algorithms to detect respiratory rate and classify breath sounds [15-21]. This research has led to a recently FDA approved acoustic respiratory monitor developed by Masimo Corporation (Irvine, CA) [22]. Continued research is needed to improve upon current designs and technologies.

1.3.2 The source of respiratory sounds at the trachea

Breathing sounds are created as the flow within the trachea becomes turbulent. The turbulence occurs at the narrowest point in the respiratory path. The glottal opening is the primary source of breath sounds [23]. The critical flow for the tracheal turbulence happens when the tracheal flow rate exceeds 7.2 l/m, which corresponds to a tracheal Reynolds number of 1000 [23].

During normal breathing, airway flow is correlated with the amplitude of the sound [15,17,21]. This relationship is not linear, but it is possible to estimate flow from the amplitude of the breath sound [21]. The relationship of flow to sound amplitude has been modeled with the equation $A=kF^{3.0\pm0.2}$, where k is a constant, A is the sound amplitude in units of volts, and F is the flow rate in liters per second [16,24]. Beck and Olson et al. [24] did not give some of the units for this equation. A more detailed explanation is given in Chapter 2. Similarly, Kraman measured the relationship between tracheal flow and sound amplitude in a model of a canine respiratory airway [25]. Kraman concluded that the sound amplitude is directly proportional to the flow rate (l/s)

squared. Calibration of the audio amplitude to the flow rate is necessary in order to make a flow rate estimate using the audio signal. This is because the size of the sub-glottal opening near the trachea varies between patients [15], as well as the amplitude of the audio with respect to where the transducer is placed on the patient relative to the sound source.

1.3.3 Measurement of respiratory rate from normal tracheal sounds

Sierra et al. [19] estimated the respiratory breath rate from the acoustic sounds in both healthy patients and those with pulmonary disease. Sierra et al. used a piezo-electric film transducer placed on the neck to record breath sounds. The methodology in Sierra et al. involved estimating the respiratory rate from the autocorrelation function of the audio envelope. Sierra et al. performed this by taking a section of the audio envelope and calculating the autocorrelation of it. Sierra et al. then found the local maxima of the auto-correlated signal. The number of samples between each local maxima was used as the number of samples between each breath and was used to calculate a breath rate. This algorithm worked well for the healthy patients, with $R^2 = 0.99$ when compared to the respiratory rate measured by a pneumotachometer, which is the current gold standard. This value dropped to $R^2 = 0.87$ for patients with pulmonary disease. Major errors were found to be related to low signal-to-noise ratio and from sounds not related to breathing, both external sounds like talking and internal sounds like swallowing. Sierra et al. also compared a manually counted respiratory rate from the audio data to that of the gold standard by manually counting the breaths of the audio envelope. This method performed with similar results to the automatic detection algorithm. Although Sierra et al. were successful in measuring an accurate respiratory rate from the audio envelope, the

ability of this algorithm to detect apnea was not determined.

1.4 Classification of Snoring and Obstruction

1.4.1 Source of snoring and obstructive sounds

Snoring occurs when loose adipose tissue near the oropharynx vibrates and slaps against other tissue [26]. This generally only partially occludes the airway but in severe cases this turns into RO where the loose tissue completely occludes the airway [27,28]. RO that occurs in nonsedated patients as they sleep is called obstructive sleep apnea (OSA). OSA poses a serious risk to an estimated 24% of the American population [26]. Because of the prevalence of OSA, research has been done to automatically record and detect periods of OSA without the need of an all-night study called polysomnography [28-31]. Benumof et al. [26] state that patients undergoing medical procedures that are suspected of having OSA are at higher risk for life-threatening occurrences during perioperative periods. The detection of apnea and classification of apnea as RO during a sedation would be a valuable information to physicians performing the procedure.

1.4.2 Literature review of methods for snore sound classification

Hara et al. [29] attempt to decipher the difference between snore sounds generated by simple snorers versus those generated by patients with OSA. Hara et al. used a Multidimensional Voice Program (MDVP) using inputs of soft phonation index, noise harmonics ratio, and power ratio in an attempt to decipher between simple snore sounds and obstructive snore sounds. Hara et al. defined simple snore sounds as snores that still allow the subject to be normally ventilated. Hara et al. concluded that the most prominent difference between simple snore sounds and obstructive snore sounds was the range

in the acoustic spectrum observed in obstructive snore sounds. Obstructive snore sounds showed multiple harmonics in the spectrum up to 4000 Hz, while simple snore sounds were observed to have spikes in the spectrum under 1000 Hz.

Jane et al. [28] used a multilayer neural network to decipher between obstructive snore sounds and simple snore sounds. This neural network received 22 temporal and spectral features from each sound segment as inputs. Jane et al. used a large database of sounds consisting of 625 pre-selected events. Using this strategy, the neural network was able to provide an 82% sensitivity and 90% specificity of OSA snore detection.

Nakano et al. [30] provided a different approach to classifying periods of OSA. Nakano et al. used the sounds recorded at the trachea to calculate the power spectral density (PSD) per Hz. This produced a single value of the summed PSD spectrum weighted by the inverse valued frequency in a specified frequency band. Obstructive snoring was classified if the peak value of the PSD per Hz was over 70 dB per Hz within the frequency band of 100-300 Hz. Obstructive snoring was defined as snore sounds that preceded a period of RO. Nakano et al. classified periods of apnea into three categories: RO, obstructive hypopnea, and RD. Nakano et al. further defined the tracheal sound-respiratory disturbance index (TS-RDI) as the number of TS dips per hour of examination. A TS dip was as a drop of 12 dB in the moving average in the PSD per Hz value in the band of 400-600 Hz.

In Nakano et al., RO was detected if the PSD per Hz value was found to be in the snoring classification area and was followed by a sudden drop in PSD per Hz. Periods of hypopnea were identified when the PSD per Hz remained in the snoring range for extended periods of time with no drop into the normal range. Periods of RD were

defined as the PSD stayed in normal range, then dropped near zero. Periods of apnea were identified and classified by performing polysomnography on the data recorded. Nakano et al. compared the performance of the TS-RDI to the apnea-hypopnea (AHI) index given by polysymnography with a high correlation coefficient of R=0.93. Nakano then specified two values of the cutoff for the AHI of 5 and 15. The sensitivity and specificity for the AHI cutoff value of 5 were calculated at 93% and 67%, respectively, and similarly with the AHI cutoff value of 15 at 79% and 95% respectively.

Ng et al. [31] used psychoacoustics of snore sounds to screen for OSA. The psychoacoustics were qualitative characteristics that were quantized using various equations. The qualitative characteristics included loudness, sharpness, roughness, fluctuation strength and annoyance. Each quality was used independently to generate a receiver operating characteristic (ROC) plot. The most successful features were loudness, annoyance, and roughness with sensitivities of 72-78% and specificities of 82-92%. The work of Ng et al. could be improved by using the combination of two or more of these qualities to improve sensitivity and specificity. Ng et al. also used a subjective method to select snore sounds before the psychoacoustic analysis was performed, rather than using the psychoacoustic data to classify a sound as a snore. Although equations for each psychoacoustic quality were given, the use of human classifications were used adding subjectivity to the data classification process.

The research focused on sleep-disordered breathing above is similar to the sleep-disordered breathing that occurs during patient sedation. Many principles are common for both situations, but sedation-induced apnea carries more risk because the patient wakes up periodically during sleep-disordered breathing to breath but is unable to do this

during a sedation procedure. The mechanisms that stop the breathing are similar though, and so the research regarding sleep disorders can be applied to sedative apnea in certain circumstances.

Although these works support the theory that breathing sounds can be used to detect breathing and apnea, a comprehensive method to detect breaths and define the source of apnea has not been performed. This has also never been done for sedated patients rather than data collected from sleep centers. This research covers data collected from subjects who were sedated and attempts to uncover the validity of novel approaches to apnea detection and classification.

1.5 Sound Detection

Detecting signals in noisy environments is an important task in signal detection theory. Gazor et al. [32] attempt to detect voiced conversations in a noisy environment. This is referred to as a voice activity detector (VAD). The VAD modeled the background noise as the sum of multiple Gaussian distributed signals with zero mean, which is equal to a single Gaussian distributed signal with zero mean. The speech portion was modeled as a Laplace distribution uncorrelated with the background noise. Gazor et al. used a maximum likelihood (ML) estimator for determining both the standard deviation of the combined Gaussian signals and the standard deviation of the Laplace signal. Gazor et al. then used a hypothesis test to determine if speech is present during that period. Gazor et al. had a goal of creating a reliable VAD with low computational complexity. Gazor's et al. method performed better than a VAD outlined in [33] and was lower in complexity.

Chapter 2 describes a method that uses a similar model to find the standard deviation of the background noise and the standard deviation of the desired signal.

Instead of using an ML estimator to find the parameters, a parameter estimation algorithm will be implemented.

Chapter 3 describes the methods used that attempt to classify breath sounds in order to classify apnea as RD or RO when compared to a standard. A standard to classify apnea as RD or RO has been developed that is based on the flow rate and the chest and abdomen movement of the subjects.

Chapter 4 describes the methods used to improve the sensitivity of the apnea detection algorithm described in Chapter 2. This was performed by testing the validity of using an adaptive filter to remove ambient sounds recorded by the microphone at the trachea.

Chapter 5 summarizes the results found in Chapters 2, 3, and 4 and discusses the success of algorithms that were developed. It also suggests directions for future work.

1.6 Project Objectives

This work will show the development of novel approaches to breath detection and apnea detection. It will also attempt to classify periods of apnea detected by the apnea detection algorithm. It will do this using a data set collected from volunteers for a sedative sleep study.

The goal is to develop a novel respiratory monitor that overcomes the shortcomings of the other monitors. It should be able to detect apnea sooner than a pulse-oximeter, be easier to use than plethysmography, and not need a tight-fitting facemask in the case of the direct flow measurement. In addition, it should be able to detect the difference between RO and RD, which none of the mentioned monitors can do.

1.7 References

[1] H. Vila, R. Soto, A.B. Cantor, and D. Mackey, “Comparative outcomes analysis of procedures performed in physician offices and ambulatory surgery centers.,” *Arch Surg*, vol. 138, Oct. 2003, pp. 991-995.

[2] D.C. Morello, G.A. Colon, S. Fredricks, R.E. Iverson, and R. Singer, “Patient safety in accredited office surgical facilities.,” *Plast Reconstr Surg*, vol. 99, May. 1997, pp. 1496-1500.

[3] D.H. Perrott, “Anesthesia outside the operating room in the office-based setting.,” *Curr Opin Anaesthesiol*, vol. 21, 2008, pp. 480-485.

[4] K. Domino, “Office-based anesthesia: lessons learned from the dlosed dlaims project.,” *ASA Newsletter*, vol. 65, Jun. 2001, pp. 9-11.

[5] J. Nagler and B. Krauss, “Capnography: a valuable tool for airway management,” *Emergency Medicine Clinics of North America*, vol. 26, Nov. 2008, pp. 881-897, vii.

[6] D.B. MacLeod, L.I. Cortinez, J.C. Keifer, D. Cameron, D.R. Wright, W.D. White, E.W. Moretti, L.R. Radulescu, and J. Somma, “The desaturation response time of finger pulse oximeters during mild hypothermia.,” *Anaesthesia*, vol. 60, Jan. 2005, pp. 65-71.

[7] H.L. Price, A.A. Lurie, R.E. Jones, M.L. Price, and H.W. Linde, “Cyclopropane anesthesia. II. epinephrine and norepinephrine in initiation of ventricular arrhythmias by carbon dioxide inhalation,” *Anesthesiology*, vol. 19, 1958.

[8] H.L. Price, “Effects of carbon dioxide on the cardiovascular system,” *Anesthesiology*, vol. 21, Dec. 1960, pp. 652-663.

[9] S. Lim, “Metabolic acidosis,” *Acta Medica Indonesiana*, vol. 39, Sep. 2007, pp. 145-150.

[10] H.L. Watson, D.A. Poole, and M.A. Sackner, “Accuracy of respiratory inductive plethysmographic cross-sectional areas.,” *J Appl Physiol*, vol. 65, Jul. 1988, pp. 306-308.

[11] M.A. Sackner, H. Watson, A.S. Belsito, D. Feinerman, M. Suarez, G. Gonzalez, F. Bizousky, and B. Krieger, “Calibration of respiratory inductive plethysmograph during natural breathing.,” *J Appl Physiol*, vol. 66, Jan. 1989, pp. 410-420.

[12] V. Kaplan, J.N. Zhang, E.W. Russi, and K.E. Bloch, “Detection of inspiratory flow limitation during sleep by computer assisted respiratory inductive plethysmography.,” *Eur Respir J*, vol. 15, Mar. 2000, pp. 570-578.

[13] R.N. Westhorpe and C. Ball, “Precordial and oesophageal stethoscopes,” *Anaesthesia and Intensive Care*, vol. 36, Jul. 2008, p. 479.

[14] R.C. Prielipp, J.S. Kelly, and R.C. Roy, “Use of esophageal or precordial stethoscopes by anesthesia providers: are we listening to our patients?,” *J Clin Anesth*, vol. 7, 1995, pp. 367-372.

[15] V.P. Harper, H. Pasterkamp, H. Kiyokawa, and G.R. Wodicka, “Modeling and measurement of flow effects on tracheal sounds,” *IEEE Journal of Biomedical Engineering*, vol. 50, 2003, pp. 1-10.

[16] R. Beck, G. Rosenhouse, M. Mahagnah, R.M. Chow, D.W. Cugell, and N. Gavriely, “Measurements and theory of normal tracheal breath sounds.,” *Ann Biomed Eng*, vol. 33, 2005, pp. 1344-1351.

[17] N. Gavriely and D.W. Cugell, “Airflow effects on amplitude and spectral content of normal breath sounds.,” *J Appl Physiol*, vol. 80, Jan. 1996, pp. 5-13.

[18] N. Gavriely, Y. Palti, and G. Alroy, “Spectral characteristics of normal breath sounds.,” *J Appl Physiol*, vol. 50, Feb. 1981, pp. 307-314.

[19] G. Sierra, V. Telfort, B. Popov, L.G. Durand, R. Agarwal, and V. Lanzo, “Monitoring respiratory rate based on tracheal sounds. First experiences.,” *Conf Proc IEEE Eng Med Biol Soc*, vol. 1, 2004, pp. 317-320.

[20] Z.M.K.M. Azadeh Yadollahi, “A robust method for estimating respiratory flow using tracheal sound entropy,” *IEEE Transactions on Biomedical Engineering*, vol. 53, 2006, pp. 662-668.

[21] A. Yadollahi and Z. Moussavi, “Comparison of flow-sound relationship for different features of tracheal sound,” *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2008*, 2008, pp. 805-808.

[22] “Masimo - Acoustic Respiration Rate (RRa).” Available http://www.masimo.com/rra/index.htm. [Accessed June 15, 2010].

[23] A.L.L. Egnel and M. Paiva, *Gas Mixing and Distribution in the Lung*, Dekker, 1985.

[24] D.E. Olson, M. Bogyi, D.B. Schwartz, and J.R. Hammersley, “Relationship of tracheal breath sounds to airflow,” *American Review of Respiratory Disease*, vol. 129, 1984.

[25] S.S. Kraman and P.M. Wang, “Airflow-generated sound in a hollow canine airway cast,” *Chest*, vol. 97, Feb. 1990, pp. 461-466.

[26] J.L. Benumof, “Obesity, sleep apnea, the airway and anesthesia.,” *Curr Opin Anaesthesiol*, vol. 17, Feb. 2004, pp. 21-30.

[27] D.W. Hudgel, “The role of upper airway anatomy and physiology in obstructive sleep apnea.,” *Clin Chest Med*, vol. 13, Oct. 1992, pp. 383-398.

[28] R. Jane, J. Sola-Soler, J.A. Fiz, and J. Morera, “Automatic detection of snoring signals: validation with simple snorers and OSAS patients,” *Proc. 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2000, pp. 3129-3131.

[29] H. Hara, N. Murakami, Y. Miyauchi, and H. Yamashita, “Acoustic analysis of snoring sounds by a multidimensional voice program.,” *Laryngoscope*, vol. 116, Mar. 2006, pp. 379-381.

[30] H. Nakano, M. Hayashi, E. Ohshima, N. Nishikata, and T. Shinohara, “Validation of a new system of tracheal sound analysis for the diagnosis of sleep apnea-hypopnea syndrome.,” *Sleep*, vol. 27, 2004, pp. 951-957.

[31] A.K. Ng and T.S. Koh, “Using psychoacoustics of snoring sounds to screen for obstructive sleep apnea,” *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2008*, 2008, pp. 1647-1650.

[32] S. Gazor and W. Zhang, “A soft voice activity detector based on a -Gaussian model,” *IEEE Transactions on Speech and Audio Processing*, vol. 11, 2003, pp. 498-505.

[33] Jongseo Sohn, Nam Soo Kim, and Wonyong Sung, “A statistical model-based voice activity detection,” *Signal Processing Letters, IEEE*, vol. 6, 1999, pp. 1-3.

CHAPTER 2

DETECTION OF APNEA

2.1 Introduction

2.1.1 Apnea and its risks during sedation

Apnea is the cessation of lung ventilation for a defined period of time, from 15-20 seconds (see Section 2.1.2). Apnea can come in two primary forms: respiratory depression (RD) and respiratory obstruction (RO). Both problems pose a serious risk to the patient if not detected and treated. Respiratory depression is caused by the lack of drive to breathe although the airway may be open. RO is caused when tissues in the upper airway relax and the upper airway is occluded. During RO the subject may be trying to breathe, but no ventilation occurs. These types of apnea are discussed in greater detail in Chapter 3. A sedated patient can suffer from either form of apnea depending on the amount and type of drug used to sedate the patient. The goal of the physician is to use enough drug to make the patient comfortable but still allow him/her to breathe on his/her own. Clinical monitors help in ensuring that the patient continues to breathe, but depending on the clinical monitors available, not all periods of apnea are detected before adverse events or even death occurs [1,2].

2.1.2 Apnea detection and definition

An acoustic apnea detection monitor has been previously investigated as a simple and effective monitor to detect apnea during procedures as described above. Several attempts have already been made to detect breath rate, as described in Chapter 1. In Sierra et al. [3], breath rate was measured from the sounds recorded on the neck near the trachea via a piezo-electric film transducer. Great care was taken to measure the breath rate very accurately, but the breath rates were not low enough that apnea would have been detected. The lowest measured breath rate by the gold-standard flow meter was ten breaths per minute in healthy individuals, and five breaths per minute in patients with pulmonary disease. Although it is important to detect breaths, it is far more important to be able to detect periods of apnea accurately.

Apnea has been defined as the cessation of breathing for at least ten seconds in [4,5]. Little physiological reasoning was given for this definition. In the case of pediatrics, apnea has been defined as a period lasting at least 15 seconds without breathing [6]. The defined length of apnea described in the literature is variable; for this study, a period of fifteen seconds without a valid breath was defined as a period of apnea.

A valid breath is defined as a breath with enough breath volume to clear the airway dead-space. The airway dead-space consists of the volume in the oral passage, nasal passage, trachea, and bronchi. In the average adult, this volume is estimated to be 150 milliliters [7].

2.1.3 Correlation of breath sounds to flow rate

The envelope of normal breathing sounds is highly correlated to the flow rate, and when properly calibrated and processed, breath sounds can predict flow rate [8,9]. Breath

sounds are generated in adults when the flow rate exceeds approximately 7.2 liters per minute given the Reynolds number of approximately 1000 for the anatomy of the trachea of an adult [10]. As the rate of the flow increases, the amplitude of the sound also increases. This relationship is not linear, and since flow during both inspiration and expiration creates turbulence, direction of flow from the sound alone cannot accurately be determined. Beck and Olson et al. [11,12] state that the relationship of flow to sound amplitude is $A=kF^{3.0\pm0.2}$ where k is a fitting constant, A is the amplitude of the sound and F is the flow rate, but they did not indicate the units of k, F or A. Using data selected from that data set that will be described in Section 2.2, the plot shown in Figure 2.1 was created that shows the comparison of flow rate in units of liters per minute and the audio amplitude in units of volts. The audio envelope was calculated from the raw audio as will be described in Section 2.3. The data between the markers at +/-7.2 l/m are observed to have little variation which supports the claim by Ludwig that absolute flow rates below 7.2 l/m would not create turbulent noise. The value between these markers is not zero and it is hypothesized that the signal that created this portion of the data is due to noise. In addition, a model was fitted to the data with equation $A=|kF^3|$ where A is the audio envelope amplitude in units of volts, F is the flow rate in units of l/min, and k is a fitting constant with units of $V/(L/min)^3$ and valued at $k=1.5\times10^{-7}$. The correlation coefficient of the data with absolute flow rate greater than 7.2 l/m is R=0.8869.

2.2 Data Set

2.2.1 Data collection process

In an IRB-approved study, 24 subjects were sedated by injecting a combination of remifentinal and propofol intravenously. Remifentinal is a short acting opioid analgesic

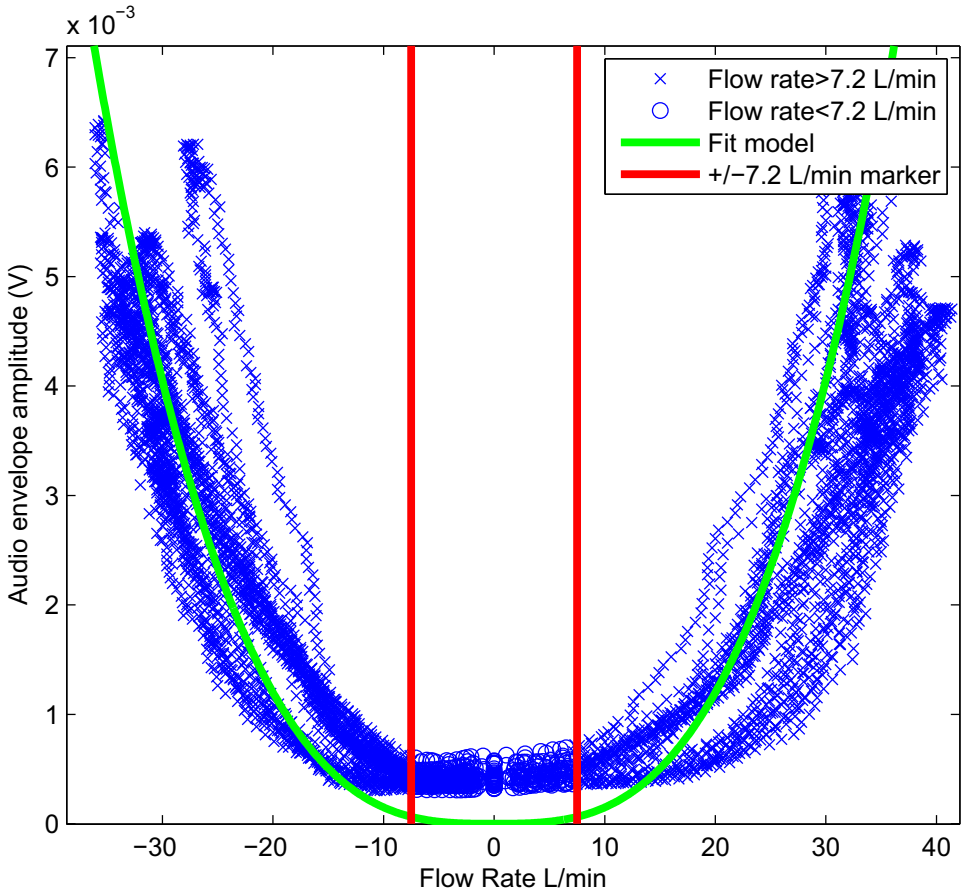


Figure 2.1 Comparison of flow rate measured by Cosmo+II pneumotachograph (Respironics,Wallingford, CT) to audio envelope amplitude.

used to reduce pain sensed by the subject. Propofol is a sedative/hypnotic used to lessen the awareness of the subject. The amount of each drug was varied according to the chart in Figure 2.2, where the concentration of one drug was held constant and the other was incremented. The drug-dosing for each subject was assigned randomly from a pre-determined drug-dosing regimen. The randomized drug-dosing regimen would randomly assign each drug as the first or second drug and the amounts of each drug to be used according to predetermined sedation strategies. Each subject underwent three periods of sedation. During the first period of sedation, the first drug concentration was incremented stepwise while the second drug concentration was held constant. In the second period of sedation, the concentration of the second drug was increased while repeating the drug-incrementation regimen of the first drug from the first period. In the third period, the concentration of the second drug was again increased while repeating the drug-incrementation regimen of the first drug from the first and second sedation period. The primary design of this study was to measure the alertness and pain tolerance of the subjects for different concentrations of the two drugs.

The sedation method used for the study was target-controlled infusion (TCI). TCI takes advantage of pharmacokinetic models to maintain a constant sedation level for the patient [13]. Although several phamacokinetic models exist, the Minto model was used for the remifentinal infusion and the Schnider model was used for the propofol infusion to maintain the sedative level of the subjects of this study.

The subjects were monitored using both standard medical monitors (described below) and by a board certified anesthesiologist to avoid any complications caused by the injected drugs. Each subject was given an initial dosage of the two drugs and tests were

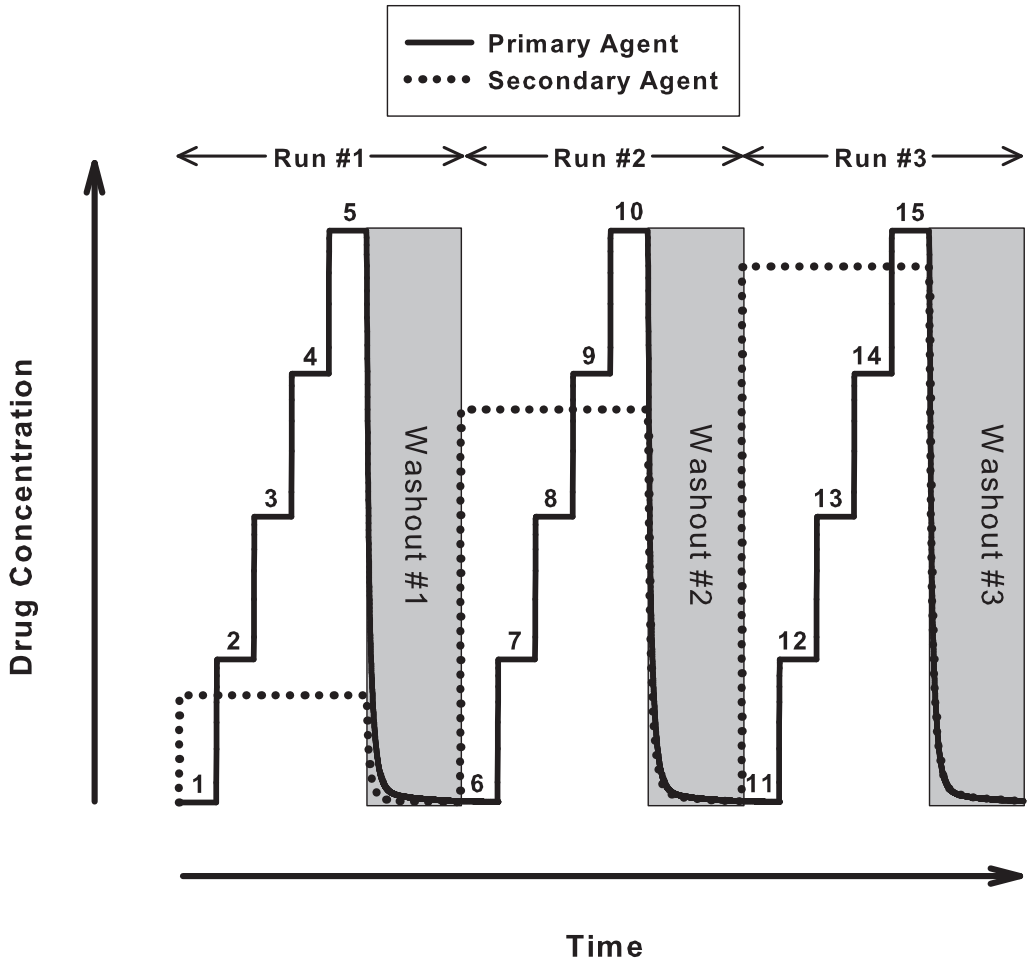


Figure 2.2 Drug-dosing regimen for sedating patients participating in the study.

performed at baseline as well as before increasing the drug amounts. The tests included the observer's assessment of alertness and sedation (OAA/S), tetanic stimulus, pressure algometry, and attempting to insert a bougie into the subjects' esophagus. OAA/S is done as a nurse tries to communicate with the subject, and the nurse rates the amount of interaction received for specific communication attempts. Tetanic stimulus is a controlled electric shock to the leg. Pressure algometry is the application of a pressure generated by a one square inch piston to the subjects' shin bone. A bougie is a rubber tube used to simulate inserting an esophageal scope. The bougie test required the removal of the facemask; thus data in these time periods did not record the flow through the tight-fitting facemask. The use of the tight-fitting facemask is described in the next section.

The primary study using the data of the sedated subjects attempted to find specific dosing combinations of remifentinal and propofol that would satisfy pain tolerances while allowing the subjects to breathe without the need of ventilation. The study undertaken in this dissertation was allowed to use the data recorded from the primary study and was allowed to record the audio data from the subjects.

Of the twenty-four subjects participating, data from four of the subjects were discarded. Two of the subjects' data were lost during data transfer of the audio data between computers. Flow and respiratory inductance plethysmography (RIP) data from another subject were lost during data transfer between two computers. One data set was recorded with a different microphone than for the other recordings, leaving twenty sets to be processed.

The monitoring equipment for each subject included respiratory flow, capnography, pulse-oximetery, RIP, bi-spectral EEG index, arterial blood pressure, and tracheal

audio recording. Only a few of these were pertinent to this project, namely the respiratory flow, RIP, and the tracheal audio recording.

2.2.2 Data measurement

2.2.2.1 Direct flow data measurement

The respiratory flow was measured by placing a tight-fitting facemask onto each subject and measuring the flow rate via a differential pressure technique called pneumotachography implemented in the CosmoII+ device (Respironics, Wallingford, CT). This device measures flow rate directly in units of liters per minute (l/m). The device calculates breath volume by integrating the flow rate on a breath-to-breath basis and records it in units of milliliters (mL). This device also produced a breath marker at the beginning of inspiration for each breath. The data were recorded directly to a computer hard drive at a sampling rate of 100 Hz. Each sample recorded was accompanied with a time-stamp.

Respiratory flow measured by the CosmoII+ device served as the gold standard for apnea detection for this study because it is a direct measurement of the subjects' ventilation.

2.2.2.2 Plethysmography data measurement

True plethysmography is the measure of the change in body volume. This is measured by placing the subject inside a chamber and having the subject breathe through a tube to the open air and measuring the change in volume in the chamber [14].

Two common ways to estimate plethysmography without the need of a chamber are thoracic impedance plethysmography and RIP. Thoracic impedance plethysmography

measures the change of skin impedance through standard ECG leads. As the chest and abdomen volumes change, the body impedance between the ECG leads changes and is measured. RIP measures the change of inductance of bands placed on the chest and abdomen of the patient. The bands are lined with elastic with a conducting element sewn into the fabric. The conducting element is sewn into the bands in a "V" pattern, inducing mutual inductance. As the chest and abdomen volumes change the circumference of the bands change thus changing the inductance of the conducting elements. The RIP device measures the changes in the inductance of the bands, which is proportional to the change in chest and abdomen volume [15].

In this study RIP was used to measure the thoracic and abdominal circumference change. The RIP device measured the change in the inductance of the bands as the subject breathed. The change in inductance of the RIP can be calibrated to the change in flow rate. The calibration is performed by comparing the change in the inductance to the respiratory breath volume as calculated by the CosmoII+ [16]. In this study the breath volume of each breath, segmented by the CosmoII+ device, was compared to the change in inductance from the RIP. If the correlation coefficient of the breath volume segment V_{flow} and the RIP segment L_{RIP} was greater than 0.75 the best fit line of the two data segments was calculated. This produced a gain m_{ccf} and offset b_{ccf}. The values of m_{ccf} and b_{ccf} were averaged for every 10 minutes of data. The values were recorded as the RIP calibration gain and offset.

The RIP measurements were recorded to the same hard drive as the flow measurements were, at a 100 Hz sampling rate.

2.2.2.3 Respiratory acoustic data measurement

Audio signals were recorded using a microphone (WM-56A103 Panasonic, NJ) placed in a metal precordial stethoscope cup (Wenger #00-390-C, AINCA, San Marcos, CA). The cup was affixed to the neck with a double-stick disc (#2181 3M, MN) just below the larynx and above the suprasternal notch of each subject. The audio signals were digitized at 22 kHz onto a computer hard drive using an audio sound card (Sound-Blaster Audigy, Creative, Singapore). Acoustic data were synchronized to the flow rate and RIP data by recording the start time of the beginning of the acoustic recording.

2.3 Methods

2.3.1 Creation of the audio envelope

The term "envelope" can have many definitions in signal processing. This work uses a basic template to build an acoustic envelope from a raw audio signal. Figure 2.3 shows a flowchart of the general steps of creating the acoustic envelope from the raw audio. The steps of audio amplification, analog low pass filter and A/D conversion were all performed by the audio sound card used to record these data (SoundBlaster Audigy, Creative, Singapore).

The first step after digitization was to apply a bandpass filter to the raw audio signal. An example of raw audio is shown in Figure 2.4a. The bandpass filter was designed to remove both high frequency noise not related to the breathing sounds and low frequency sounds, such as heartbeat and electronic noise. For this work the bandpass filter passed frequencies from 75 Hz to 1500 Hz. 75 Hz was chosen as the low cutoff frequency because of its ability to remove most of the heartbeat sounds [11] and common

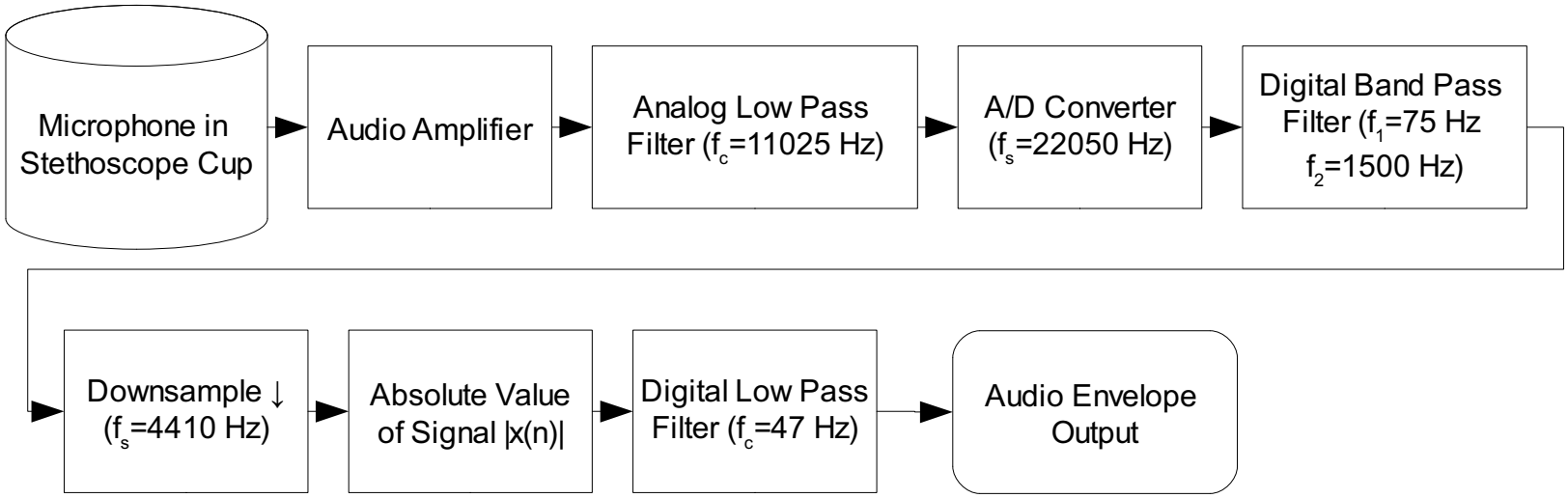


Figure 2.3 Flowchart showing steps used to create an acoustic respiratory monitor.

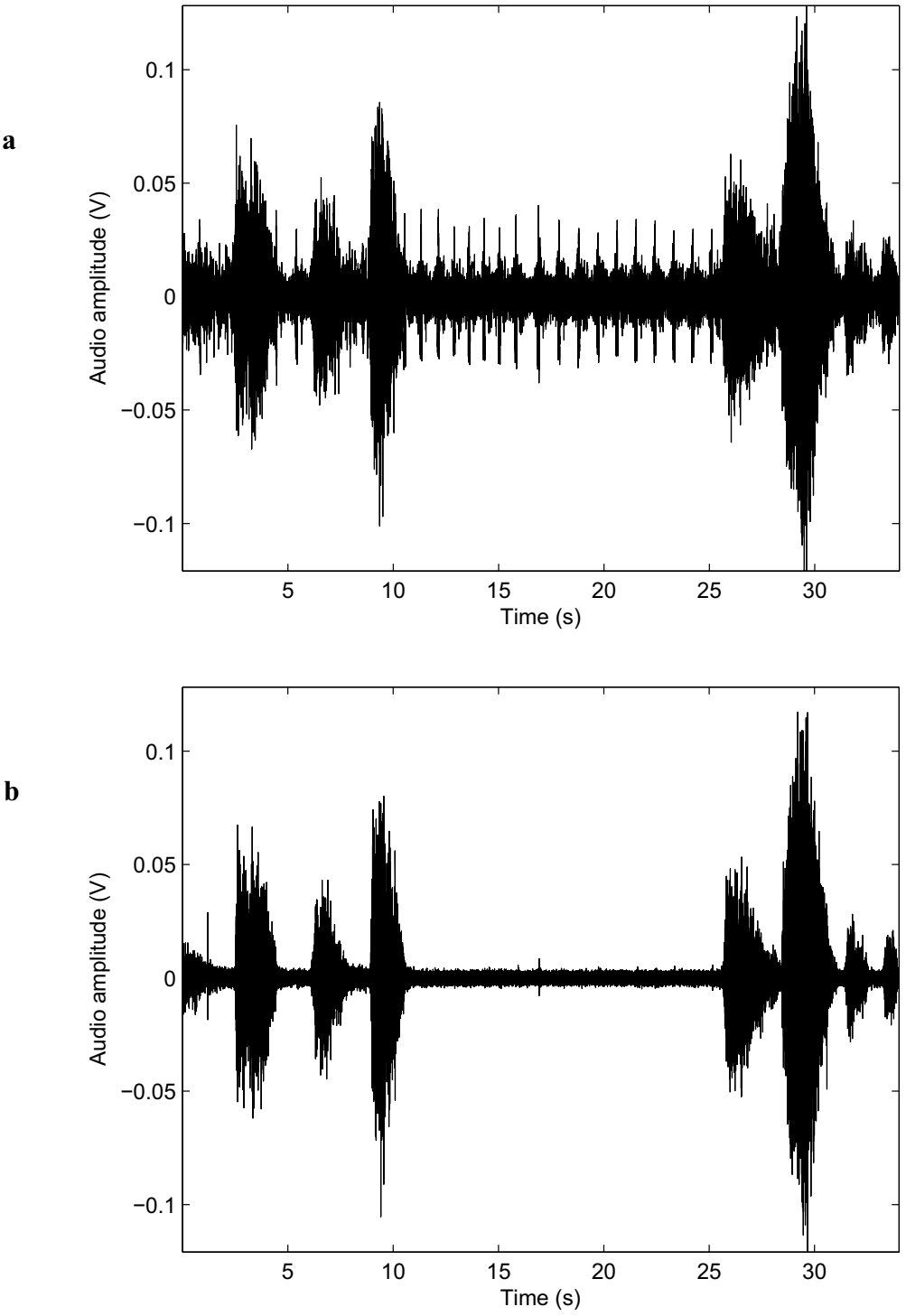


Figure 2.4 Section of audio data that shows a period of silence. (a) Raw audio (b)Filtered digitized audio signal.

c

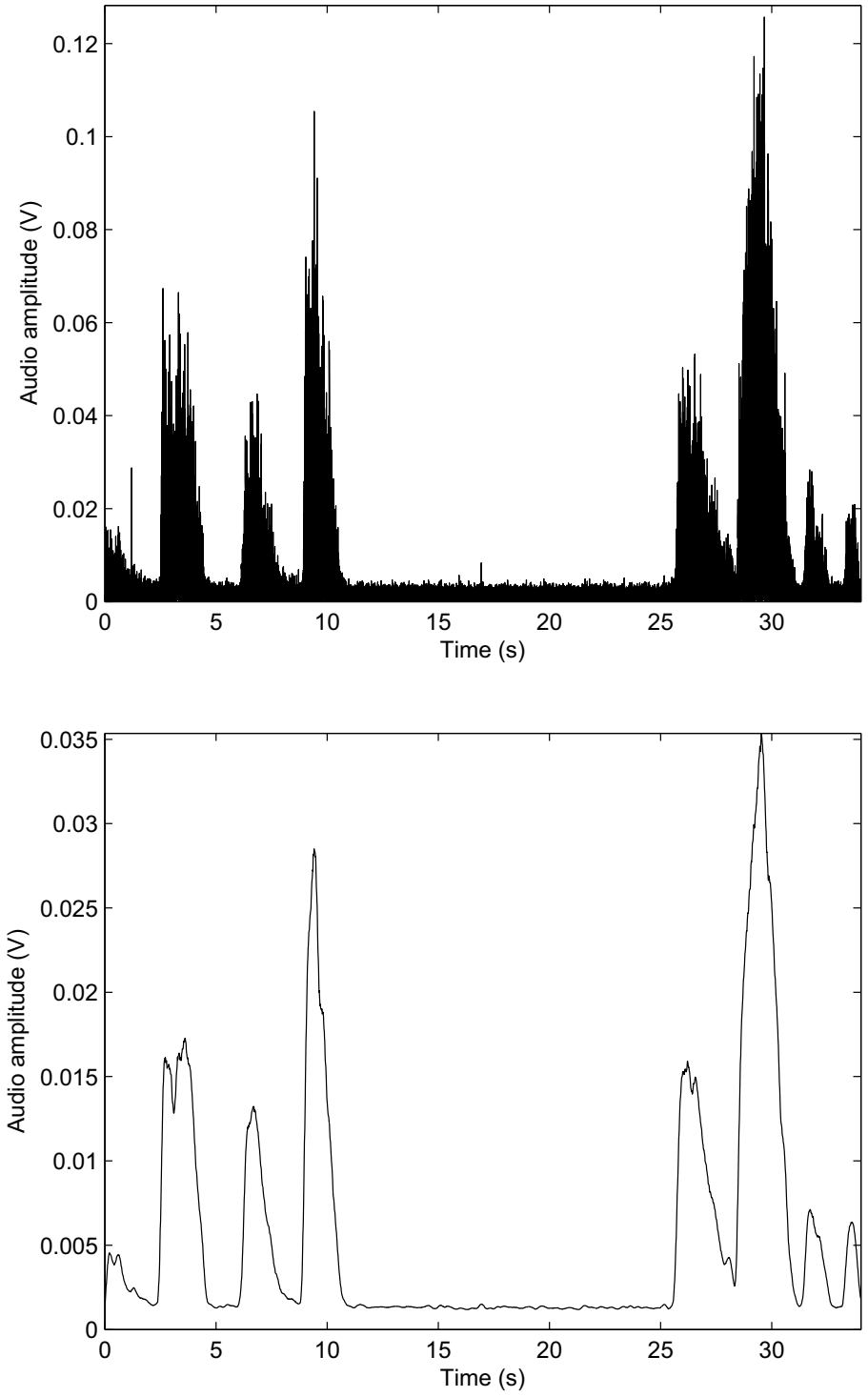


Figure 2.4 cont'd (c) Absolute value of the filtered signal (d)Envelope of the signal.

electronic noise at 60 Hz. The high cutoff at 1500 Hz was chosen because most information about the breath sounds is within the range of 100-1400 Hz [17]. The resulting signal is shown in Figure 2.4b.

The next step was to downsample the data. To avoid aliasing, this must done with a sample rate at least two times the high cutoff frequency of the bandpass filter. The audio was downsampled by a factor of five to 4410 Hz. Downsampling was performed by keeping every fifth sample in the filtered signal.

The absolute value was then applied to the resulting signal. The absolute value-doubles the frequency of the signal with an additional DC component. The DC component is an indicator of signal amplitude in the band from 75 Hz to 1500 Hz. This also creates a positive signal ensuring a positive envelope. The resulting signal is shown in Figure 2.4c.

Finally, a lowpass filter with corner frequency of 43 Hz was applied. This filter removed all high frequency content and left only a weighted average of the signal. The value of 43 Hz was chosen because it is less than the Nyquist frequency for a signal sampled at 100 Hz. This signal was then downsampled to 100 Hz to match the sample rate of the accompanying flow and RIP data. The resulting envelope is shown in Figure 2.4d.

2.3.2 Synchronization of acoustic data with flow and RIP data

Data recorded from the RIP and the CosmoII+ were recorded on the same computer, and were thus synchronized. Data from the microphone were recorded at a different rate and on a different computer. The start times of the audio recordings were recorded at the beginning of data collection for each recording period. Each recording

period lasted between one and four hours, and small errors in sampling rate or buffer errors in the audio data were found to cause errors between the synchronization of the audio data and the flow and RIP data. In order to ensure that the data were properly synchronized, short sections of the audio envelope were compared to the absolute value of the flow rate. Because of the high correlation of the flow rate to the audio envelope [8,9,11] manual synchronization was possible. This was performed as the researcher compared the envelope of the audio signal to the absolute value of the flow rate. Pauses between breaths and significantly larger breaths allowed for a high level of confidence that the synchronization was performed properly. In retrospect, the audio signal should have been synchronized to the flow rate and RIP data by recording the time stamp from the flow rate and RIP computer to a second channel of the audio data.

The data were divided into 10 minute segments to allow for ease of processing and ease of finding and sorting of data.

2.3.3 Data separation for blinded testing

Of the 20 viable data sets, 10 were randomly selected to be held out for testing purposes. The subjects' data sets were assigned an index number (1-20) in the order that they were recorded. Twenty nonrepeating random numbers U_{RN} were generated using a pseudo-random number generator, and were assigned an index number I_{RN} (1-20) in the order that they were generated. I_{RN} was then sorted by the value of U_{RN} from lowest to highest. The first 10 sorted values of I_{RN} were used to select the data sets of the subjects that would be held out as the testing set. The remaining 10 subjects' data were used as the training set. The training set allowed for the researcher to develop algorithms to detect apnea. The testing set allowed half of the data to be blinded against researcher bias

before processing it with the algorithms.

2.3.4 Acoustic threshold modeling

The acoustic sounds recorded were composed of two primary sources, the noise source *n(t)* and the signal source *s(t)*. During times of breathing, vocalization, or snoring the recorded signal *x(t)* was the summation of the two sound sources; otherwise the recorded signal was made up entirely of the noise source as described by:

$$\begin{cases} x(t)=n(t)+s(t), & \textit{for periods of detected signals} \\ x(t)=n(t), & \textit{for periods of silence} \end{cases} \tag{2.1}$$

Equation 2.1 is the basis for the model that will be used to differentiate the noise from the signal.

2.3.4.1 Description of the noise source

The noise was a wide sense stationary (WSS) signal recorded at the microphone. This signal consisted of several additive sources. The two main sources of noise at the microphone were thermal noise and ambient WSS room noise that was transmitted through the precordial stethoscope. Thermal noise was the most significant source of this signal. This was because the stethoscope cup shielded the microphone from any major noise coming from outside the subject. Because the thermal noise made up the majority of the noise, the noise was considered WSS and Gaussian [18].

The first attempt at finding the threshold of the noise involved modeling the histogram of the audio envelope. The process of finding the parameters of this model are described below. Although this method was eventually abandoned, it helped in the development of the modeling of the raw audio signal to determine a threshold of the noise.

2.3.4.2 Audio envelope histogram model

Using the methodology of calculating the sound envelope described above, an envelope was created for each 10 minute segment of data.

Initially it was noticed that the histogram of the envelope amplitude had a distinct pattern that would help in identifying noise threshold. A typical histogram of the envelope is shown in Figure 2.5a. The portion shown by the thick line was hypothesized to be noise of the audio envelope. This was hypothesized because it was observed that this portion of the distribution would show little change after changing the window of the data used to generate the histogram. The portion of the distribution with envelope exceeding the presumed noise threshold, shown as a thin line, was hypothesized to be the desired signal. This was hypothesized because this portion of the histogram fluctuated in amplitude and samples in each bin as the audio envelope changed in amplitude, and it also had a higher amplitude than the portion of the histogram attributed to noise. In most cases it was observed that the portion of the distribution attributed to noise had a Gaussi-an appearance. To test this, a portion of audio was found that contained no detected sounds. This portion was identified by observation and confirmed that no breathing occurred when compared to the flow rate measurement. The histogram of an example of this signal with a best fit Gaussian curve is shown in Figure 2.5b. From the figure shown, the histogram of the noise segment does not match the distribution of the fitted Guassian primarily because the histogram is not symmetric about the mean.

The part of the distribution attributed to the detected sounds took on many forms. The most commonly observed distribution was similar to an exponential distribution as

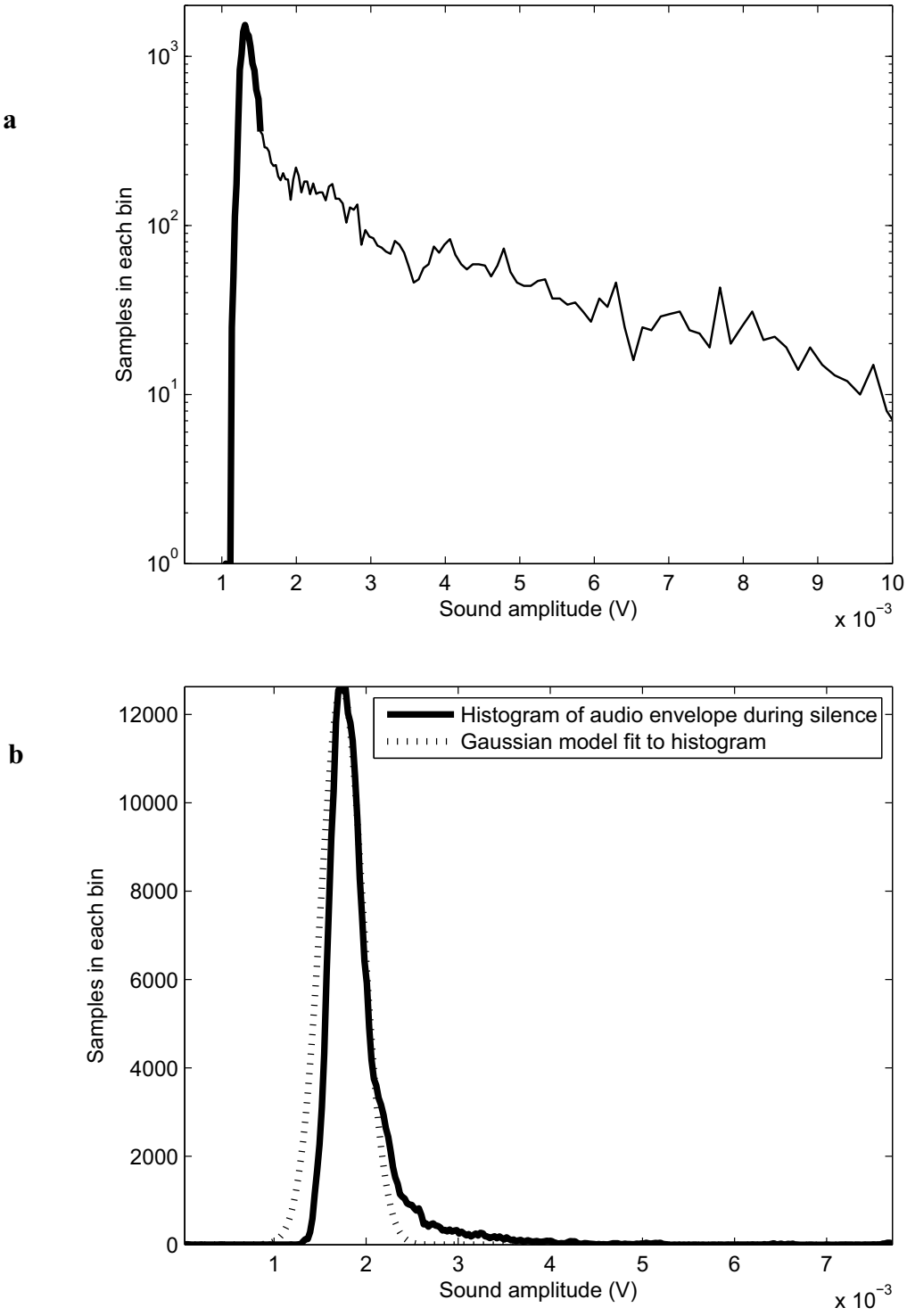


Figure 2.5 Plot of the histogram of an audio envelope. (a) Semilog plot of the histogram of an example audio envelope(b) During a period of silence with best fit Normal distribution.

seen in Figure 2.5a. This observation was based on the histogram displayed with a logarithmic plot similar to the one shown in Figure 2.5a from ten normal breathing segments. An exponential distribution on a logarithmic plot is a sloped line and was easy for the primary investigator to identify histograms with a similar plot. This distribution was observed on the ten normal breathing segments above the portion of the distribution attributed to noise.

Using the model of Gaussian distributed noise with an exponentially distributed breathing signal, a threshold can be determined from the standard deviation of the noise. The mean of the Gaussian signal μ_{NF} was estimated as the bin number with the maximum value of the entire histogram. This estimation is based on the assumption that the most common signal during a recording was the noise.

Using the value of μ_{NF} as the center of the distribution, bins lower in amplitude than μ_{NF} were summed until they made up 34% of the values between 0 and μ_{NF}. The value of 34% was used because in a Gaussian distribution 34% of the area of the curve is found between the mean and one standard deviation of half of the distribution. The difference between μ_{NF} and the lowest valued bin described was used as the estimate of σ_{NF}. The threshold to divide the noise signal from the breathing signal was calculated to be $\mu_{NF} + 3\sigma_{NF}$ such that the breathing portion of the signal would have to be at least three standard deviations higher in envelope than the noise threshold for it to be detected.

The model described was developed using a small number of data sets. Initially this appeared to be a very effective method of finding the noise threshold, but using a broader pool of data it was observed that the model did not estimate the parameters of the distributions as well as for the initial data sets. Common examples of this were when the

portion of the histogram attributed to breathing sounds did not have an exponential distribution. This was a common occurrence when the amplitude of the signals recorded saturated the microphone. Another reason for poor estimation of the parameters was when the assumption that the noise was most common sound was not true. In these cases the portion of the histogram attributed to breathing sounds had a higher amplitude than the portion of the histogram attributed to noise. This caused the estimate of μ_{NF} and σ_{NF} to be too large and breathing sounds were missed.

The failure of this model and algorithm can be attributed to several reasons. The first reason was that the maximum point of the distribution of a sound was not always the maximum point of the Gaussian distribution as described above. The second reason was the histogram of the noise of the envelope was assumed to be a Gaussian distribution, but when compared to a Gaussian distribution it became apparent that this assumption was not correct. In addition the calculation of the distribution of the envelope of the noise could not be found. The third reason was the distribution of the detectable signal was not always distributed as an exponential distribution as described above.

Another reason that this method was abandoned in favor of the method described in Section 2.3.4.3 was that the sample rate of the audio envelope at 100 Hz required several minutes of data to initially populate a histogram to be able to estimate the model parameters when compared with the method that will be described.

2.3.4.3 Raw audio histogram model

The experimentation with the model of the distribution of the audio envelope led to an investigation of the distribution of the raw audio. This investigation was prompted because the raw audio would be able to populate a histogram using a shorter period of

data, and the mean of the distribution would already be known to be zero. The distribution of the noise portion of the raw audio data could also be directly attributed to be Gaussian signal rather than just assumed to be. This is because primary source of noise is thermal noise and can be modeled as a Gaussian distribution. To further prove this model a period was found that was confirmed by the flow meter to be a period of noise only. The histogram of these data was calculated and a Gaussian distribution was fitted to it as shown in Figure 2.6. The standard deviation of the Gaussian distribution fitted to this model was $\sigma=3.6\times10^{-4}$. This histogram shows a high correlation to the Gaussian distribution with a correlation coefficient of R=0.9984.

The raw audio signal was first filtered using the bandpass filter and down-sampled to a rate of 4410 Hz as shown in Figure 2.4b. An example of a histogram of this signal can be seen in Figure 2.7. The noise has been modeled as a Gaussian random signal with zero mean. The detected breathing signal was observed to have a distribution with a much higher standard deviation and heavier tails than a Gaussian distribution. A segment of the audio used to generate this histogram is shown in Figure 2.8.

The heavy-tailed distribution of signal that has been attributed to the breathing signal, was modeled as a Laplace distribution with zero mean, based on observation. Observations showed that the histograms of the sound attributed to breathing sounds observed were very similar to the Laplace distribution. In addition, sounds such as snoring and vocalization created heavy-tailed histograms. Gazor et al. [19] also modeled the voiced periods of his voice-activity detector as a Laplace distribution. The noise and breath signals can be considered nonoverlapping in time, since during periods of detectable signal it is assumed that the breathing signal has a standard deviation a factor of 10

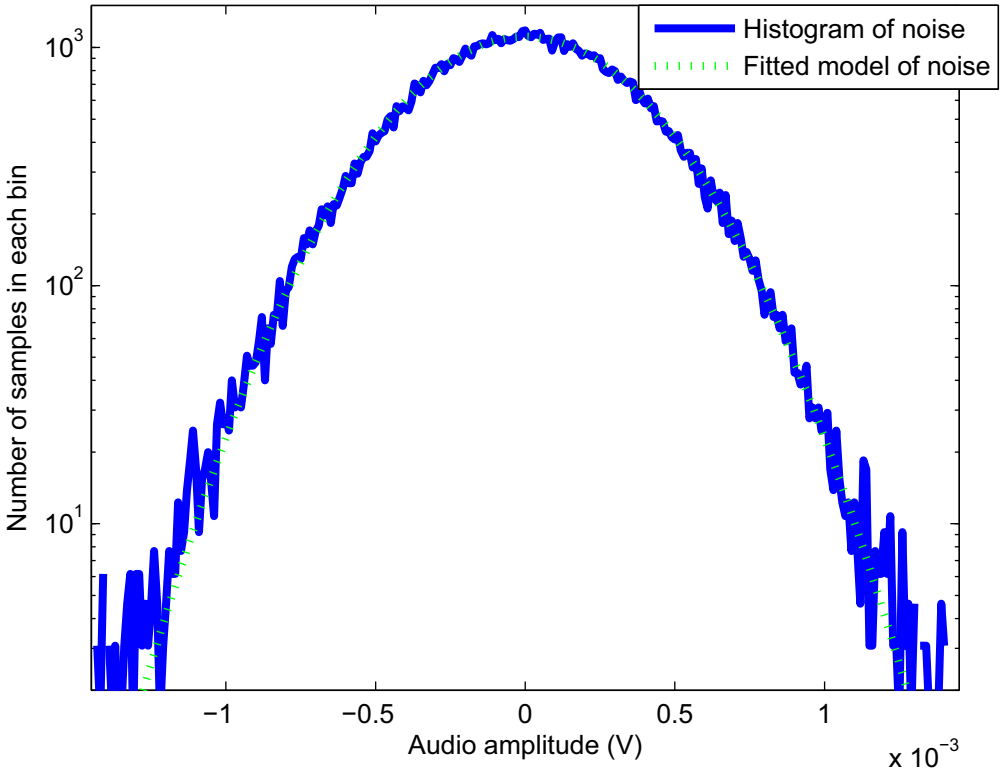


Figure 2.6 Histogram of a typical audio signal during a period of noise.

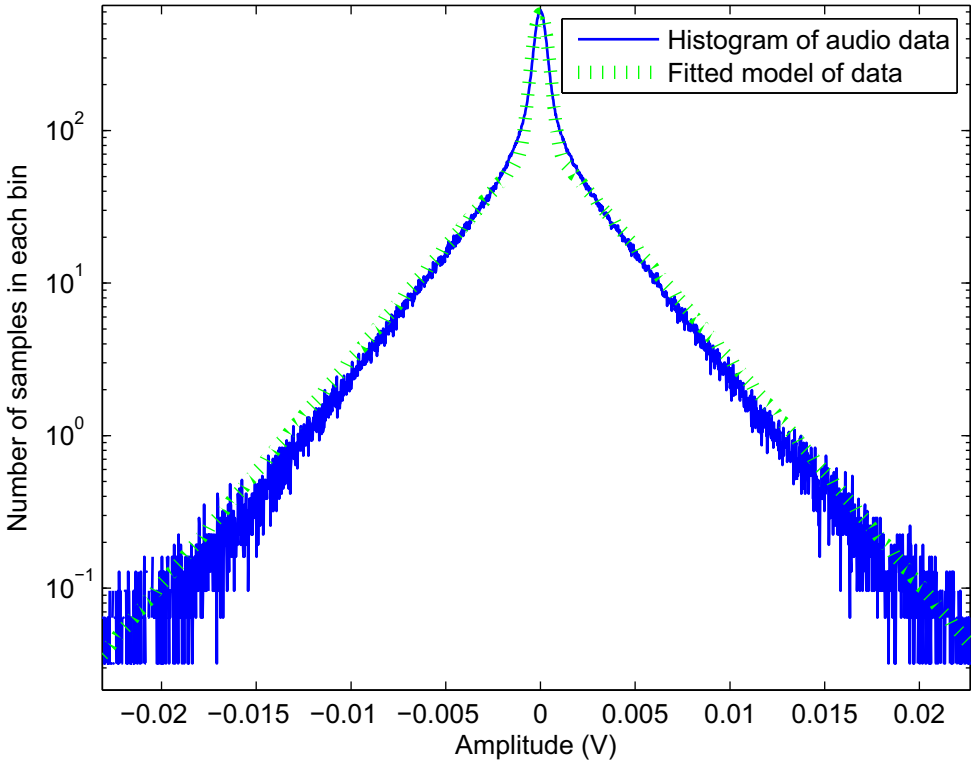


Figure 2.7 Histogram of a typical audio signal including breathing and the model fit to it using the EM algorithm.

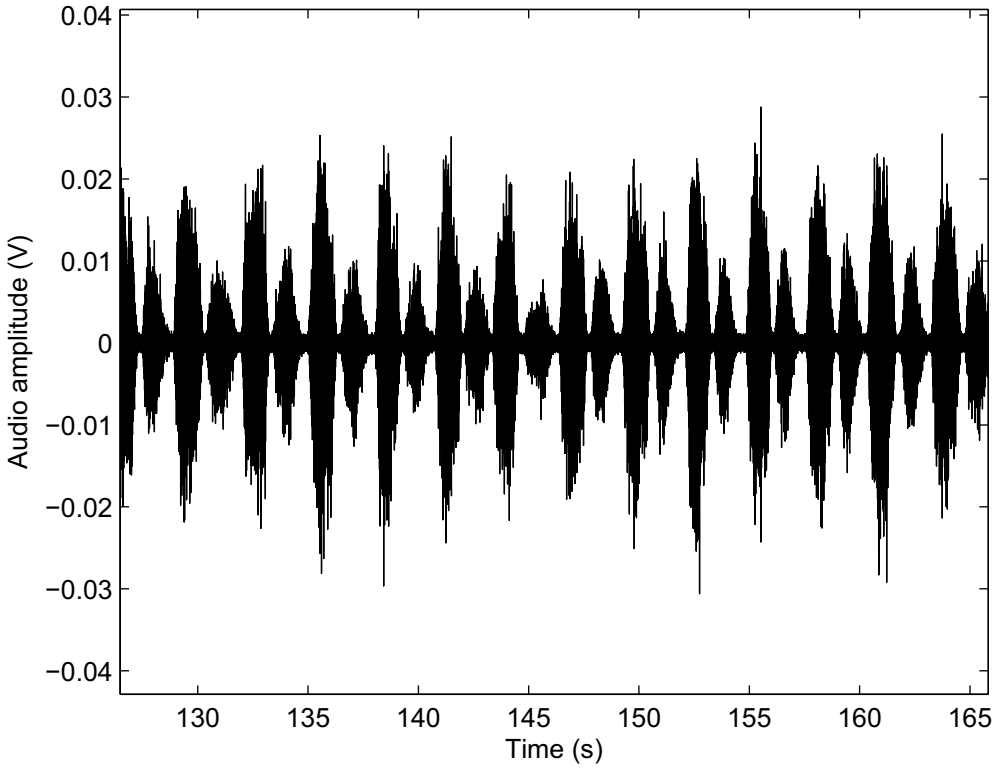


Figure 2.8 Audio signal used to generate histogram and model.

greater than the standard deviation of the noise that it does not contribute significantly to the Laplace distribution. During times of breathing silence, the noise signal is the only signal present. Because these two signals can be considered to be nonoverlapping in time, the convolution of the two distributions does not need to be performed. The assumption that the noise is insignificant in comparison to the detected signal modifies the acoustic model to be:

$$\begin{cases} x(t)=s(t), & \textit{for periods of detected signals} \\ x(t)=n(t), & \textit{for periods of silence} \end{cases} \tag{2.2}$$

The mixture of the models can be modeled simply as the summation of the two distributions in proportion to each other with the equation:

$$f_m(x)=\frac{p}{\sqrt{2\times\pi\sigma^2}}e^{\frac{-x^2}{2\sigma^2}}+(1-p)\frac{1}{2\mathrm{b}}e^{\frac{-|x|}{b}} \tag{2.3}$$

where $f_m(x)$ is the distribution of the mixture, x is the random variable in units of volts, p is the probability that the signal only contains noise, σ is the standard deviation of the Gaussian distribution, and b is the parameter of the Laplace distribution. The standard deviation of the Laplace distribution is $\sigma_L=\frac{\sqrt{2}\,b}{2}$.

This model is based upon the assumption that the standard deviation of the Laplace distribution σ_L is much greater than the standard deviation of the Gaussian distribution σ. If σ_L is less than or equal to σ the signal to be detected would have less amplitude than the noise and would not be audible. A breath sound signal with absolute amplitude a factor of 10 times that of the standard deviation of the noise ensures that that signal is greater than more than 99% of the noise samples. This is based on the assumption that the noise is WSS Gaussian and that the standard deviation of the noise has been

fitted accurately.

Using the raw signal was an advantage over the envelope for two reasons. The number of samples was dramatically increased for a given period of time. This allows for the histogram to be populated in a shorter period of time. This model bases the distribution of the noise signal on the known distribution of thermal noise and has further shown that the noise can be accurately fit with a Gaussian distribution. This model also uses the nature of the audio signal as a zero mean distributed signal to remove the need for calculating the mean of the two distributions.

2.3.5 Finding the model parameters

2.3.5.1 Finding the model parameters directly

Finding the parameters of the distribution mixture, namely p, σ, and b, was the next task. Because there are three unknowns to the above equation, three equations are needed to find all of the parameters.

The initial attempt was to find the second, fourth, and sixth moments of a signal and compare them to the mathematical model of the distributions. This method was eventually abandoned, but shows the step made towards model parameterization.

The odd moments on a symmetric distribution are zero, thus the first three even moments were selected. For the equations below, A is the second moment, B is the fourth moment and C is the sixth moment calculated from the audio signal.

$$\begin{aligned} A &= p\sigma^2 + (1-p)2b^2 \\ B &= p3\sigma^4 + (1-p)24b^4 \\ C &= p15\sigma^6 + (1-p)720b^6 \end{aligned} \tag{2.4}$$

This set of equations does not provide a close estimate for the parameters because the

estimates of the fourth and sixth moments are noisy and that amplifies the amount of noise in the estimates. In addition, the signal modeled by the Laplace distribution would have to be nearly perfectly Laplace distributed for these equations to estimate the proper values for *p, σ,* and *b*. Because the actual signals can be distributed differently from the model it is possible that this parameterization technique would fail. For these reasons a different approach to finding the parameters was pursued.

2.3.5.2 Finding the model parameters using the expectation maximization algorithm

The expectation maximization (EM) algorithm [20] was used to estimate the values of p, σ, and b. The EM algorithm is an iterative method that uses initial estimates of the parameters and iterates until the estimates of the parameters converge on a value. The generalized EM algorithm has two major steps, the expectation step (E-step) and the maximization step (M-step).

The general equations for the EM algorithm are given by Moon et al. [20]. The first equation is the general form of the E-step:

$$Q(\Theta \vee \Theta^{[k]}) = E[\log f(\boldsymbol{x}|\Theta)|\boldsymbol{y}, \Theta^{[k]}] \tag{2.5}$$

where Q is defined as the *Q-function*, $f(\boldsymbol{x}|\Theta)$ is the differentiable pdf of the mixture, $\boldsymbol{y}$ is a vector of observed values, $\boldsymbol{x}$ is the vector of the combination of unobserved values and observed values, and $\Theta^{[k]}$ is set of estimated parameters given the k^{th} iteration of the algorithm. The general equation for the M-step is given as:

$$\Theta^{[k+1]} = \arg\max_{\Theta} Q(\Theta|\Theta^{[k]}). \tag{2.6}$$

The EM algorithm is based on the assumption that an observed value y_i has an accompanied unobserved value z_i. The theoretical combination of these two values is x_i

where $x_i=[y_i, z_i]$ and $\boldsymbol{x}=[x_1, x_2...x_n]$.

Considerable simplifications can be made if it can be proven that the distribution model is in the exponential family of distribution. Moon et al. [20] define a distribution as being part of the exponential family if it has the form:

$$f(\boldsymbol{x}|\Theta)=b(\boldsymbol{x})e^{(\left[\boldsymbol{C}(\Theta)^T\boldsymbol{t}(\boldsymbol{x})\right])}/a(\Theta) \tag{2.7}$$

where the superscript T denotes the vector transpose, a and b are scalar functions, and $\boldsymbol{C}$ and $\boldsymbol{t}$ are vector functions that need to be identified in the distribution in question. The distribution used for this method is the Gaussian-Laplace Mixture:

$$f(\boldsymbol{x}|\Theta)=\Pi_{i=1}^{n}\{\boldsymbol{I}(z_i=0)\frac{p}{\sqrt{2\pi\sigma^2}}e^{\frac{-y_i^2}{2\sigma^2}}+\boldsymbol{I}(z_i=1)\frac{(1-p)}{2\mathrm{b}}e^{\frac{-|y_i|}{b}}\} \tag{2.8}$$

where $\boldsymbol{I}$ is the indicator function that has value of 1 if the contents are true and zero if the contents are false, and z_i is the unobserved value that the observed value of y_i is from the distribution indicated by the indicator function. If $z_i=0$ the accompanying value of y_i is created by the Gaussian process and if $z_i=1$ the accompanying value of y_i is created by the Laplace process. In order to get equation 2.8 into the form of of equation 2.7, the exponential form of all values within the product term were found as:

$$f(\boldsymbol{x}|\Theta)=\Pi_{i=1}^{n}\{e^{\boldsymbol{I}(z_i=0)}e^{\log(\frac{p}{\sqrt{2\pi\sigma^2}})}e^{\frac{-y_i^2}{2\sigma^2}}+e^{\boldsymbol{I}(z_i=1)}e^{\log(\frac{(1-p)}{2\mathrm{b}})}e^{\frac{-|y_i|}{b}}\} \tag{2.9}$$

where $\Theta=[\sigma,b,p]$. The exponent was then taken outside of the product creating a summation shown as:

$$\begin{aligned} f(\boldsymbol{x}|\Theta)=\exp\Bigg[&\Sigma_{i=1}^{n}\boldsymbol{I}(z_i=0)(\log\left(\frac{p}{\sqrt{2\pi\sigma^2}}\right)+\frac{-y_i^2}{2\sigma^2}) \\ &+\boldsymbol{I}(z_i=1)(\log\left(\frac{(1-p)}{2\mathrm{b}}\right)+\frac{-|y_i|}{b})\Bigg]. \end{aligned} \tag{2.10}$$

From equation 2.10 the parameters of a, b, $\boldsymbol{C}$, and $\boldsymbol{t}$ can be assigned with $a=1$, $b=1$, and:

$$\boldsymbol{C}(\Theta)=\left[\log\left(\frac{p}{\sqrt{(2\pi\sigma^2)}}\right),\quad \frac{-1}{2\sigma^2},\quad \log\left(\frac{1-p}{2\mathrm{b}}\right),\quad \frac{-1}{b}\right], \tag{2.11}$$

$$\boldsymbol{t}(\boldsymbol{x})=\left[\Sigma_{i=1}^{n}\boldsymbol{I}(z_i=0),\quad \Sigma_{i=1}^{n}\boldsymbol{I}(z_i=0)\,y_i^{2},\quad \Sigma_{i=1}^{n}\boldsymbol{I}(z_i=1),\quad \Sigma_{i=1}^{n}\boldsymbol{I}(z_i=1)\,|y_i|\right]. \tag{2.12}$$

With this proof that the distribution model is in the exponential family, the simplified version of the EM algorithm for this special case was pursued. Using the simplification the E-step calculation becomes [20]:

$$\boldsymbol{t}^{[k+1]}=E\left[\boldsymbol{t}(\boldsymbol{x})\middle|y,\Theta^k\right] \tag{2.13}$$

and the M-step is [20]:

$$\Theta^{[k+1]}=\underset{\Theta}{argmax}\left[\boldsymbol{C}(\Theta)^T\boldsymbol{t}^{[k+1]}-\log a(\Theta)\right] \tag{2.14}$$

Since the value of $a=1$ the log of that term becomes zero.

Applying equation 2.12 to equation 2.13 results in four values:

$$t_1^{[k+1]}=E\left[\Sigma_{i=1}^{n}\boldsymbol{I}(z_i=0)\middle|y_i,\Theta^k\right]=\Sigma_{i=1}^{n}\left[\frac{\frac{p}{\sqrt{(2\pi\sigma^2)}}e^{\frac{-y_i^2}{2\sigma^2}}}{\frac{p}{\sqrt{(2\pi\sigma^2)}}e^{-y_i^2}+\frac{1-p}{2\mathrm{b}}e^{\frac{-|y_i|}{b}}}\right], \tag{2.15}$$

$$t_2^{[k+1]}=E\left[\Sigma_{i=1}^{n}\boldsymbol{I}(z_i=0)\,y_i^2\middle|y_i,\Theta^k\right]=\Sigma_{i=1}^{n}y_i^2\left[\frac{\frac{p}{\sqrt{(2\pi\sigma^2)}}e^{\frac{-y_i^2}{2\sigma^2}}}{\frac{p}{\sqrt{(2\pi\sigma^2)}}e^{-y_i^2}+\frac{1-p}{2\mathrm{b}}e^{\frac{-|y_i|}{b}}}\right], \tag{2.16}$$

$$t_3^{[k+1]}=E\left[\Sigma_{i=1}^{n}\boldsymbol{I}(z_i=1)\middle|y_i,\Theta^k\right]=\Sigma_{i=1}^{n}\left[\frac{\frac{1-p}{2\mathrm{b}}e^{\frac{-|y_i|}{b}}}{\frac{p}{\sqrt{(2\pi\sigma^2)}}e^{-y_i^2}+\frac{1-p}{2\mathrm{b}}e^{\frac{-|y_i|}{b}}}\right], \tag{2.17}$$

$$t_4^{[k+1]} = E\left[\Sigma_{i=1}^{n} \boldsymbol{I}(z_i=1)\,|\,y_i \middle|\, y_i, \Theta^k\right] = \Sigma_{i=1}^{n} |y_i| \left[\frac{\frac{1-p}{2\mathrm{b}} e^{\frac{-|y_i|}{b}}}{\frac{p}{\sqrt{(2\pi\sigma^2)}} e^{-y_i^2} + \frac{1-p}{2\mathrm{b}} e^{\frac{-|y_i|}{b}}} \right]. \tag{2.18}$$

Since the values of $\boldsymbol{y}$ are known, and the values of σ, b, and p are the estimates from the initial condition or the previous iteration of the algorithm, $\boldsymbol{t}$ becomes a row vector of four values.

Applying equations 2.11 and 2.12 to equation 2.14, the M-step becomes:

$$\begin{aligned} \Theta^{[k+1]} = \underset{\Theta}{argmax} \Bigg(& t_1^{[k+1]} \log\left(\frac{p}{\sqrt{(2\pi\sigma^2)}} \right) + t_2^{[k+1]} \frac{-1}{2\sigma^2} \\ & + t_3^{[k+1]} \log\left(\frac{1-p}{2\mathrm{b}} \right) + t_4^{[k+1]} \frac{-1}{b} \Bigg). \end{aligned} \tag{2.19}$$

In order to find the values of the parameters σ, b, and p, the derivative of the contents of the *argmax()* needs to be calculated and equated to zero. For the value of p this calculation resulted in:

$$\frac{d}{dp}\left(t_1^{[k+1]} \log\left(\frac{p}{\sqrt{(2\pi\sigma^2)}} \right) + t_2^{[k+1]} \frac{-1}{2\sigma^2} + t_3^{[k+1]} \log\left(\frac{1-p}{2\mathrm{b}} \right) + t_4^{[k+1]} \frac{-1}{b} \right) = 0 \tag{2.20}$$

and because two of the terms do not contain p this simplifies to:

$$\frac{d}{dp}\left(t_1^{[k+1]} \log\left(\frac{p}{\sqrt{(2\pi\sigma^2)}} \right) + t_3^{[k+1]} \log\left(\frac{1-p}{2\mathrm{b}} \right) \right) = \left(t_1^{[k+1]} \frac{1}{p} + t_3^{[k+1]} \frac{-1}{1-p} \right) = 0 \tag{2.21}$$

and results in:

$$p_{EST} = \frac{t_1^{[k+1]}}{t_1^{[k+1]} + t_3^{[k+1]}} \tag{2.22}$$

where the values of $t_1^{[k+1]}$ and $t_3^{[k+1]}$ are the values found by the E-step. Similarly the derivative with respect to σ was performed:

$$\frac{d}{d\sigma}\left(t_1^{[k+1]}\log\left(\frac{p}{\sqrt{(2\pi\sigma^2)}}\right)+t_2^{[k+1]}\frac{-1}{2\sigma^2}+t_3^{[k+1]}\log\left(\frac{1-p}{2\mathrm{b}}\right)+t_4^{[k+1]}\frac{-1}{b}\right)=0 \qquad (2.23)$$

and because two of the terms do not contain σ this simplifies to:

$$\frac{d}{d\sigma}\left(t_1^{[k+1]}\log\left(\frac{p}{\sqrt{(2\pi\sigma^2)}}\right)+t_2^{[k+1]}\frac{-1}{2\sigma^2}\right)=\frac{t_1^{[k+1]}}{\sigma^3}+\frac{t_2^{[k+1]}}{\sigma}=0 \qquad (2.24)$$

and results in:

$$\sigma_{EST}=\sqrt{\frac{t_2^{[k+1]}}{t_1^{[k+1]}}}. \qquad (2.25)$$

The derivative with respect to b was also performed:

$$\frac{d}{db}\left(t_1^{[k+1]}\log\left(\frac{p}{\sqrt{(2\pi\sigma^2)}}\right)+t_2^{[k+1]}\frac{-1}{2\sigma^2}+t_3^{[k+1]}\log\left(\frac{1-p}{2\mathrm{b}}\right)+t_4^{[k+1]}\frac{-1}{b}\right)=0 \qquad (2.26)$$

and because two of the terms do not contain b this simplifies to:

$$\frac{d}{db}\left(t_3^{[k+1]}\log\left(\frac{1-p}{2\mathrm{b}}\right)+t_4^{[k+1]}\frac{-1}{b}\right)=\frac{-t_3^{[k+1]}}{b}+\frac{t_3^{[k+1]}}{b^2}=0 \qquad (2.27)$$

and results in :

$$b_{EST}=\frac{t_4^{[k+1]}}{t_3^{[k+1]}}. \qquad (2.28)$$

The implementation of this algorithm is straightforward by first finding initial estimates of the three parameters σ_{INIT}, b_{INIT}, and p_{INIT}. These were made by finding the standard deviation of the given audio signal $\sigma_{audio}=std(\mathbf{y})$ and setting $\sigma_{INIT}=\frac{\sigma_{audio}}{10}$ and $b_{INIT}=\sqrt{2}\,\sigma_{audio}$ and $p_{INIT}=0.5$. These values were chosen because the breathing signal is assumed to be highest amplitude of the two signals. The standard deviation of the noise is assumed to be a factor of less than the standard deviation of the total sound signal. The

probability p_{INIT} was arbitrarily chosen assuming that the probability of being in a breath sound signal or a pause sound was equally likely.

The equations to find the E-step and resulting vector $\boldsymbol{t}^{[k+1]}$ were then calculated directly. The M-step was then calculated to find the new estimates of the parameters σ, *b*, and *p*. Because this is an iterative process the above steps were iterated until the convergence criteria were met. Convergence was achieved when the difference of five consecutive estimates of the parameters σ, b, and *p* were within 5% of the estimates of the values of σ, *b*, and *p*. Moon et al. [20] state that convergence may be determined in this way.

2.3.5.3 Finding the model parameters using the Gaussian estimation algorithm

The initial attempt at calculating the EM algorithm resulted in an iterative method to find the parameters of σ, *b*, and *p* in a similar manner to the EM algorithm. Because this method is not the EM algorithm, but is based on the intuitive definition of the EM algorithm in finding the standard deviation of the noise of the signal it will be referred to as the Gaussian estimation (GE) method. This method is being discussed because it was the method used in the final algorithm to detect breaths. As will be discussed in Section 2.3.5.5, the GE algorithm was desirable for its ability to more closely estimate the standard deviation of the noise portion of the signal even when the histogram of the breath sound signal was not a good fit to the Laplace distribution.

The GE algorithm used two primary steps in an iterative technique just as in the EM algorithm. Initial estimates of σ, *b*, and *p* were found in the same way as for the EM algorithm. Using the estimates for σ, *b*, and *p* the expected distribution of each sample of

the input vector $\boldsymbol{y}$ was calculated by Bayesian estimation. The distribution mixture used for this model is:

$$f(\boldsymbol{x}|\Theta)=\frac{p}{\sqrt{2\times\pi\,\sigma^2}}e^{\frac{-x^2}{2\sigma^2}}+(1-p)\frac{1}{2\mathrm{b}}e^{\frac{-|x|}{b}}. \tag{2.29}$$

This model is made of two primary components. The first is the Gaussian distribution:

$$f_{XG}(x)=\frac{1}{\sqrt{(2\pi\,\sigma^2)}}e^{\frac{-x^2}{2\sigma^2}} \tag{2.30}$$

and the Laplace distribution:

$$f_{XL}(x)=\frac{1}{2\mathrm{b}}e^{\frac{-|x|}{b}}. \tag{2.31}$$

Applying Bayesian estimation to the Gaussian portion of the signal results in the equation:

$$P\left[z_i=0\right]=\frac{p_{EST}\times f_{XG}(y_i|\sigma_{EST})}{p_{EST}\times f_{XG}(y_i|\sigma_{EST})+(1-p_{EST})f_{XL}(y_i|b_{EST})} \tag{2.32}$$

where z_i is unobserved parameter that the sample y_i is from the Gaussian distribution, and σ_{EST}, b_{EST}, and p_{EST} are the current estimates of those parameters. If this probability was greater than 0.5, the sample y_i was placed in the Gaussian group of samples $\boldsymbol{G}_{VEC}$, otherwise it was placed in the Gaussian group $\boldsymbol{L}_{VEC}$.

After all of the values were sorted into the vectors of $\boldsymbol{G}_{VEC}$ and $\boldsymbol{L}_{VEC}$, the estimates of the parameters of σ, b, and p were found. The value of σ_{EST} was calculated by finding the maximum likelihood (ML) estimate of the Gaussian distribution. This is done by taking the derivative of the log of the distribution and equating it to zero. In the case of the Gaussian distribution this is calculated by:

$$\frac{\delta}{\delta\,\sigma}\log\left(\frac{1}{\sqrt{(2\pi\,\sigma^2)}}e^{\frac{-x^2}{2\sigma^2}}\right)=\frac{\delta}{\delta\,\sigma}\left(\log\left(\frac{1}{\sqrt{(2\pi\,\sigma^2)}}\right)-\frac{x^2}{2\,\sigma^2}\right)=\frac{-1}{\sigma}+\frac{x^2}{\sigma^3}=0 \tag{2.33}$$

and is then simplified further by calculating the mean of the vector $\boldsymbol{x}^2$ and has equation:

$$\sigma^2 = \frac{1}{N} \Sigma_{n=0}^{N-1} \boldsymbol{x}^2. \tag{2.34}$$

The input vector $\boldsymbol{x}$ of equation 2.34 is the vector of samples of the overall input signal that are determined to be in the Gaussian portion of the distribution as calculated by equation 2.32.

The ML estimate of the Laplace distribution is calculated in a similar manner, by finding the derivative of the log of the distribution and equating it to zero, and is calculated by:

$$\frac{\delta}{\delta b} \log\left(\frac{1}{2\mathrm{b}} e^{\frac{-|\boldsymbol{x}|}{b}}\right) = \frac{\delta}{\delta b}\left(\log\left(\frac{1}{2\mathrm{b}}\right) - \frac{|\boldsymbol{x}|}{b}\right) = \frac{-1}{b} + \frac{|\boldsymbol{x}|}{b^2} = 0 \tag{2.34}$$

and is then simplified to:

$$b = \frac{1}{N} \Sigma_{n=0}^{N-1} |\boldsymbol{x}|. \tag{2.35}$$

The implementation of the GE algorithm was performed as follows. Figure 2.9 shows a block diagram of the general steps of the algorithm.

The first step was performed in the following manner. The initial estimates of the three parameters were used to estimate which samples belonged more to the Gaussian or Laplace distribution using equation 2.32.

The second step was performed using the ML estimate from the two groups of samples as described above. The ML estimate of the standard deviation of the Gaussian signal was calculated as $\sigma_{EST} = \left(\frac{1}{K} \Sigma_{k=1}^{K} \boldsymbol{G}_{\boldsymbol{VEC}}^{\boldsymbol{2}}\right)^{\frac{1}{2}}$ where K is the number of samples in $\boldsymbol{G}_{\boldsymbol{VEC}}$. The ML estimate of the parameter of the Laplace distribution was calculated as

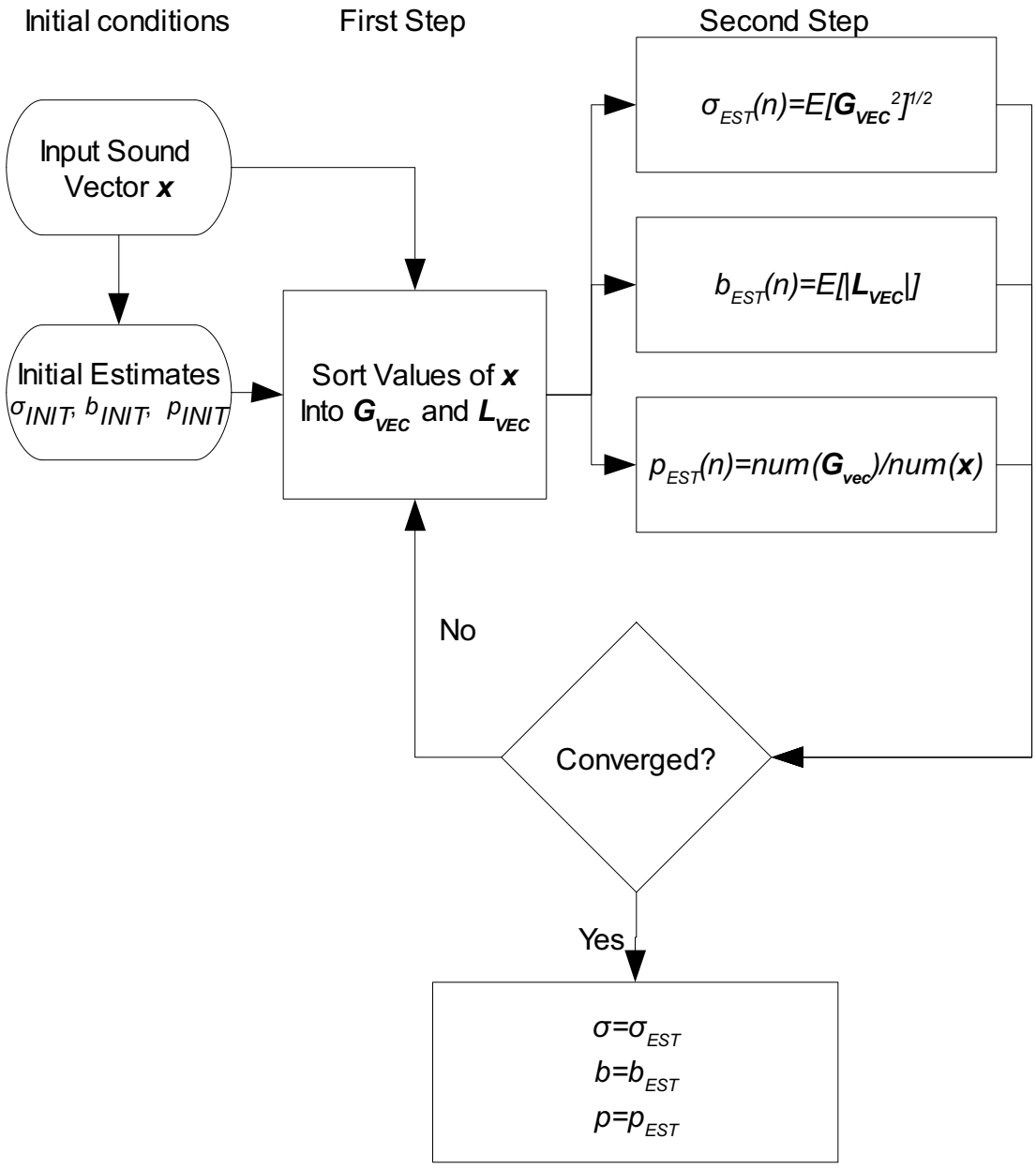


Figure 2.9 Block diagram of the GE algorithm to estimate the parameters of σ, b, and p.

$b_{EST} = \frac{1}{L}\Sigma_{l=1}^{L}|\boldsymbol{L}_{VEC}|$ where L is the number of samples in $\boldsymbol{L}_{VEC}$. The value of p was estimated by taking the ratio of the number of samples in the Gaussian group to the overall number of samples.

The E-step and M-step were then iterated until all of the parameters converged to a value. Convergence for this algorithm was defined when the estimate of the parameter changed by less than 5% of its value four consecutive times and was nonmonotonic. When the iterations converged for all parameters the estimates were used as the measured values of p, σ, and b.

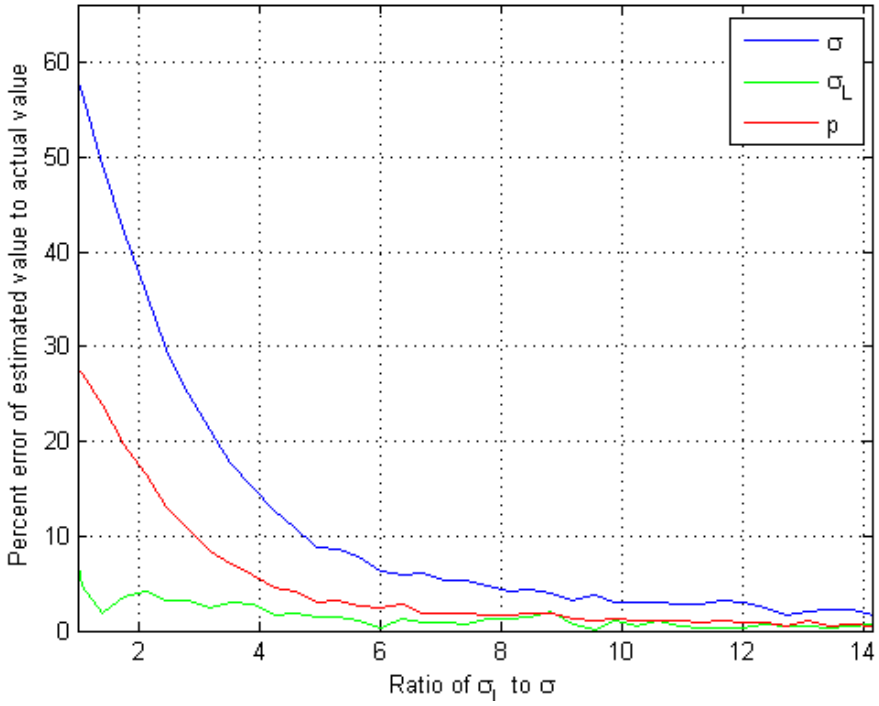
2.3.5.4 Simulation testing the accuracy of the GE and EM algorithm

In order to determine the accuracy of the described GE and EM algorithms, a simulation of a Gaussian-Laplace mixture random signal was created and then measured using the GE and EM algorithms. The simulation was made by providing the three parameters of p, σ, and b and using them to create a vector of data. The data consisted of two parts; the first was a pseudo-Gaussian signal produced from a random number generator with standard deviation σ that contained $N \times p$ values, where N is the number of values in the vector. This was concatenated with an additional string of $N \times (1-p)$ values of a Laplace distributed signal with standard deviation $\sigma_L = \frac{\sqrt{2}\,b}{2}$. The value of p was held constant at 0.5. This simulation was also performed for values of p other than 0.5 as will be described later. Because both the GE and EM algorithms calculate the expected distribution of each sample without respect to order, the order of the concatenation of the input vector does not affect the estimates of the parameters.

Using the simulated Gaussian-Laplace mixture signal, the value of σ was set to 1 while the value of b was varied from 1.4 in steps of 0.1 to a value of 28. This range was chosen to make the ratio of the standard deviation of the Laplace distribution to the Gaussian distribution from 1 to 20 . The percent errors of p_{EST}, σ_{EST}, and b_{EST} to the true values of p, σ, and b were then calculated and plotted in Figure 2.10a for the GE algorithm and 2.10b for the EM algorithm. These plots show that for the estimates from the GE algorithm to be within 10% of the actual value, the ratio of σ_L to σ must be greater than 5. The ratio can be less than this but the estimates will have a higher margin of error. For the EM algorithm the ratio needs to be greater than 3.5 to achieve the same accuracy.

This simulation was calculated for different values of p. This was done as p was varied from 0 to 1 in steps of 0.1 as the ratio of σ_L to σ was varied from 1 to 20 as performed above. The percent errors of p_{EST}, σ_{EST}, and b_{EST} to the true values of p, σ, and b were then calculated. The percent errors that were less than 10% are shown in white, the percent errors that were between 10% and 20% are shown in gray, and the percent errors greater than 20% are shown in black. These values were calculated using both the EM and GE algorithms. Figure 2.11a shows the percent error in estimating the parameter σ using the EM algorithm. Figure 2.11b shows the percent error in estimating the parameter σ using the GE algorithm. Figure 2.11c shows the percent error in estimating the parameter b using the EM algorithm. Figure 2.11d shows the percent error in estimating the parameter b using the GE algorithm. Figure 2.11e shows the percent error in estimating the parameter p using the EM algorithm. Figure 2.11f shows the percent error in estimating the parameter p using the GE algorithm. From these figures it is evident that the EM algorithm estimates the parameters within 10% of the actual values for a wider

a



b

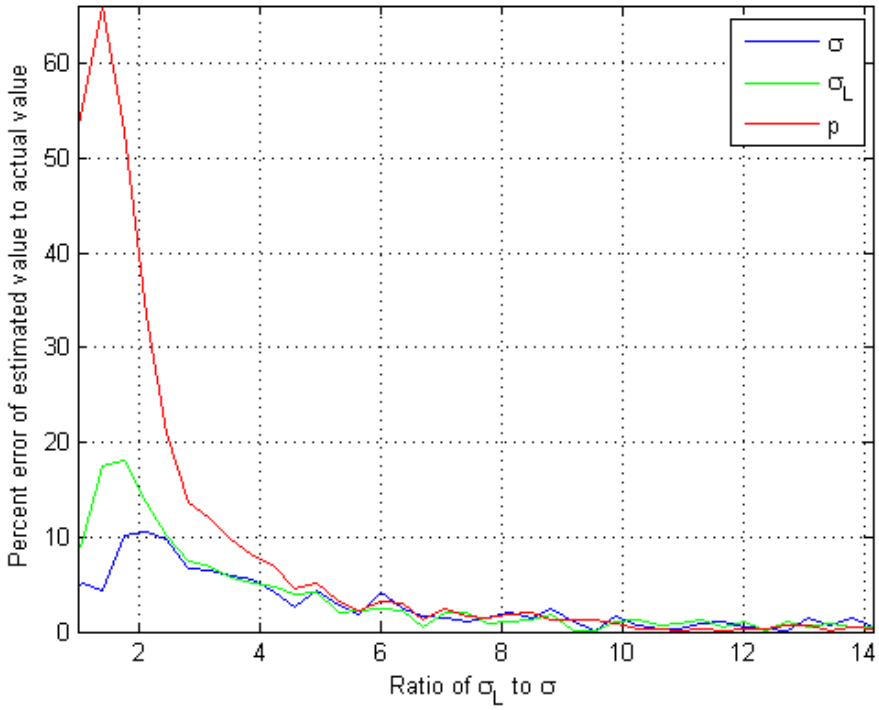
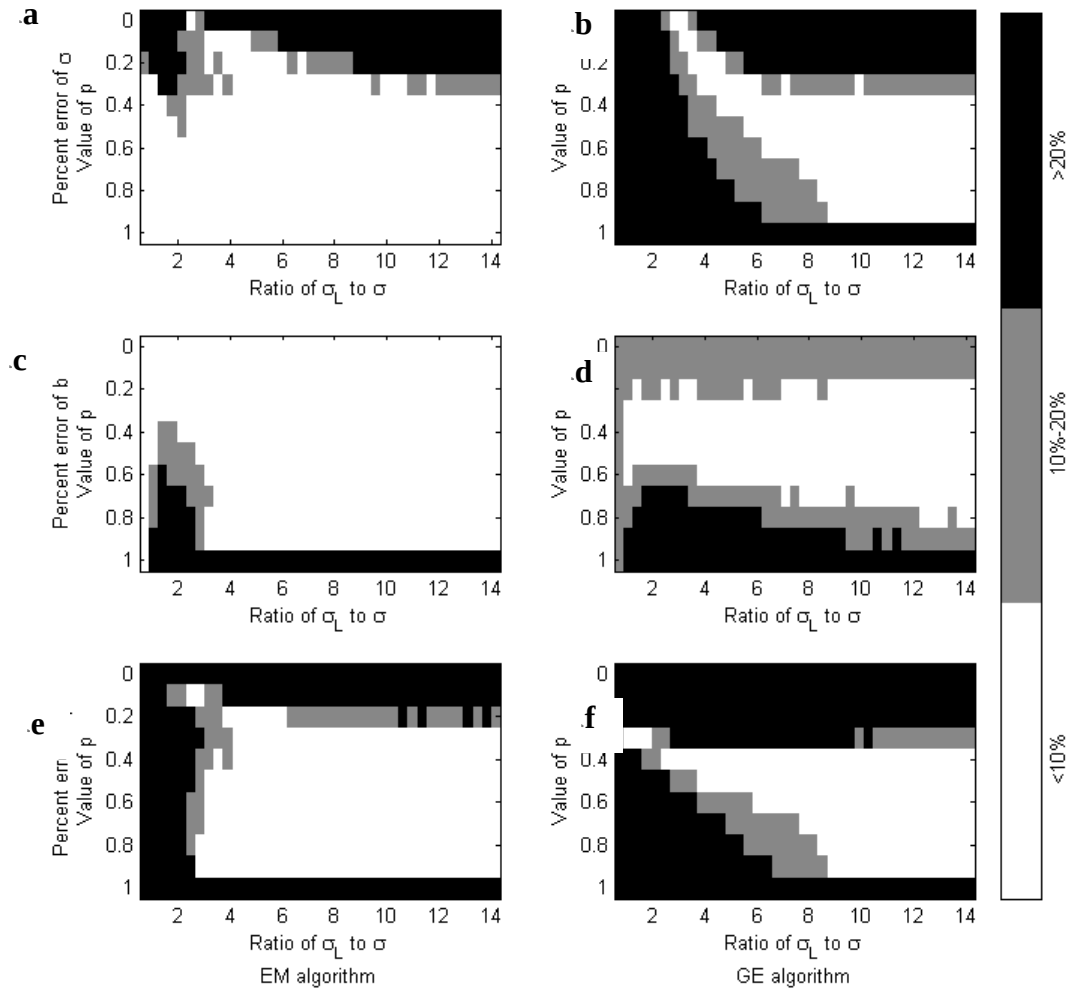


Figure 2.10 Percent error of the estimate to the actual value when calculating the parameters σ, b, and p from a simulated model (a) for the GE algorithm (b) for the EM algorithm.



- Figure 2.11 Error of the estimates of parameters of σ, b and p using the EM and GE algorithms (a) EM algorithm estimating σ (b) GE algorithm estimating σ (c) EM algorithm estimating b (d) GE algorithm estimating b (e) EM algorithm estimating p (f) GE algorithm estimating p.

range of values of the ratio of σ_L to σ and p. The EM algorithm has the most difficulty estimating the parameter of σ when the value of p is close 0, meaning that few of the samples in the simulated input were generated by a Gaussian process. Similarly the EM algorithm estimates the value of b with the highest amount of error when the value of p is close to 1, meaning that few of the samples in the simulated input were generated by the Laplace process.

Using the GE algorithm the estimate of the standard deviation of the Gaussian distribution is more accurately estimated when p is between 0.3 and 0.9. Similarly the value of b is more accurately estimated when p is between 0 and 0.7. For the GE algorithm to estimate the values of p, σ, and b with 10% accuracy does require tighter tolerances for the input data. Since the parameters of the input data cannot be chosen this means an acceptance of a higher error for signals with a low ratio of σ_L to σ for the GE algorithm. Since the parameters of the input stream are unknown this raises some question if the GE algorithm is suitable for this application. Despite this, the GE algorithm was considered because the estimated values of the standard deviation of each distribution would be limited to the standard deviation of the input data. This might affect the threshold measurement, but because the signal would be in a period of detected sounds, the variability in the amplitude of the detected sounds would ensure a threshold that would detect sounds. Periods in which p is very high would be consistent with a period of pause in breathing. This period would measure the value of σ with a higher amount of accuracy than that of b. The error that would occur in this situation would be the detection of signals during a period of apnea because the threshold was too low. A method to ensure that the noise threshold did not fall too low was created and is discussed in the

next section.

2.3.5.5 Choice of parameter estimation algorithm

The EM algorithm performed much better than the GE algorithm in simulation when the input data were known to be Gaussian and Laplace distributed signals. The estimation of these parameters was performed on data from the data set described. Figure 2.12 shows the audio envelope of a typical period of breathing with an anomalous breath sound with a much higher amplitude than the rest. The moving window parameters estimated from both the GE and EM algorithms are shown. The GE algorithm estimates the standard deviation of the noise with little variation despite the disturbance by the anomalous breath sound in the detected signal. The EM algorithm shows that the estimate of the standard deviation of the noise jumps significantly due to the disturbance of the detected signal caused by the anomalous breath sound. This is likely due to the fact that the Laplace distribution is not a good model for the histogram of the detected signal during this period. Because the signal attributed to breathing is not a Laplace distributed signal the EM algorithm cannot accurately estimate the standard deviation of the noise.

Because the amplitude of breathing signals are unpredictable and anomalous sounds were observed to be frequent, the EM algorithm may cause detections of apnea when there is none, causing more frequent false positive alarms. The GE algorithm appears to be able to estimate the standard deviation of the noise with a higher level of consistency despite sudden changes of the detected signal. For these reasons the GE algorithm was used to estimate the parameters for the breath detection algorithm.

The GE algorithm was used to fit the mixture distribution model to the histogram in Figure 2.7. The estimates for this data were $p=0.44$, $\sigma=3.58\times10^{-4}$, and $b=2.9\times10^{-3}$.

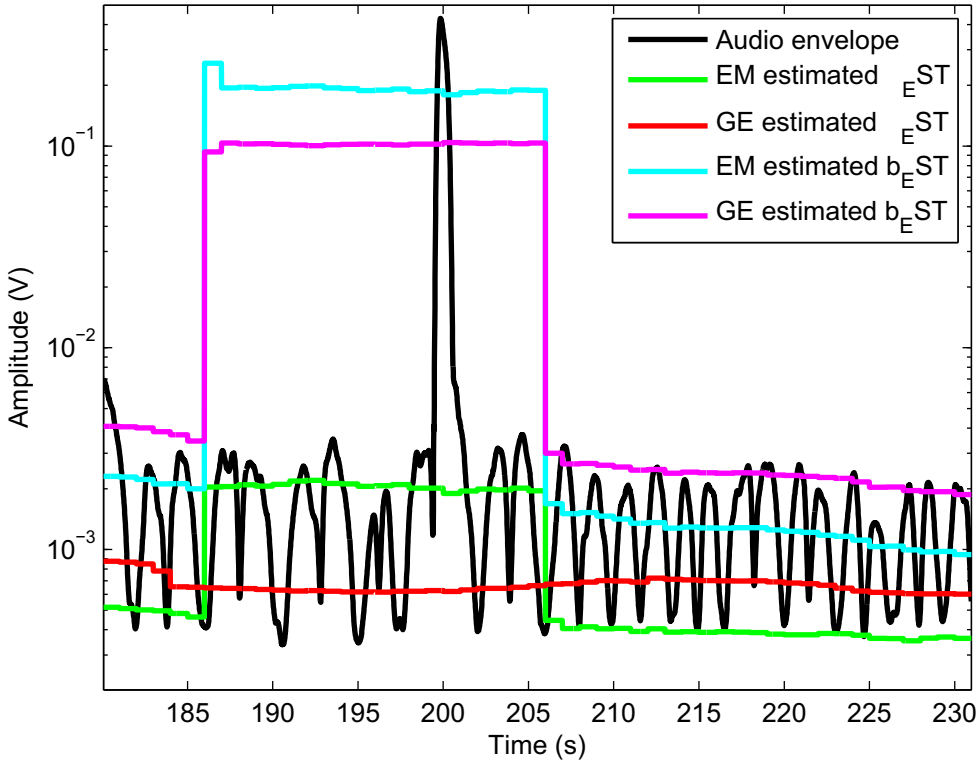


Figure 2.12 Comparison of the GE algorithm to the EM algorithm in estimating parameters for a segment of audio.

The distribution model was then plotted over the histogram of the data in Figure 2.7 using these estimates.

2.3.6 Using the model parameters to find an audio threshold

The threshold to decipher breath sounds from the noise can be determined using just the parameter σ alone or by using it in conjunction with the value of b. Using σ alone, a confidence interval of 96% can be achieved if the threshold is set to two times standard deviation of the noise.

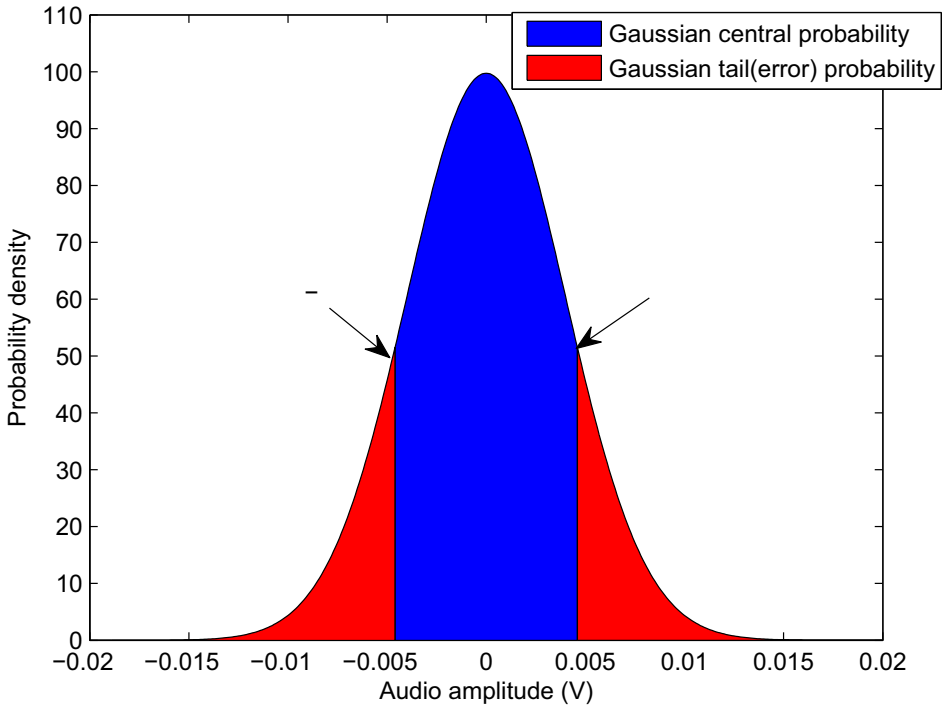
Using both parameters, estimation theory can be used to determine threshold. In order to minimize risk and maximized detection the minimax algorithm was used to determine the threshold. The minimax algorithm makes no assumptions about the prior probability of either distribution [21]. This is similar to a method in voice detection used in [19].

Using the minimax algorithm, the error from both distributions needed to be minimized. The tail probability e_{NF} of the Gaussian distribution is

$e_{NF} = \int_{\tau}^{\infty} \frac{1}{\sqrt{(2\pi\sigma)}} e^{\frac{-x^2}{2\sigma^2}} dx$, where τ is the threshold between the two distributions. Only the positive tail of the distribution has been calculated because both the zero-mean Gaussian distribution is symmetric about zero. Figure 2.13a shows the error probabilities attributed to the tail probabilities in red. The probability e_{BS} of the Laplace distribution is

$e_{BS} = \int_{0}^{\tau} \frac{1}{2b} e^{\frac{-|x|}{b}} dx$ where the error is attributed to the central probability. Because this distribution is symmetric about zero only the positive half needs to be calculated. The

a



b

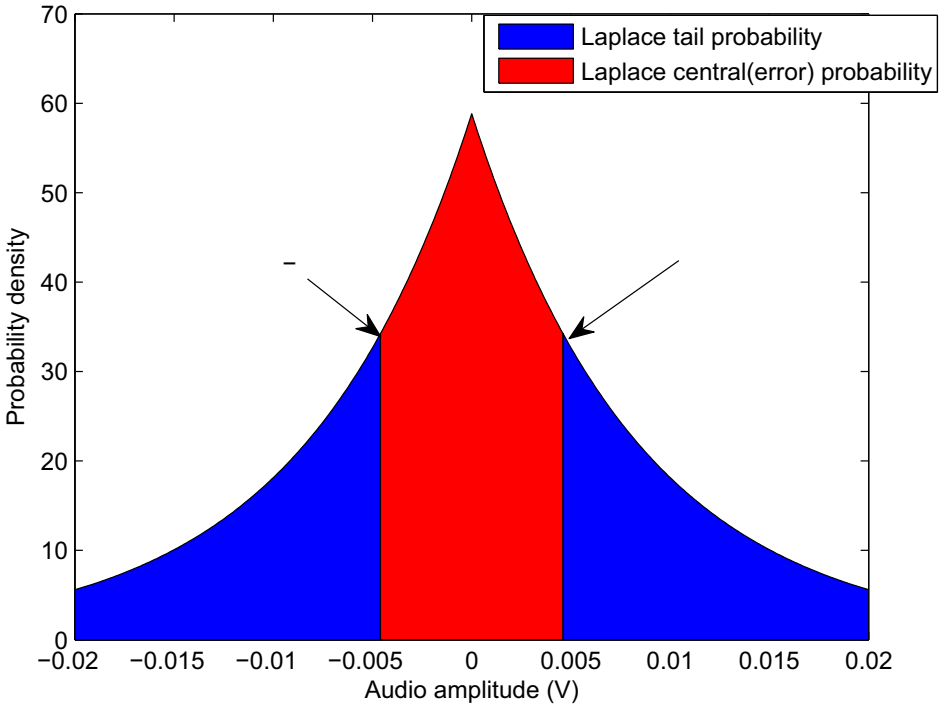


Figure 2.13 Probability distribution function showing error probabilities (a) Gaussian with tail error probabilities (b) Laplace with central error probability

error of the Laplace distribution is illustrated in Figure 2.13b where the error is the sum of the central probabilities. The minimax algorithm attempts to minimize the tail probability e_{NF} and probability e_{BS} by equating them and solving for τ. Equating the two errors results in the equation:

$$\int_{\tau}^{\infty} \frac{1}{\sqrt{(2\pi\sigma)}} e^{\frac{-x^2}{2\sigma^2}} dx = \int_{0}^{\tau} \frac{1}{2\mathrm{b}} e^{\frac{-|x|}{b}} dx \tag{2.36}$$

which can be simplified to

$$-erf\left(\frac{-\tau}{\sigma}\right) = e^{\frac{-\tau}{b}} \tag{2.37}$$

where *erf( )* is defined as the error function with equation $erf(x) = \frac{2}{\sqrt{(\pi)}} \int_{0}^{x} e^{-t^2} dt$.

An analytic solution for τ in equation 2.37 is not possible because the equation is already in its simplest form and is transendental. Instead of finding a direct value for τ, a parameter estimation technique was used to find a reasonable value for τ. A vector of values which would be common to τ was generated as will be described. The vector created, called τ_f, ranged from 0 to 1 in steps of 0.0001 with units of volts. This range and stepsize was chosen because the known range of the audio envelope is below 1 and the lowest noise observed was 0.0005. The values of τ_f were applied to the equations $-erf\left(\frac{-\tau}{\sigma}\right)$ and $e^{\frac{-\tau}{b}}$ using the estimated values σ_{EST} and b_{EST}. This resulted in two vectors of equal length. The difference of the two vectors was found and the value of τ_f with the absolute minimum error was chosen as the estimate for τ. The values of the two vectors calculated crossed for the values of τ_f because the error probability for the Gaussian distribution is zero for $\tau_f=0$ and the error probability for the Laplace distribution is zero for $\tau_f=1$.

Using the range of values for τ_f the vector for the error of the Gaussian distribution and the error of the Laplace distribution meet between 0 and 1 given the limitation that the input data is limited to be between -1 and 1.

The value for τ_f is a good estimate for the noise threshold in most circumstances. However, in periods in which there are no detectable signals, the estimate of the standard deviation of the Laplace distribution becomes similar to the standard deviation of the Gaussian distribution. Thus the assumption that $\sigma << b$ is not valid during long periods of silence. Using the value of τ in this circumstance would allow for a lower breath detection threshold, which could possibly start detecting noise sounds as breaths. Also during times of extremely loud sounds the value of b dominates the inequality thus making it possible to raise the threshold above the envelope of some of signals that should be detected. In order to avoid these two problems the value of τ is limited by $1.44 \times \sigma_{EST}$ on the lower end and $2.4 \times \sigma_{EST}$ on the upper end. These values were chosen as values with confidence intervals of 85% and 98% of the Gaussian distribution.

In order to determine the validity of the noise threshold τ chosen, a receiver operating characteristic (ROC) curve was generated from experimental data for the detection of apnea found by the acoustic detection method and that of the gold standard. The determination of apnea periods for the acoustic detection method and the gold standard will be described in later sections. For this curve, the value of the noise threshold was multiplied by a value from a vector g_a that included values from 0.5 to 10 in steps of 0.1. The probability of false alarm (p_{FA}) and probability of miss (p_{MISS}) were calculated from the training data set for each value of g_a, with the resulting ROC curve shown in Figure 2.14. The optimal value of g_a was 1.0 and it occurred where the value of $p_{FA}=0.094$ and

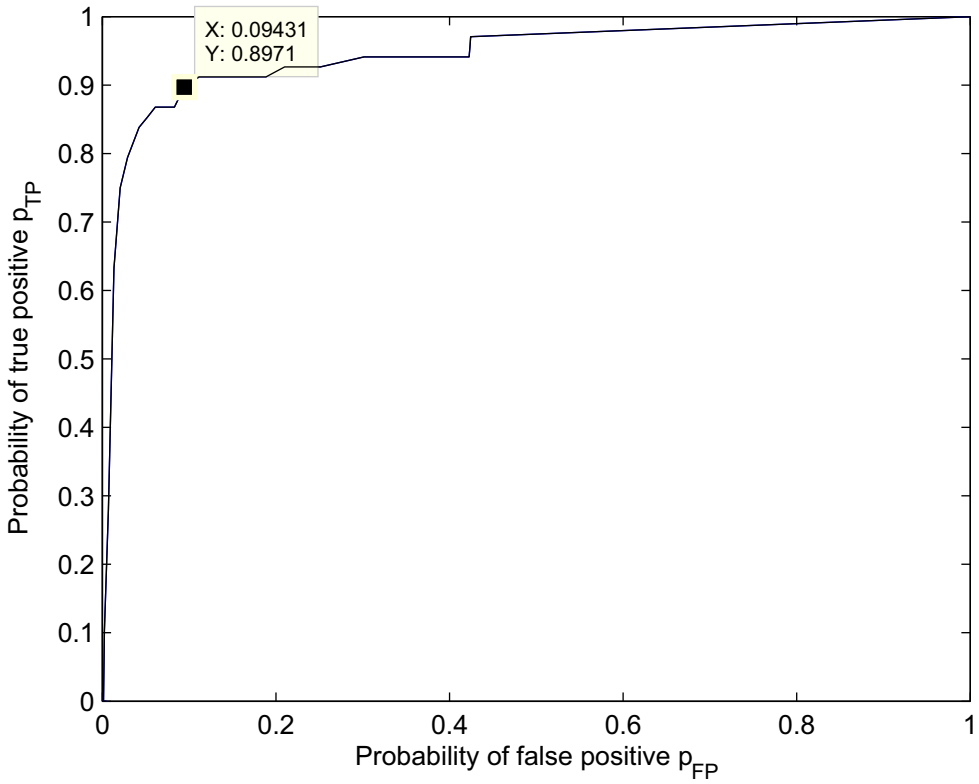


Figure 2.14 Receiver operating characteristics of acoustic breath detection to flow meter breath detection.

the value of $p_{MISS}=0.103$. The value of p_{FA} is shown directly on the plot of Figure 2.14, and the value of p_{MISS} is the value $p_{MISS}=1-p_{TP}$ from Figure 2.14.

2.3.7 Using the acoustic threshold to detect breath sounds

Using the threshold value and audio envelope as described above, crossings of the audio envelope through the breath detection threshold were found. A crossing was defined when two consecutive samples of the audio envelope were found in which one was less than the breath detection threshold and the other was greater than the breath detection threshold. Each crossing in which the slope of the envelope was positive indicated the beginning of a detected sound, and each crossing in which the slope of the envelope was negative indicated the end of a detected sound. Detected sounds shorter than 0.3 seconds were discarded. The value of 0.3 seconds was calculated from the maximum respiratory rate of a spontaneously breathing adult of 75 breaths per minute [22]. A rate of 75 breaths per minute means that each breath would be 0.8 seconds in length. Allowing for a 0.2 second expiratory pause and assuming equal inspiratory and expiratory times, the minimum breath sound length is 0.3 seconds. The remaining sounds were considered valid breaths.

Note that the the estimated parameters used to find the threshold were not derived from the envelope. It was assumed that because both the parameters and the envelope were directly derived from the filtered audio signal, they could be used together.

2.3.8 Detection of apnea by the gold standard

2.3.8.1 The need for multiple forms of apnea detection

The CosmoII+ monitor measured the flow rate directly in units of l/min. This monitor also marked the beginning of each breath with a marker. Using the breath marker and the flow rate, a breath volume signal was calculated by the CosmoII+ device for each breath in units of mL. The breath volume signal is simply the integral of the flow rate signal. The flow rate signal was not perfect and included some artifacts as described next. In addition, the flow rate signal occasionally had a DC offset that could be up to 5 l/min, as shown in Figure 2.15a. The DC offset in the flow rate signal caused the breath volume signal to integrate continuously, as shown in Figure 2.15b. Another problem with the breath volume signal was that the initial value of the volume was not known, causing a DC offset in the reported breath volume. The value of the breath volume was periodically reset by the CosmoII+ monitor, but variations in the initial breath volume were often observed. The largest artifact observed on the flow rate signal was caused as the heartbeat pushed air through the trachea in rates up to 10 l/min as seen in Figure 2.16. Other artifacts such as electronic noise were not as significant when measuring the flow rate signal.

Apnea has been defined in this work as a period of at least 15 seconds where the breath volume does not exceed 150 mL. The value of 150 mL was used because this is the average adult airway dead-space [7]. The example shown in Figure 2.15b shows the problem with this definition given the data set. Specifically this example shows the the case when volume could exceed the threshold of 150 mL while no real gas exchange occurs. For this reason the flow rate signal was used in addition to the breath volume

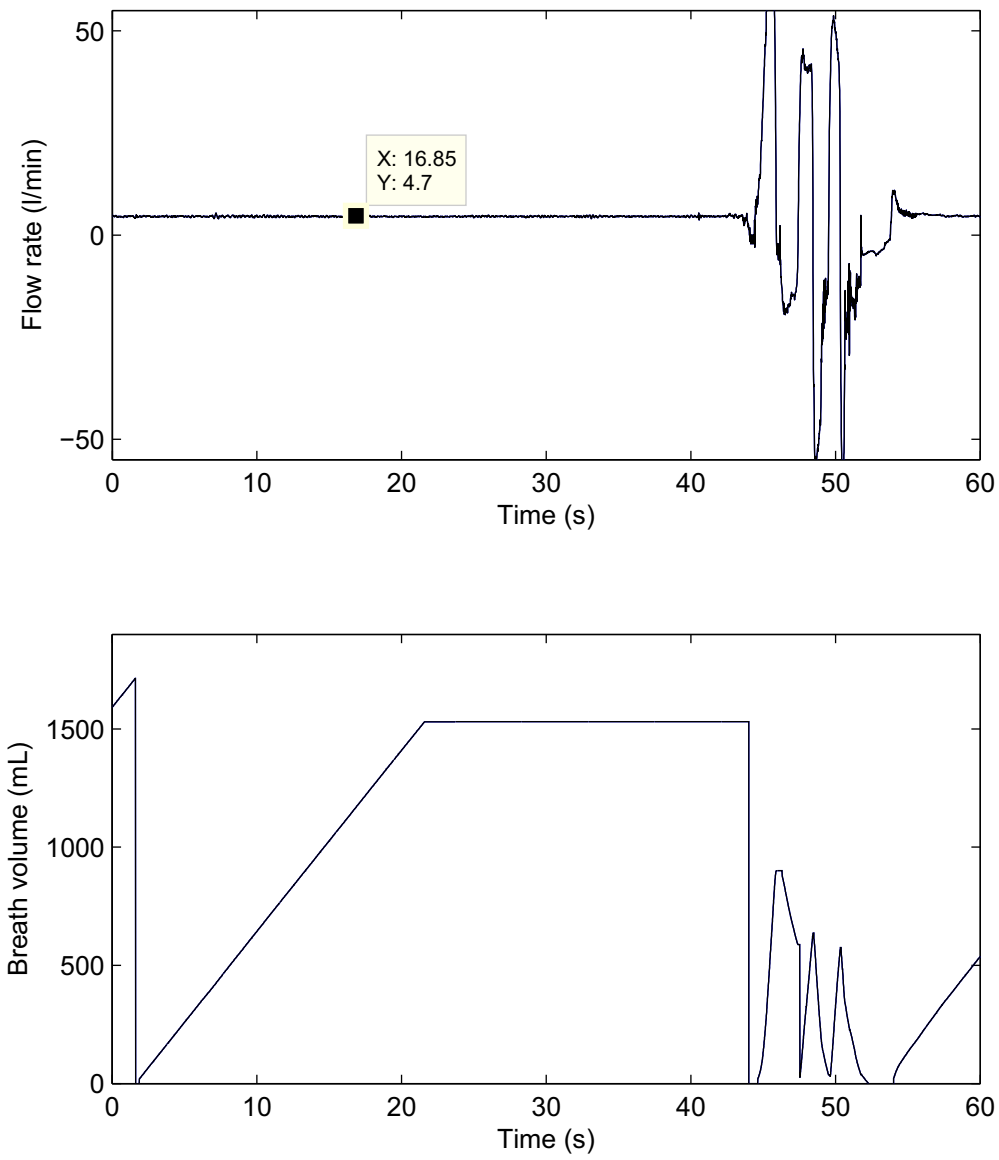


Figure 2.15 Example of the affect of a DC offset on the flow rate and the integrated breath volume, (a) Flow rate (l/min) (b) Flow Volume (mL).

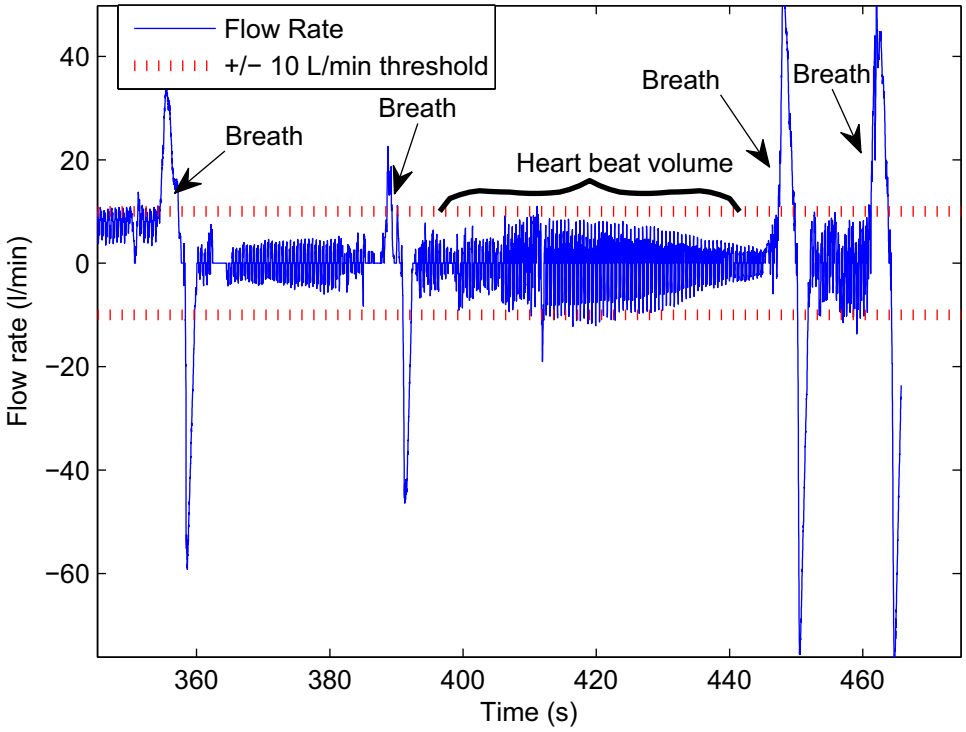


Figure 2.16 Example of a flow rate affected by a strong heartbeat flow.

signal in detecting periods of apnea. The threshold set on the flow rate signal was 7.5 l/min. This value was chosen to eliminate the majority of the heartbeat flow rate artifacts, and to detect periods of apnea even when a DC offset up to 7.5 l/min on the flow rate signal was present. More importantly, the peak value of 7.5 l/min corresponded to a tidal volume (peak breath volume) of 150 mL for the subjects in this study as shown in Figure 2.17. Figure 2.17 shows the comparison of the peak flow rate of a breath of the subjects and the tidal volume for that breath. The relationship on this plots shows the majority of the breaths with volume less than 150 mL having a peak flow rate of 7.5 l/min. The outliers observed in this plot are attributed to flow volumes calculated from a DC offset of the flow rate as shown in Figure 2.14b. Additionally this is close to the flow rate at which sound can be detected at the trachea of 7.2 l/min [10].

2.3.8.2 Flow rate apnea detection and breath volume apnea detection

Using the flow rate signal and breath marker, apnea was detected where the absolute flow rate was below the threshold of 7.5 l/min for longer than 15 seconds.

Using the breath volume signal, apnea was detected where the breath volume was below 50 mL for more than 15 seconds. The value of 50 mL was used because of the DC offset associated with the calculation of the breath volume signal. If the value of 150 mL were used some apnea would be detected where there actually was none because the initial volume of the breath was below -100 mL. The decision in this case was to err on the side of missing a period of apnea rather than detecting a period of apnea that did not exist.

If either of the apnea detection methods described above detected a period of apnea, that period was considered apnea. This was done to detect periods of apnea by the

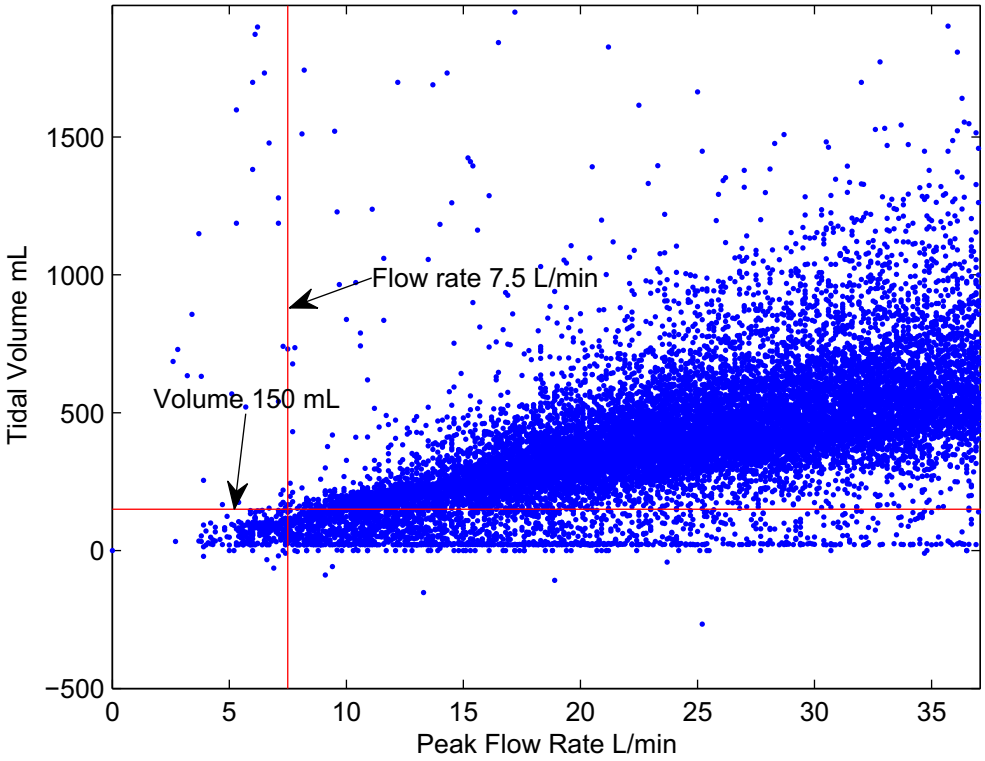


Figure 2.17 Peak flow rate compared to tidal volume for the training data set.

breath volume method when artifact on the flow rate signal caused it to miss, and to detect periods of apnea when the breath volume signal integrated a DC offset of the flow rate signal, but the flow rate signal detected the period of apnea.

Periods detected as apnea but determined to be times at which the tight-fitting-facemask was not present were eliminated from the data set, as will be described in Chapter 3.

2.3.9 Comparing the acoustic apnea detection to the flow apnea detection method

Apnea detected by the acoustic method and the flow method were compared directly as follows. First the periods of apnea detected by the flow method were compared to the acoustic method. If a period of apnea detected by the acoustic method overlapped a period detected by the flow method, this period was considered a true positive. If a period of apnea detected by the flow method did not overlap a period detected by the acoustic method, the period of apnea was considered a false negative. Next the periods of apnea detected by the acoustic method were compared to the periods detected by the flow method. If the acoustic method detected a period of apnea but the period did not overlap with a period detected by the flow method, the period was considered a false positive.

In order to calculate a value for the number of true negatives, the average length of all of the periods of apnea detected by the flow method was calculated. The amount of time found in which neither the acoustic apnea detection method or the flow apnea detection method detected apnea was divided by the average length of a period of apnea to find the equivalent number of periods of true negative apnea detections.

2.4 Results

The number of true positives, false negatives, false positives, and calculated true negatives of detected apnea were calculated for both the training data set and later the testing data set. The results of the training data set are shown in Figure 2.18. The results of the testing data set are shown in Figure 2.19. The data shown in the Figures 2.18 and 2.19 are summarized in Tables 2.1 and 2.2, respectively.

2.5 Discussion

The results shown indicate that the apnea detection by the acoustic method is relatively effective. The training data set had a 92.8% sensitivity and a 91.8% specificity. This is below the goal of 95% sensitivity and specificity as desired in the project proposal of this work. The testing set had a 98% sensitivity and a 100% specificity. The improvement of both specificity and sensitivity is likely due to the differences in the two data sets. The training set consisted of 21% apnea and the testing consisted of only 13% of apnea. Because the original data were randomly divided into two sets, the discrepancy in the percentage of apnea in each set can probably be attributed to the random selection process of the testing and training sets.

When the false positive apnea periods were observed, it was determined that a significant number were caused by the acoustic breath signals being almost inaudible in comparison to the noise signal. This observation took place as the investigator listened to sound segments during periods of apnea classified as a false negatives. The breath sounds recorded during these periods were faint and in some cases inaudible. This could have been because of poor placement of the stethoscope or because of an excess of noise

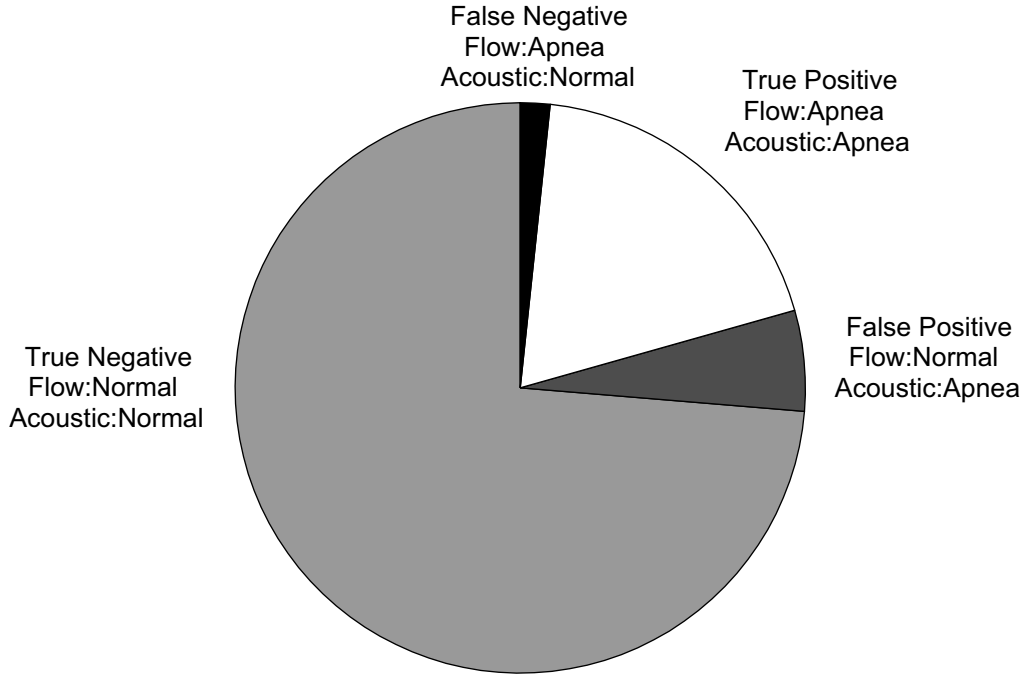


Figure 2.18 Results of the comparison of the acoustic method of apnea detection to the flow method of apnea detection from the training data set.

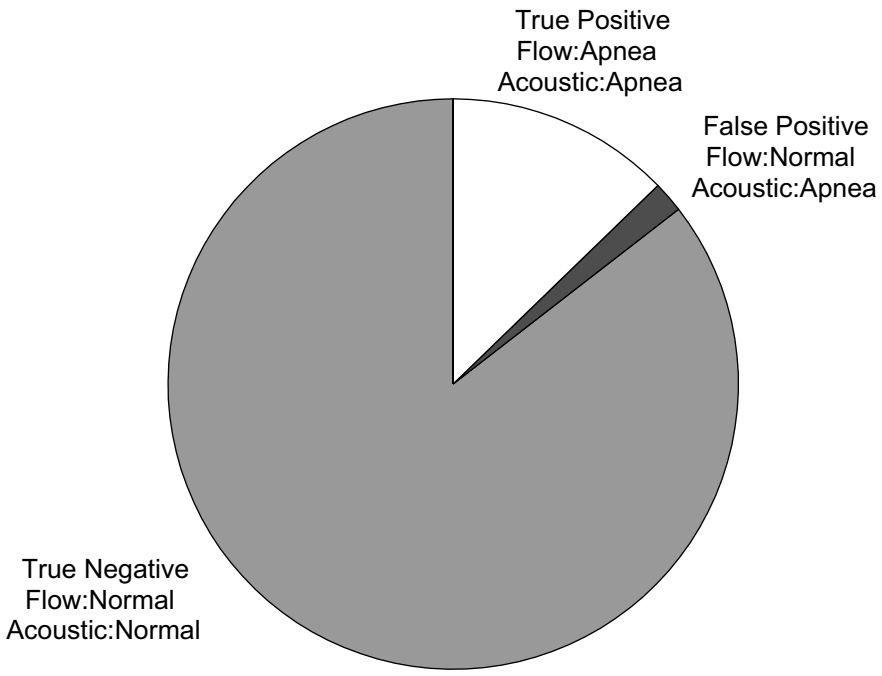


Figure 2.18 Results of the comparison of the acoustic method of apnea detection to the flow method of apnea detection from the testing data set.

Table 2.1 Training set apnea detection data comparing flow method and acoustic apnea detection.

		Flow Meter	
		Apnea	***Normal***
Acoustic	***Apnea***	575 TP=91.8%	173 FP=6.2%
	Normal	51 FN=8.2%	2238 TN=92.8%

Table 2.2 Testing set apnea detection data comparing flow method and acoustic apnea detection.

		Flow Meter	
		Apnea	***Normal***
Acoustic	***Apnea***	472 TP=100%	65 FP=2%
	Normal	0 FN=0%	3167 TN=98%

on the microphone. The probability of this happening could be reduced by identifying optimal places to place the microphone that would increase the amount of breath sound detected by the microphone.

The periods of false negative were caused primarily when a source of noise from outside of the stethoscope was detected as a breath sound. Depending on the placement of the stethoscope, some external sounds were able to be recorded within the stethoscope cup. This was determined when the researcher listened to the audio during the apnea periods identified as false negatives, and common ambient sounds were determined to be the sounds that were detected as a breath. A typical sound heard was the nurse assessing the OAA/S score. The nurses voice could be clearly heard in some instances. The number of periods of apnea missed by the acoustic method could be decreased in several ways. The first method would be to use an adaptive filter to reduce the ambient noise by placing a reference microphone outside the stethoscope cup. This method will be tested in Chapter 4. The other option to improve this is to classify the source of the sounds based on their acoustics.

Because of the relatively large number of periods of false negative from the training data set, the acoustic method alone would be a risky method to determine patient ventilation, especially in a noisy environment. The periods of false positive are not as dangerous, but do pose a threat of a clinician silencing the device if it repeatedly shows a false positive alarm. This method may be more useful in a clinical setting where the patient is not as closely monitored. During the data recording of this set, the subjects were not allowed to be apneic for long periods of time due to the close observation of the anesthesiologist present and the large amount of monitoring equipment. It may be that if

the subjects were allowed to have been apneic for longer periods of time, the number of periods of false negatives by the acoustic method would have gone down. This is because the periods of apnea would be longer, thus allowing the acoustic method to detect the periods of apnea. The periods of apnea would be longer in a clinical setting because the data collection setup during this study allowed for the anesthesiologist to give his entire attention to the subject and so periods of apnea were not as long, as opposed to a clinical setting where the patients are not monitored as closely by the physician. In addition, a clinical setting does not normally monitor respiration directly, but relies on the pulse-oximeter to determine apnea. As noted in the introduction the time difference between the onset of apnea and the alarm of the pulse-oximeter can be over two minutes.

It is difficult to compare the results of this work with those of other acoustic respiratory monitors because no other monitor tried to identify apnea. Other acoustic apnea monitors have only reported the success in identifying a respiratory rate [3].

2.6 References

[1] D.C. Morello, G.A. Colon, S. Fredricks, R.E. Iverson, and R. Singer, "Patient safety in accredited office surgical facilities.," *Plast Reconstr Surg*, vol. 99, May. 1997, pp. 1496-1500.

[2] H. Vila, R. Soto, A.B. Cantor, and D. Mackey, "Comparative outcomes analysis of procedures performed in physician offices and ambulatory surgery centers.," *Arch Surg*, vol. 138, Oct. 2003, pp. 991-995.

[3] G. Sierra, V. Telfort, B. Popov, L.G. Durand, R. Agarwal, and V. Lanzo, "Monitoring respiratory rate based on tracheal sounds. First experiences.," *Conf Proc IEEE Eng Med Biol Soc*, vol. 1, 2004, pp. 317-320.

[4] C. Guilleminault, F.L. Eldridge, F.B. Simmon, and W.C. Dement, "Sleep apnea syndrome. Can it induce hemodynamic changes?," *The Western Journal of Medicine*, vol. 123, Jul. 1975, pp. 7-16.

[5] C. Guilleminault, A. Tilkian, and W.C. Dement, "The sleep apnea syndromes," *Annual Review of Medicine*, vol. 27, 1976, pp. 465-484.

[6] D.E. Weese-Mayer, A.S. Morrow, L.P. Conway, R.T. Brouillette, and J.M. Silvestri, "Assessing clinical significance of apnea exceeding fifteen seconds with event recording," *The Journal of Pediatrics*, vol. 117, Oct. 1990, pp. 568-574.

[7] R. Woolmer and B. Lind, "Rebreathing with a semi-closed system.," *Br J Anaesth*, vol. 26, Oct. 1954, pp. 316-322.

[8] A. Yadollahi and Z. Moussavi, "Comparison of flow-sound relationship for different features of tracheal sound," *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2008*, 2008, pp. 805-808.

[9] V.P. Harper, H. Pasterkamp, H. Kiyokawa, and G.R. Wodicka, "Modeling and measurement of flow effects on tracheal sounds," *IEEE Journal of Biomedical Engineering*, vol. 50, 2003, pp. 1-10.

[10] A.L.L. Egnel and M. Paiva, *Gas Mixing and Distribution in the Lung*, Dekker, 1985.

[11] R. Beck, G. Rosenhouse, M. Mahagnah, R.M. Chow, D.W. Cugell, and N. Gavriely, "Measurements and theory of normal tracheal breath sounds.," *Ann Biomed Eng*, vol. 33, 2005, pp. 1344-1351.

[12] D.E. Olson, M. Bogyi, D.B. Schwartz, and J.R. Hammersley, "Relationship of tracheal breath sounds to airflow," *American Review of Respiratory Disease*, vol. 129, 1984.

[13] A. Borgeat and J. Aguirre, “Sedation and regional anesthesia,” *Current Opinion in Anaesthesiology*, vol. 22, Oct. 2009, pp. 678-682.

[14] M.J. Jaeger and A.B. Otis, “Measurement of airway resistance with a volume displacement body plethysmograph,” *Journal of Applied Physiology*, vol. 19, Jul. 1964, pp. 813-820.

[15] H.L. Watson, D.A. Poole, and M.A. Sackner, “Accuracy of respiratory inductive plethysmographic cross-sectional areas.,” *J Appl Physiol*, vol. 65, Jul. 1988, pp. 306-308.

[16] M.A. Sackner, H. Watson, A.S. Belsito, D. Feinerman, M. Suarez, G. Gonzalez, F. Bizousky, and B. Krieger, “Calibration of respiratory inductive plethysmograph during natural breathing.,” *J Appl Physiol*, vol. 66, Jan. 1989, pp. 410-420.

[17] N. Gavriely, Y. Palti, and G. Alroy, “Spectral characteristics of normal breath sounds.,” *J Appl Physiol*, vol. 50, Feb. 1981, pp. 307-314.

[18] A.C. Bovik, *Handbook of image and video processing*, Academic Press, 2005.

[19] S. Gazor and W. Zhang, “A soft voice activity detector based on a Laplacian-Gaussian model,” *IEEE Transactions on Speech and Audio Processing*, vol. 11, 2003, pp. 498-505.

[20] T.K. Moon, “The expectation-maximization algorithm,” *IEEE Signal Processing*, vol. 13, 1996, pp. 47-60.

[21] H.V. Poor, *An introduction to signal detection and estimation*, Birkhäuser, 1994.

[22] W.D. McArdle, F.I. Katch, and V.L. Katch, *Exercise physiology*, Lippincott Williams & Wilkins, 2007.

CHAPTER 3

SOUND CLASSIFICATION

3.1 Introduction

3.1.1 Goals of sound classification

Chapter 2 focused on using the audio signal to detect apnea when compared to the flow meter. This chapter focuses on determining the source of the apnea, whether it be from respiratory depression (RD) or respiratory obstruction (RO) based on the acoustic signal. The classification based on the acoustic signals is then compared to a standard that has been developed using the CosmoII+ flow meter and the Respitrace respiratory impedance plethysmography (RIP) device to determine the source of apnea.

The goal of the resarch work is to be able to determine the source of apnea detected with a 95% sensitivity and 95% specificity when compared to the standard. The standard will be described in Section 3.2.1.

3.1.2 Overview

This chapter will first discuss the current research into classification of breath sounds. This chapter will then discuss the basis and reasoning behind sound classification. It will also introduce the causes of RO and RD and the data set used to create and test the classifier. Section 3.2.1 then discusses the initial challenge of developing a standard to differentiate between RD and RO. Because of the lack of a recognized standard

that differentiates between RD and RO, a method using the classifications of three researchers as they viewed the flow rate and RIP data will be discussed. Section 3.2.2 discusses the proposed models used to classify the different sounds recorded at the trachea. The models were created using observations of features of each sound. The models were also created using explanations of the physical aspects of what is creating the sounds. Section 3.2.3 then describes the methods used to classify the sounds based on the models described. Finally Section 3.3 shows the results and comparison of the standard to the method of classification and Section 3.4 discusses the results and conclusions of this chapter.

3.1.3 Overview of breath sound classification techniques

The development of an algorithm to process tracheal sounds and classify them has been pursued in recent years by several groups. Four techniques and approaches will be discussed, and the validity of applying each technique to the current application will be discussed.

Ng et al. [1] describe a technique that uses psychoacoustic measures of sounds to classify snoring sounds in an attempt to classify a subject as an obstructive or benign snorer. Ng et al. use five calculations of individual sounds independently to classify the subject. The five calculations of psychoacoustics were loudness, sharpness, roughness, fluctuation strength, and annoyance. Ng et al. give the equation used to calculate each measure from the segmented audio signal. Twenty-five researchers were asked to classify each snoring sound for each of the five psychoacoustic factors based on the sound alone of each subject. The classification of the subjects and the classification of the calculated value were then compared using a receiver-operating characteristic chart.

Using these, Ng et al. concluded that the highest correlating psychoacoustic measures were loudness, annoyance, and roughness. Although this research classifies snore sounds, no mention is made of normal breath sounds and also how they would be classified. In addition, one of the primary signals used was loudness, which is arbitrary because there is no absolute measure of loudness at the trachea. The classification of a subject being at risk for obstructive snoring is also not applicable in the case of sedation because all subjects are at risk of obstruction given the concentrations of drugs being used. Since this research focused on nonsedated patients, it may not predict how sedated patients sound or the changes that occur during sedation.

Jane et al. [2] used a neural network to classify sounds into two categories of presence or absence of snore. Jane et al. trained the neural network using 625-selected events including snores and other common sounds recorded at the trachea from sleeping patients. The neural network used 22 unspecified inputs from each sound. Jane et al. reported that the neural network classified on average 82% of the sound events correctly. Although this work is relevant, Jane et al. failed to describe the inputs of the network. In addition, simple snoring was not differentiated from obstructive snoring. The ultimate goal of this dissertation is to classify periods of apnea as obstructive or depressive, so an algorithm to translate the classification of individual sounds to the classification of an entire period of apnea needs to be developed.

Hara et al. [3] recognized that the sounds from simple snorers might be easily distinguishable from obstructive snorers. Hara et al. then compiled snore sounds from patients that were simple snorers and patients that were obstructive snorers and observed the frequency spectrum produced by each. The subjects were recorded during sleep

studies. Hara et al. noted that there were significant differences in the spectrum above 800 Hz between simple snorers and obstructive snorers. Hara et al. concluded that using a multidimensional voice program could be a viable solution to differentiating between simple snore sounds and obstructive snore sounds. Although the observation in this work is valuable it did not find a threshold to use to differentiate the two types of sound, but rather used observations and calculations to show that this may be possible.

Nakano et al. [4] took a different approach to classifying breath sounds. Nakano et al. used data from patients undergoing polysomnography (PSG) in conjunction with audio data recorded at the trachea of each patient. PSG is an analysis performed on patients undergoing a sleep study to diagnose sleep apnea. The PSG was performed for each patient and produced the apnea-hypopnea index (AHI). Instead of segmenting and classifying each individual sound, Nakano et al. created a continuous index of what is described as the tracheal sound-respiratory distrubance index (TS-RDI). This was calculated by summing the power spectra from 400 to 600 Hz of the audio data with a rectangular windowing function of 0.2 seconds. This value was placed in a moving average and sharp changes of 12 dB per 18 seconds were defined as tracheal sound dips. The number of tracheal sound dips per hour was calculated and called the TS-RDI. Nakano et al. used the TS-RDI to differentiate between obstructive apnea, hypopnea (defined as shallow breathing), and central apnea (described in this dissertation as respiratory depression). Nakano et al. then compared the AHI to the TS-RDI, finding that it had a correlation coefficient of $r=0.945$. Nakano et al. went on to use receiver-operating characteristic curves to find the best cutoff for the TS-RDI to predict the AHI. The threshold used yielded a sensitivity of 93% and specificity of 67% in diagnosing

sleep apnea. Nakano's et al. approach was distinguishably different because it classified periods of apnea rather than an individual breath. This approach is more useful because it gains a more broad perspective into the patients' breathing state rather than just for a short breath sound. Despite the fact that the primary factor in diagnosing sleep apnea by Nakano et al. was the amplitude of the sound, Nakano et al. did not provide a method for calibration of the sound amplitude. This could lead to unreproducible results if the sound transducer was placed differently on different subjects.

In this research the audio has already been segmented as described in Chapter 2. In contrast to the first three approaches shown, the individual sound classifications will be used to classify an entire period of apnea. In contrast to the approach of Nakano et al., this paper uses the individual breath sounds prior to apnea to classify that period as RD or RO rather than a continuous index of the sound. In addition, the research mentioned has been done on nonsedated patients. Since the patients in the data set for this dissertation were sedated, the characteristics of the snoring sounds may be different.

3.1.4 Basis for sound classification

Personal observation during data collection has shown that the human ear can decipher the difference between normal breath sounds and partially obstructed breath sounds with little training. If an algorithm could be developed to decipher the difference between these sounds, the automation of apnea classification could help in not only identifying that a patient is not breathing but also in identifying the source of the cessation of breathing.

An assumption of the breath detection algorithm described in Chapter 2 is that every sound detected above the noise threshold is a valid breath. The data show that not

all periods of apnea detected by the respiratory flow meter were detected by the acoustic detection method. This is a result of sounds not related to breathing creating a sound exceeding the noise threshold. Examples of sounds not generated by breathing are background sounds such as ambient talking or machine sounds. Additionally, some sounds which are related to breathing, such as partial obstruction, indicate that sound is detected but in fact the flow may be inadequate to clear the respiratory deadspace. The respiratory deadspace is the volume of air between the avioli and the mouth. Ventilation is not achieved if the respiratory deadspace is not cleared in a breath.

During partial obstruction, some air passes through the respiratory tract causing sounds to be detected at the trachea. These sounds are louder than normal breathing sounds with similar respiratory flow rates. This is due to a smaller airway, which generates a higher amplitude of sound. In addition, adipose tissue slaps against other tissues to produce more sound. These sounds are produced by the same mechanisms that produce snore sounds. Because of the relationship of snore sounds to obstruction, snoring is considered a precursor to obstruction [3,1,2].

For this research snore sounds have been separated into two cases to distinguish between snores that do not lead to obstruction and snores that do. The first case is the productive snore. This happens as the upper pharynx closes partially so that adipose tissue begins to slap against itself to produce the snore sound, but flow continues and has a high enough volume to clear the respiratory deadspace during the breath. The second case is the obstructive snore. This occurs as the pharynx becomes more closed and does not allow sufficient ventilation to clear the respiratory deadspace.

Snoring is more commonly observed during the inspiratory phase of breathing.

This is because the negative airway pressure produced by the contraction in lung volume causes loose tissue in partially occluded airways to become more constricted. This constriction is either alleviated as the airway opens because of the force of the pressure, or the constriction is not alleviated and no gas exchange takes place [5]. During expiration the pressure within the lungs increases. If the airway is occluded during this time the positive pressure opens the airway allowing for gas exchange. This physiologic explanation has led to the use of continuous positive airway pressure (CPAP) devices, the most common solution to obstructive sleep apnea syndrome (OSAS) [6]. CPAP devices supply a continuous positive airway pressure by forcing air into the nose and/or mouth as the patient sleeps. The patient's airway then stays open even during inspiration because the air pressure within his/her airway is greater than the ambient air pressure. This is an important factor to the sounds generated in a subject not using a CPAP device. During a productive snore breath the inspiratory breath sound has characteristics of a snore, whereas the expiratory sound is a normal breath sound because it exerts a positive pressure on the pharynx closure. During a partially obstructed breath the inspiratory breath sound still has characteristics of a snore but the expiratory sound is too quiet to be detected. During periods of complete obstruction neither the snore sound nor the breath sound can be detected on either phase of breathing because there is no flow through the trachea.

3.1.5 Data set

The data set used for the sound classification portion of this project was the same data set described in Chapter 2. In addition, this portion of the project uses the audio segmentation markers generated by the algorithms described in Chapter 2. It is also important to note that during the recording of the data set the author listened to the audio

data and observed the other data recorded for the majority of the subjects such as measured respiratory flow rate, blood-oxygen saturation, and RIP values. These observations gave the author insight into the sounds that occur during RO and RD before the data were processed and algorithms developed.

3.2 Methods

3.2.1 RD and RO classification standard

3.2.1.1 Definition of respiratory obstruction

The current standard for identifying a period of RO is a physician's observation of the patient and monitors at the time of the obstructive period or during polysomnographic (PSG) analysis [7]. Due to the cost of having a physician perform PSG on all 20 full data sets, an alternative method of differentiating RD and RO was performed.

Identification of periods of obstruction was based on a physiological definition of RO using respiratory inductance plethysmography (RIP) and flow rate monitors. The RIP device measures the change in chest and abdomen circumference by measuring the inductance of wires sewn into elastic bands placed on the chest and abdomen. Physiologically, RO occurs when the airway becomes constricted but the drive to breathe continues. RO is identified when the flow rate shows no valid breaths but the RIP monitor shows volume changes up to approximately 180 mL. The use of 180 mL will be explained in Section 3.2.1.3

3.2.1.2 Definition of respiratory depression

Respiratory depression occurs as the patient loses the drive to breathe. In the case of this study RD occurred as a result of the combination of drugs used to sedate the

subject. Physiologically, RD becomes evident as breath volume of the subject decreases. The breath rate may also decrease, but the airway remains open. This continues until the subject does not take a breath for the designated period of time of fifteen seconds.

Using the flow rate measurement and chest and abdomen measurements, RD is identified when both signals show no valid breaths for the designated period of fifteen seconds or longer. Before the period of apnea, the flow volume and RIP signals show a correlated signal. In addition, the calibrated volume measured by the RIP remains correlated with the calculated volume from the flow rate.

3.2.1.3 Lung volume change during obstruction

During a period of RO, the RIP can measure changes in the volume of the chest and abdomen that are not detected by the flow meter. This is believed to be caused by the change in volume due to the change in pressure exerted by the subject. The maximum amount of lung volume change by suction pressure of gas during inspiratory obstruction has been estimated to be 180 mL, corresponding to the maximum suction pressure of 60 cmH_2O that the lungs will exert in normal circumstances [8]. Using this value converted to atmospheres results in the value of 0.058 atmosphere change of pressure within the lungs. The data were recorded in Salt Lake City, UT at an approximate altitude of 1288 m. The atmospheric pressure at this altitude is approximately 0.857 atmospheres. Assuming that the air breathed can be modeled as an ideal gas and that the temperature of the gas was constant, the ideal gas law was used to calculate the change in lung volume due to suction pressure. The ideal gas law follows the equation $PV=nRT$, where P is the pressure within the lungs in atmospheres, V is the volume within the lungs in liters, n is the number of moles in the lung volume, R is the gas constant with units of Joules per

Kelvin mole, and T is the absolute temperature in Kelvin. The functional residual capacity (FRC) or the lung volume after expiration in an adult can range from 2 to 4 liters [9]. For this calculation an FRC value of 2.5 liters was used. Since the initial values of V_1 and P_1 were known and n, R, and T did not change, the values of n, R, and T were considered one constant with the value $nRT=2.5\times0.857=2.1425$ atmosphere liters. The maximum change in volume was calculated using this equation:

$$V_1-V_2=nRT/P_1-nRT/P_2 \quad (3.1)$$

Substituting the nRT value in equation 3.1 results in a change of 181 mL as the maximum volume change due to suction pressure within the lungs.

3.2.1.4 Manual classification of apnea

It was hypothesized that, using data from the flow meter and the RIP, a manual classification of the periods of apnea would be feasible. The classification was blinded to the author for the testing set. The physiology of RO and RD as explained in Section 3.1.4 was used as the basis of this classification.

Using the breath detection markers described in Chapter 2 all periods of apnea detected that were longer than 15 seconds in length were classified as RO, RD, or period of "mask off" (see below) in the following manner: For each period of apnea the flow rate, breath volume, and calibrated RIP volume were viewed, including 5 seconds prior to the apnea and 5 seconds after the apnea. Three researchers who were not involved in developing any of the algorithms in this research, independently classified each period of apnea as described next.

RO was the classification if, during the period of apnea, the investigator determined that the flow rate signal showed no breaths but RIP volume showed breath volume

change up to 180 mL. RD was classified if, during the period of apnea, the investigator determined the flow rate signal showed no breaths and the RIP volume also showed no significant volume change that would be consistent with a breath. A period of "mask off" was classified if the flow rate showed no breaths but the RIP volume showed volume changes greater than 180 mL in volume, since a RIP volume change greater than 180 mL during a period deemed apnea by the flow meter indicates that the mask that measures respiratory flow was removed. These periods were removed from the data set.

The resulting classifications of the apnea from the three researchers were then analyzed in the following way: If a classified period was classified the same way by two or more of the researchers, the classification given by those researchers was chosen for that period of apnea. Periods that were classified differently by all three researchers were displayed before all three researchers at the same time. The three researchers then decided as a group which of the three classifications was the best fit. In the event that the researchers did not agree, the period in question would be removed from the data set. Of the 645 periods to be classified the three researchers agreed 395 times and at least two agreed a total of 622 times. Of the 23 periods that at least two researchers did not agree, all were able to be agreed upon in a collaborative setting.

The results of the classification performed by the researchers was compared to the statistical probabilities of three randomly assigned classifications. Given the three possible options, each with three independent trials and all equally likely, there are 3^3 classification possibilities. First the probability of three independent classifiers choosing the same classification will be found. Three of the 27 possibilities result in a unanimous decision. A reasonable assumption was that those classifying the sounds did not have a

prior expectation of each of the three possibilities because those classifying these events were not told the likelihood of each of the three possibilities. Using this assumption each decision can be treated as a binomial random process with two possibilities: unanimous decision or non-unanimous decision. The probability mass function (PMF) is:

$$P_K(k)=\begin{cases}1/9, k=3 & Unanimous \\ 8/9, k<3 & Non-unanimous\end{cases} \tag{3.2}$$

where $P_K(k)$ is the PMF of the binomial distribution, and k is the number of agreements from the three classifiers, and K is the binomial random variable denoting unanimous decision. The mean value for k=*3* is given by the mean of the binomial distribution $E[X]=np$, where n is the number of independent trials p is the probability of the unanimous decision, which is 1/9 in this case, and X is the random variable [10]. Given this, the mean value is 71.66 out of the 645 trials. The variance of a binomial distribution is $VAR(X)=np(1-p)$ where n is the number of independent tirals, p is the probability of unanimous decision and X is the random variable [10]. For this case the standard deviation is 7.98. The number of periods of apnea classified unanimously by the researchers was 395 times out of the 645. This value is over 40 standard deviations better than for independent uniform random trials.

Since the classification was based on two or more researchers agreeing, the statistical probability of the having at least two independent sources agree was also calculated and compared with the results from the researchers. The probability of two or more of three independent trials agreeing given three equally likely possibilities is seven out of nine. Treating this as a binomial random process as in the previous case the PMF of this is:

$$P_K(k)=\begin{cases}7/9, k\geq 2\\ 2/9, k<2\end{cases} \tag{3.3}$$

where $P_K(k)$ is the PMF of the binomial distribution and k is the number of classifiers in agreement. The mean number of at least two classifiers in agreement given 645 independent trials is 501.66, and the standard deviation is 10.56. At least two of the researchers classified 622 of the 645 periods in agreement, which is over 11 standard deviations better than for independent uniform random trials.

The assumption that each classification is equally likely was made because the actual probabilities of how often each classification occurs is not known. Since these probabilities are not known, the number of classifications found in the training set will be used as a secondary statistical comparison. From the training set, the total number of apnea periods classified as RD were 564 out of a total of 821. The total number of apnea periods classified as RO were 28 out of a total of 821. The total number of apnea periods determined to be a period of mask off were 195 out of a total of 821. Comparing the number of unanimous agreements made by the researches to three independent trials of the probability of randomly choosing unanimous agreement results in the probability of

$$P_K(k)=\begin{cases}p_1^3+p_2^3+p_3^3 & k=3 \quad Unanimous\\ 1-(p_1^3+p_2^3+p_3^3) & k<3 \quad Non-unanimous\end{cases} \tag{3.4}$$

where $P_K(k)$ is the PMF of the binomial distribution, k is the number of classifiers in agreement, p_1 is the probability of choosing RD as the classification, p_2 is the probability of choosing RO as the classification, and p_3 is the probability of choosing mask off as the classification. Given the statistics of the training set classification, the probability of unanimous decision is 40%. In order to reach the number of unanimous decisions made by the researchers the value of p_1 would have to be increased to 89.44 while setting the

values of $p_2=0$ and $p_3=1-p_1$. The values of p_2 and p_3 are adjusted this way to maximize the probability of unanimous decision given the random variables.

For the case of having two out of three classifiers correct, and using the same statistics as for the unanimous decision, the probability of choosing two out of three in agreement is 96.4%. This is compared to the percentage of agreements made by the researchers two out of three times of 96.4%. Because the probability of randomly choosing two out of three to agree is the same percentage found by the researchers, this shows a statistical weakness of this method of classification.

The comparison of the statistics of actual classification to the probabilities of random classification shows that the human classifier performed significantly better than if periods were randomly classified for unanimous decision but the same for a random classifier for getting two out of three to agree. Because the unanimous classification by the researchers performed significantly better than the random classification, it is believed by the author that this shows an overall reliability of this classification method.

3.2.2 Proposed sound source models

The three most prominent sounds recorded during the data collection procedure were normal breathing, snoring, and vocalization. In order to differentiate between these three sources, the cause of the sound being generated must be understood. To better understand these sounds, each was modeled based on observations of sound features, and simulation models were created based on these observations of the features and an understanding of the physics of what created it. One feature that was hypothesized to differentiate sounds based on the physics of how it was created was the histogram of that sound. It was hypothesized that a normal breath sound normalized by its standard deviation

would have a Gaussian distribution because of the source of the sound, as will be described in Section 3.2.2.1, whereas a snore sound would have a distinguishably different distribution. Vocalization would also have a distinct distribution. Thus, each sound was modeled using a distribution model of the dominant sound that was observed. Because the histogram of a sound removes all time dependence of the input signal, an additional method to find the periodicity of the incoming signal was developed. Periodicity is believed by the author to be a feature that can differentiate between productive snore sounds and obstructive snore sounds. In Section 3.2.3.5 the combination of these two feature classifications will be discussed.

Sounds recorded at the trachea are different from sounds used in speech recognition because when the sounds pass through the tracheal wall they are filtered. Like most physical barriers, the tracheal wall acts as a low-pass filter and, thus, higher frequencies used in speech recognition are not able to be detected. In addition, the articulation that occurs above the pharynx during vocalization cannot be heard at the trachea because it is filtered by the stethoscope cup and tracheal wall. For this reason, methods of classifying the breath sound using nontraditional features were developed and used. The choice of features was based on the differences of the sounds observed during the data collection process that will be described in the next section.

Each of the three sound categories described will now be discussed. In addition, a model for each sound was created and will also be discussed.

3.2.2.1 Normal breath sound model

The most common sound recorded in the data set was the normal breath sound. Both inspiratory and expiratory sounds in a normal breath are very similar, and are

caused by the same mechanism. These sounds are created by the turbulent air flow through the trachea when the flow rate exceeds the threshold associated with the Reynold's number for the tracheal geometry. The flow rate corresponding to this in adults is estimated at 7.2 l/min [11]. Although the amplitude of the audio envelope of normal breathing sounds is highly correlated with the flow velocity [12], sounds such as vocalization and snoring do not have this same correlation. This is because the amplitude of the sounds generated by vocalization and snoring are much louder than normal breaths for the same flow rate. This means that sounds with similar amplitudes can have a wide range of breath volumes.

Kompis et al. [13] use a model that simulates breath sounds as band-limited white noise modulated by the amplitude of a breath phase. Kompis et al. did not explicitly define the waveform for the amplitude of the breath phase. Kompis et al. also did not give an equation for this model, rather a description and block diagram. In this project, normal breath sounds were modeled as Gaussian-distributed noise multiplied by a half-cycle sinusoid, as shown in Figure 3.1. A half-cycle sinusoid was chosen for simplicity in modeling one respiratory phase of breathing. Both inspiration and expiration were modeled using the same model. The simulation of the sound for a normal breath y_{nb} had the following equation:

$$y_{nb}(t)=a_b \sin(2\pi f_b t)\cdot N_{t,1}(0,1) \tag{3.5}$$

where t is time in seconds, a_b is the peak amplitude of the sinusoid in volts, f_b is the respiratory phase rate in Hz, and $N_{t,1}(0,1)$ is a vector of normally distributed random samples with a mean of zero and standard deviation of one.

In the case of normal breathing, the only other sound that would be additive

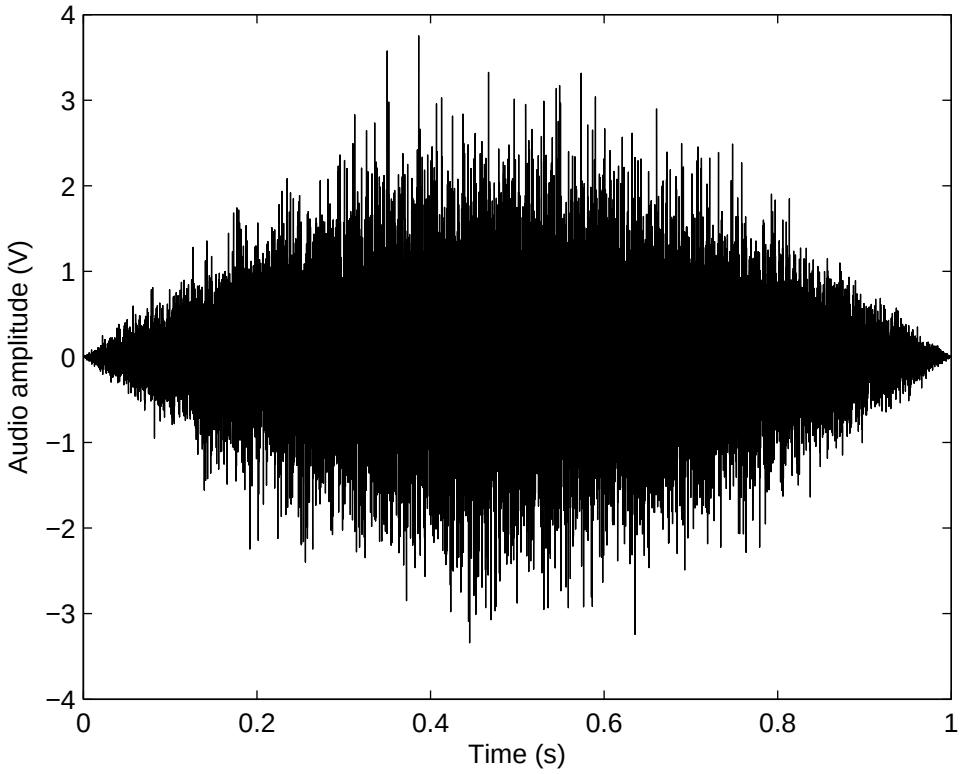


Figure 3.1 Simulated normal breath sound.

would be background noise due to a combination of electronic noise and ambient sound filtered by the stethoscope cup. The background noise was modeled in Chapter 2 as a wide-sense stationary Gaussian-distributed model. A more complete model of normal breathing including the background noise would be:

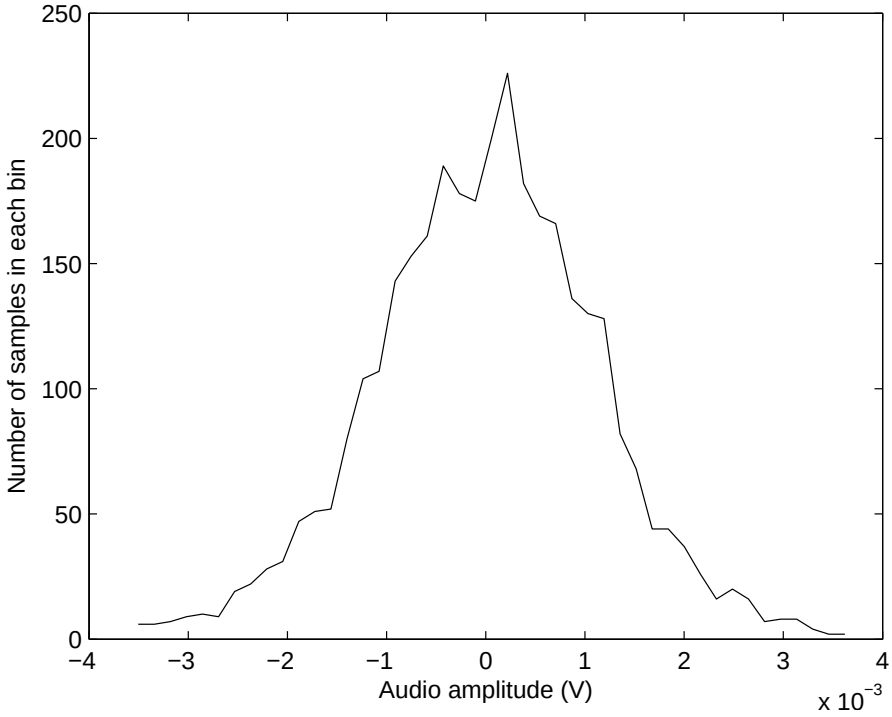
$$y_{nb}(t)=a_b sin(2\pi f_b t)\cdot N_{t,1}(0,1)+N_{t,2}(0,a_n^2) \quad (3.6)$$

where a_n is the standard deviation of the background noise and $N_{t,2}(0,a_n^2)$ is a vector of normally distributed samples with zero mean.

This model was created to simulate the band limited white noise heard when listening to normal breath sounds as they are modulated by the respiratory cycle. This is very similar to the Kompis et al. model with the exception that the modulating signal is specified in this case as a half-cycle sinusoid and the additive white noise is added.

Normal breathing heard at the trachea appears to be “white noise” to the casual observer, similar to sounds generated by rushing wind or water. The observation of the breath sound appearing “white” led to finding the actual distribution and band energy of a normal breath. Data were used from the data set described in Chapter 2. Some typical breaths were analyzed to find what type of distribution and band energy they would display. In most instances the histogram of normal breath sounds from raw signals formed a Gaussian distribution, as seen in Figure 3.2a. In other instances, however, they were heavier tailed, as shown in Figure 3.2b. The change in the distribution seemed to change from being Gaussian to more heavy-tailed than Gaussian from breath to breath. During observation of the envelope of the signal and the distribution of the signal, a pattern emerged. It was observed that sounds that had a higher variance of the audio envelope created a heavier tailed distribution, and breaths which had a relatively uniform

a



b

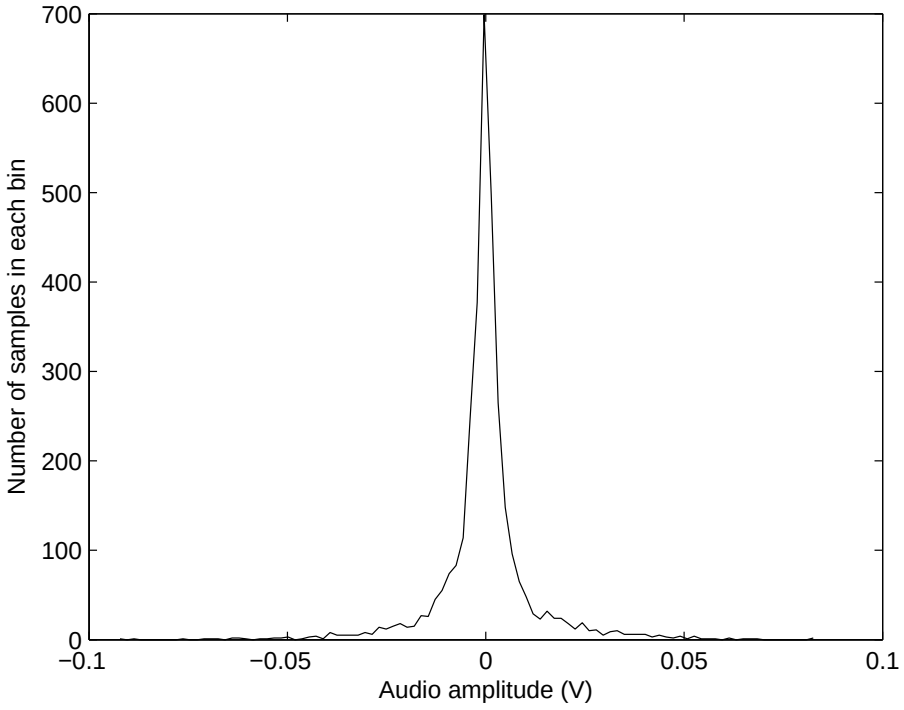


Figure 3.2 Distributions of normal sounding breaths displaying a distribution similar to (a) a Gaussian distribution and (b) a Laplace distribution.

audio envelope amplitude demonstrated a more Gaussian distribution. The heavier tailed distributions were observed to be similar to a Laplace distribution. Flow rates during a single breath vary, and since during normal breathing the flow rate is related to the sound amplitude, the amplitude of the breath sounds varies over the course of a breath sound. It was hypothesized that since the amplitude of the flow rate signal was varying, the distribution of the audio signal would not be Gaussian.

These observations led to an investigation of normalizing the amplitude of a breath sound to improve classification by means of histogram matching as will be described later in Section 3.2.3.1.

Because the band energy and shape is a common feature to differentiate sounds, the band energy of a normal breath will be discussed, even though it was not used as a method of classification. The band energy observed in a normal breath is not exactly white because it is not a uniform amplitude across all frequencies. An example of the band energy of a normal breath is shown in Figure 3.3. In most cases the band energy had a similar spectrum to the one shown. There were rarely any significant spikes in the band energy, and the signals' maximum frequencies were around 800 Hz. This was initially explored as a classification feature but was often indistinguishable from the spectrum created by a snore and was abandoned.

In Chapter 2 it was hypothesized that the breathing signal formed a Laplace distribution. This hypothesis seems to be contradictory to the hypothesis that normal breath sounds have a Gaussian distribution. The variation in the amplitude of the normal breath sound causes the distribution of the breath sound to become more heavy-tailed. Appendix A gives the calculation of distribution of the modulation of a Gaussian random

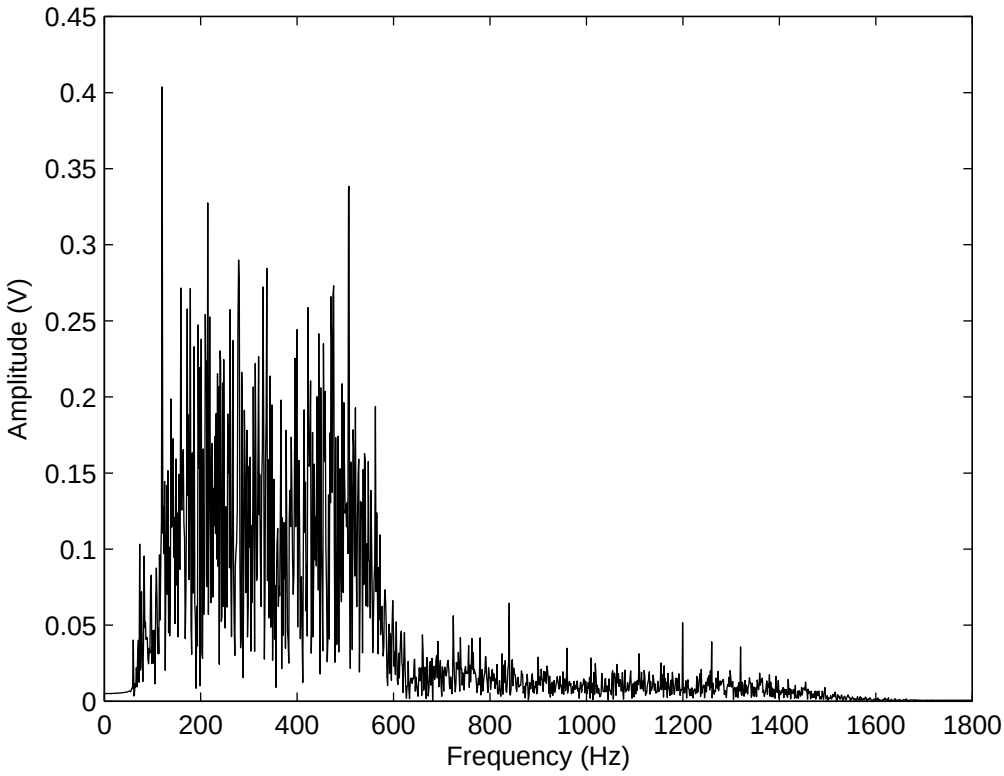


Figure 3.3 Spectrum of a typical normal breath sound.

signal by a sinusoidal signal. As it will be discussed later in this section, the histograms of the breathing sounds do appear to have a Gaussian distribution after the envelope of the sound has been used to normalize the sound or by using a smaller section of the breath sound where there is less variation of the amplitude of the sound. The calculations shown in Appendix A and Section 3.2.2.2 support the claims made in Chapter 2 that the continuous breathing sound can be modeled by a distribution with a central peak and heavier tails than a Gaussian distribution. These distribution models will also be defined and used in the automated classifier in Section 3.2.3.3.

3.2.2.2 Breath sound histogram simulation

To test the effects of modulation on a Gaussian signal distribution, a simulated breath signal was generated using the model of equation 3.5. A pseudo-random number generator (MATLAB) was used to create a vector $N_{sim}(n)$ with a length of $M=50,000$ samples and a Gaussian distribution. This length was chosen because of the length of a normal breath (approximately 2.2 seconds) at a sampling rate of 22 kHz. A second vector $S_{sim}(n)$ that was a single period of a sine wave was created with the same number of samples as $N_{sim}(n)$. The dot-product of the two vectors was calculated as $B_{sim}(n)=S_{sim}(n)\cdot N_{sim}(n)$. The envelope $E_{sim}(n)$ of $B_{sim}(n)$ was then calculated as described in Chapter 2. The simulated breath sound $B_{sim}(n)$ was then divided by the envelope to create an amplitude-normalized signal, $L_{an}(n)=B_{sim}(n)/E_{sim}(n)$. This method of dividing the original signal by its envelope will be referred to as the amplitude normalization method. The histograms of the signals $N_{sim}(n)$, $B_{sim}(n)$, and $L_{an}(n)$ were calculated using 1000 bins. This number of bins allowed for an average of 50 samples per bin. The histograms of the three signals can be seen in Figure 3.4. This shows that a Gaussian distributed signal

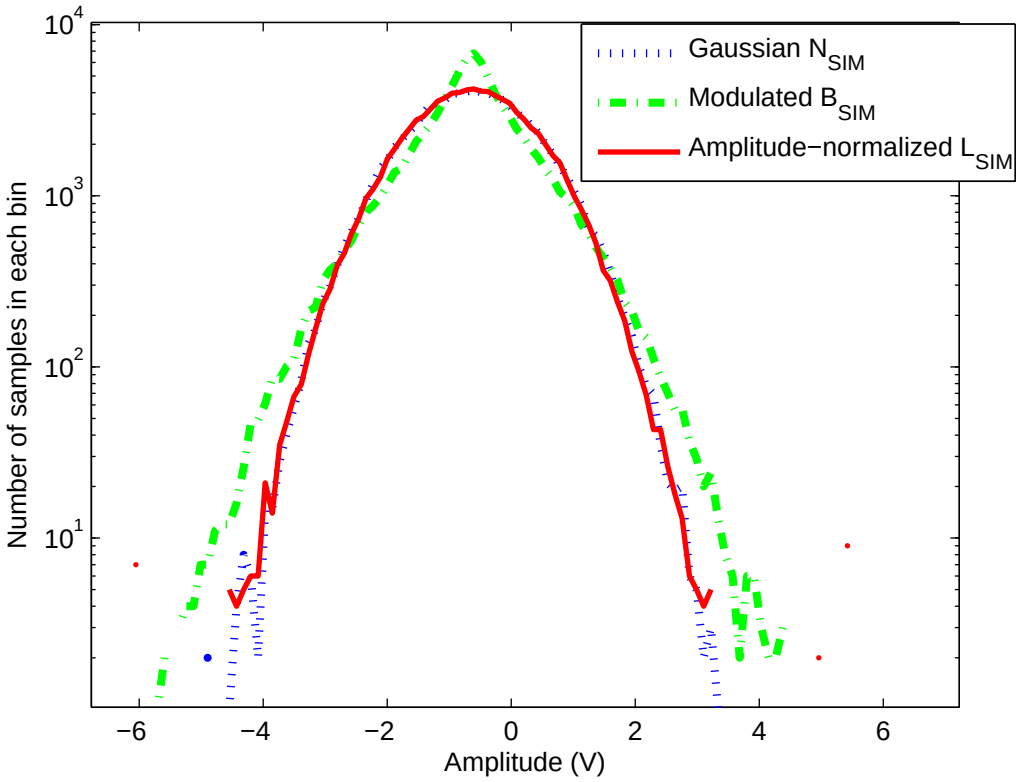


Figure 3.4 Histograms of the simulation of a Gaussian signal, sinusoidally modulated Gaussian signal and amplitude-normalized Gaussian noise.

modulated by a sinusoid can return to a Gaussian distribution through amplitude normalization. This is important because, if the histogram of a signal is to be classified, the signal should be either divided by its envelope to show the amplitude normalized histogram of the signal, or be divided into short sections to be classified individually to avoid large changes of amplitude over a breath sound.

Another option to amplitude normalization is the use of short segments of a breathing sound in which the variability of the envelope of a breath sound would be small. This method of using short segments from a breath sound was used for the classifier described in Section 3.2.3.3.

To test the segmentation method, the vector $B_{sim}(n)$ was divided into 22 sections of 2205 samples each. The histograms of each of the segments appeared to have a Gaussian distribution except for those where the signal $S_{sim}(n)$ included a segment passing through zero. The absolute slope of a sinusoid is greatest at the zero crossings, thus causing the variation in the amplitude to be greatest during the segments that contain zero crossings of the modulating signal. Because of the background noise and the breath detection method described, the nonsimulated breath sounds will not have this problem. The histograms of the segments from the simulation are displayed in comparison to a Gaussian distribution in Figure 3.5.

The distribution of a Gaussian signal modulated by a sinusoid can be calculated analytically. The distribution of this signal is [14]:

$$f_Z(z)=\frac{1}{\pi\sqrt{(2\pi)}}[e^{\frac{-z^2}{4}}K_0(\frac{z^2}{4})], \tag{3.7}$$

where z is the random variable, and K_0 is the order zero modified Bessel function of the second kind. The method for deriving this distribution is given in Appendix A. An

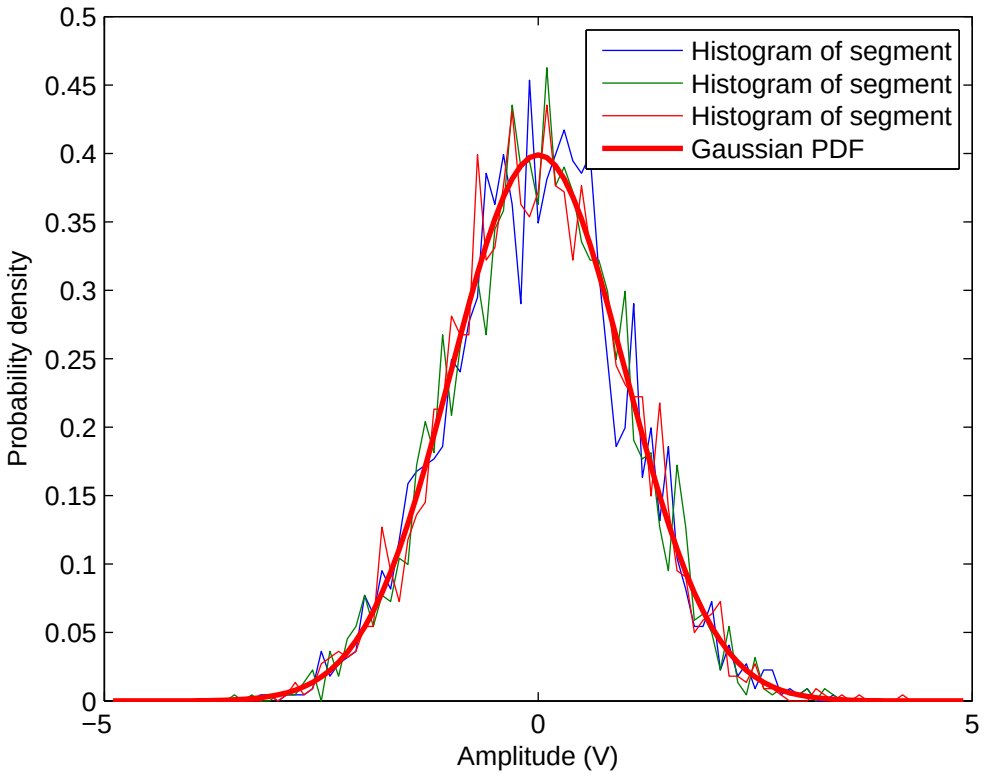


Figure 3.5 Comparison of these segments of modulated Gaussian signals to a Gaussian probability density function.

example of the distribution with a simulation is shown in Figure 3.6. This distribution is heavier-tailed than the distribution of the original signal $N_{sim}(n)$. In addition, this distribution has a pointed central peak when compared to a Gaussian distribution. These two differences are important because it will be shown in Section 3.2.2.3 that the distribution of a snore sound is heavy-tailed when compared to a Gaussian distribution and has a central peak, even when amplitude normalization is performed. These are important differentiators between normal breath sounds and snore sounds.

In contrast to the simulation of a normal breath sound, actual data was used to test the amplitude-normalization method. An example of a heavy-tailed normal breath sound was found, as shown in Figure 3.7a. The envelope of the sound and the original signal was divided by the envelope of the sound. The histogram of the amplitude normalized breath sound is shown in Figure 3.7b compared to a Gaussian distribution. The amplitude normalization technique does help the normal breath sound appear to be Gaussian.

A second important feature used to classify sounds as normal breath sounds was the periodicity of the sound. Normal breath sounds are not periodic because there is nothing resonating or vibrating to create a repeated sound. A method to determine the periodicity of the sounds will be described in Section 3.2.3.4.

3.2.2.3 Snore sound model

As noted in the introduction there are two different kinds of snore sounds that are important to classification. The first is a productive snore that allows adequate gas exchange to the lungs. The second is an obstructive snore sound that does not allow for adequate gas exchange to the lungs. Both are caused by the same mechanism and have

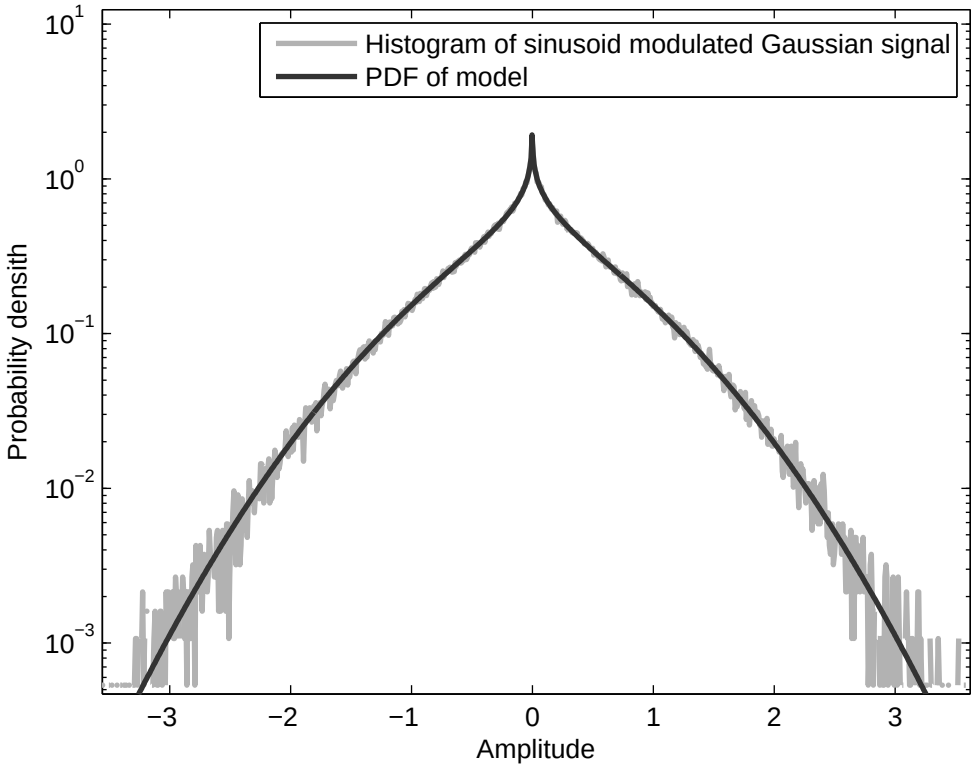
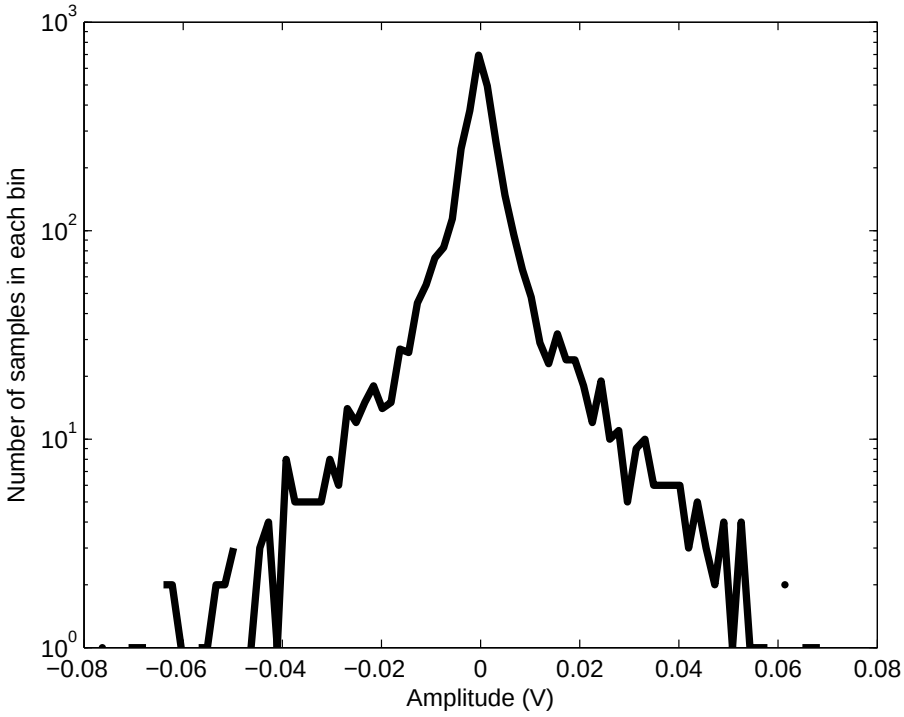


Figure 3.6 Normalized histogram of a sinusoid modulated Gaussian random signal overlaid with the PDF for the signal.

a



b

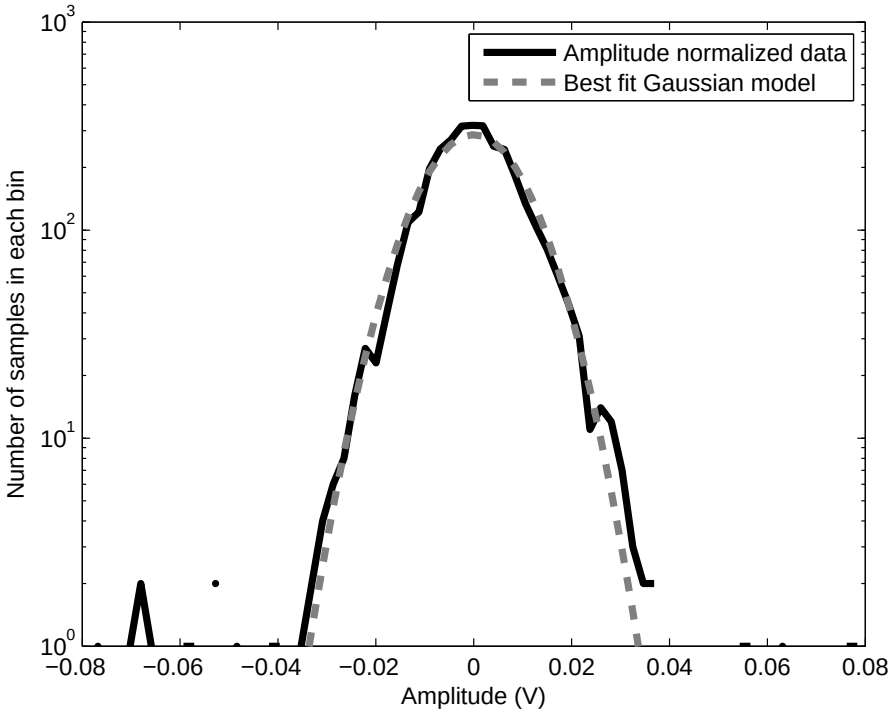


Figure 3.7 Distribution of a normal breath sound (a) before envelope normalization and (b) after envelope normalization with Gaussian model.

been modeled similarly, as will be described.

The loudest sound produced during a snore sound is the slap sound of the adipose tissue. This sound was modeled in this dissertation as a repeated drumbeat sound with variable repetition rate and amplitude. Each drumbeat was modeled as an exponential decay modulated by a sinusoid. This model is based on the sound produced by the oropharynx as the tissue slaps against itself. This tissue has a harmonic component after the initial slap similar to that of a drum. The simulation of the sound for a snore slap y_{db} has the following equation:

$$y_{db}(t) = c\,e^{\frac{-t}{d}} \cdot a \sin(2\pi\, f_n t) \tag{3.8}$$

where t is time in seconds, c is the amplitude of the individual slap in volts, d is the decay time constant of the slap, a is the amplitude of the sinusoid, and f_n is the frequency of the slap in Hz. One possible improvement of this model is if f_n is not constant. A varying value of f_n would mean that the frequency would change over time, thus producing a spectrum that is not a single peak. An example of a snore signal from the recorded data set displaying this property is shown in Figure 3.8. This models a single snore slap; multiple snore slaps can be simulated by concatenating several slaps together with differing values of c assuming that the other parameters stay constant through one breath.

Normal breath sounds are present in all sounds that have a breath flow rate that becomes turbulent. This is because sounds that have a high enough flow velocity to create turbulence will generate the sounds that create the normal breath sound in addition to the primary source of the sound that is modeled.

If there is a high enough flow velocity to create turbulence, the normal breath sounds as described above also contribute to the model. In addition the background noise

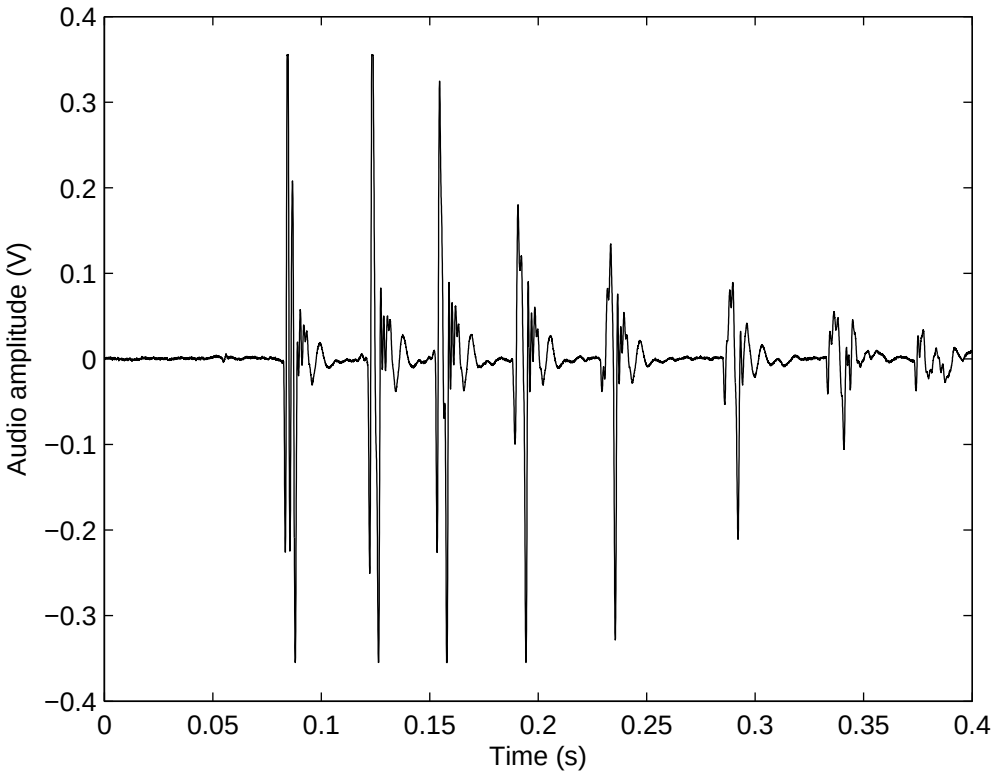


Figure 3.8 Repeated snore slap waveform from a snoring signal.

is present in every sound. A more complete model including the turbulent noise and the background noise with the snore slap is

$$y_{db}(t) = c e^{\frac{-t}{d}} \cdot a \sin(2\pi f_n t) + N_{t,1}(0, (a_n^2 + a_{nb}^2)) \tag{3.9}$$

where a_{nb} is the standard deviation of the noise generated by the normal breath and the other parameters are the same as previously defined. It is important to note that the modulation of the normal breath by a sinusoid is not present. This is because it is assumed that the length of a snore slap is short enough that the amplitude of this sound for the period of a snore slap can be considered constant.

From observations early in the research process, the distribution of a snore sound commonly had a heavier-tailed distribution when compared to the Gaussian distribution. Two features that differentiated the distribution of a snore sound from the normal breath sound distribution after amplitude normalization were a peak at the mean rather than a rounded point, and tails that were more pronounced than a Gaussian distribution. These differences were more easily identified when viewing a semilog plot of the distributions as shown previously in Figure 3.4. The model chosen for this distribution was a Laplace distribution. This was chosen for several reasons. It displays many of the same characteristics observed in the snore sound distributions and it is based in the exponential family of distributions like the Gaussian distribution. Because the Gaussian and Laplace distributions are in the exponential family, mixtures and other derivations are more easily calculated.

To explain the distribution associated with snoring, the distribution of an individual snore slap of equation 3.8 was calculated as

$$h(v)=\begin{cases}\dfrac{d}{\pi lv}\left(real\left(\sin^{-1}\left(\dfrac{v}{ace^{-l/d}}\right)\right)-\sin^{-1}\left(\dfrac{v}{ac}\right)\right), & for\ \ -ac<v<ac \\ 0, & otherwise\end{cases} \tag{3.10}$$

where c is the initial amplitude of the snore slap, d is the decay time constant of the exponential decay, v is the random variable, a is the amplitude of the sinusoid portion of the signal, and l is the length of the snore slap. The derivations of this distribution is given in Appendix B. This distribution is a combination of a distribution in the Pareto distribution family and an inverse sine distribution. In cases where l is much greater than d, it displays a distribution heavier tailed than a Gaussian distribution with a peak in the center of the distribution. In cases where l is much smaller than d it displays a distribution similar to a sinusoidal distribution. Examples of these distributions are shown in Appendix B. It is assumed from these distributions of the snores and from Figure 3.8 that, under normal circumstances, l is greater than d. This distribution has a more similar behavior to the Laplace distribution chosen for the histogram classification algorithm than to the Gaussian distribution chosen for the normal breath sound model.

As noted, both a productive and obstructive snore slap sound have been modeled by equation 3.9. The difference between the two has two aspects. The first difference is the length of time of each slap sound. In a productive snore the length of time of each slap is nearly uniform because of continuous flow being interrupted by the slaps. In an obstructive snore the slaps are not uniform because of the lack of continuous flow. In addition the time between the slaps is longer and can be more variable. The second difference is the normal breath sound portion of the model in equation 3.9. During a productive snore sound, the amplitude of the normal breath portion of the model is audible just before and after the snore because of the relationship of flow to audio

amplitude. During an obstructive sound the normal breath portion of the model is inaudible because there is not adequate flow. In the case of the productive snore, the distribution would be formed by the summation of the signals that create both the distribution of equation 3.10 and the Gaussian distribution signal before and after the snore sound.

When either the amplitude normalization method or segmentation method was used on a distribution from a snore sound, the distribution appeared much the same as before the transformation. This was both observed and tested in Section 3.2.3.3. The distribution not changing through segmentation or amplitude normalization may be due to the fact that the amplitude of the envelope of a snore slap for an individual slap remained unchanged because of the short length of the slap.

The second feature used to classify snore sounds was the periodicity of the snore sound. As described above, a productive snore sound has a regular slap period and would be periodic. An obstructive sound would not be periodic because of the lack of a continuous flow that would regulate the repetition of the snore slap. This was an observation made during data recording by the author. In addition this assumption will be tested by the comparison of the classification standard described to that of the two-dimensional classifier that will be discussed further in Section 3.2.3.5.

3.2.2.4 Vocalization model

Vocalization is the loudest sound recorded at the trachea. Vocalization was so loud that it often saturated the microphone in the stethoscope cup. Vocalization is caused as the vocal chords are stretched across the trachea. Vocalization from an individual other than the subject being recorded can also be detected at the microphone, but this sound is not as loud as the vocalization from the subject and is mixed with other sounds

such as noise. The vocal chords vibrate in a harmonic pattern, depending on the tightness, and produce harmonic sounds. There are two distribution models used in this dissertation for vocalization, one for when the microphone is not saturated and another when the microphone becomes saturated.

The first model is a modification of the model proposed by Gazor et al. [15] as a Laplace distribution for speech. However, the histogram observed from vocalization had a distribution that appeared to be Gaussian, and the distribution model used for vocalization was the Gaussian distribution. This apparent inconsistency is explained by the signal recorded at the trachea being filtered physically by the tracheal wall and skin, and digitally by the filter described in Chapter 2. Both of these filters are considered low-pass filters. To demonstrate how this would effect the distribution of a vocalization, the filter described in Chapter 2 was applied to a simulated Laplace signal. Figure 3.9 shows the normalized histogram of the simulated Laplace random signal, the normalized histogram of the signal after it has been filtered and downsampled, and a Gaussian distribution with a mean of zero and standard deviation of the filtered signal. From this figure it is evident that the Gaussian distribution is a good fit for the filtered signal. This also supports early observations of nonsaturating vocalizations that have a histogram that appeared to be Gaussian distributed.

As noted, the vocalization is not the only sound present, but because the other two additive signals from the background noise and normal breath sound also have Gaussian distributions, a combination of these sounds with the Gaussian distributed vocalization signal is simply a Gaussian distribution.

The second distribution model occurs when the microphone is saturated by the

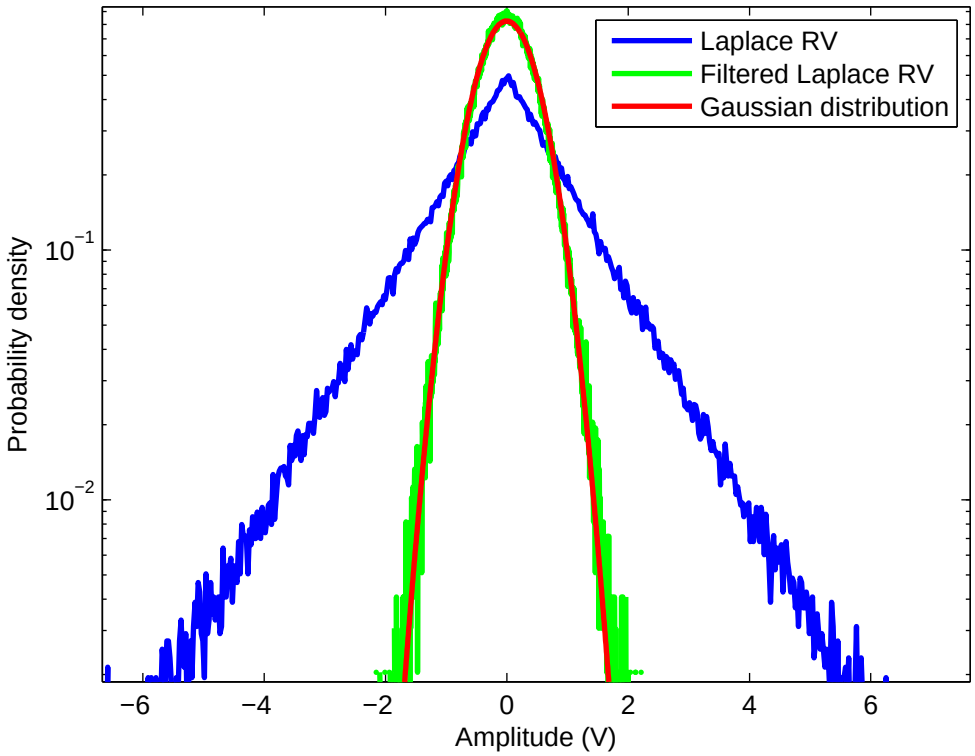


Figure 3.9 Comparison of the distributions of a Laplace random signal, a filtered Laplace random signal, and a Gaussian distribution.

amplitude of the sounds recorded at the trachea. A strong saturation creates a signal similar to a square wave with 50% duty cycle and positive value of *m* and negative value of -*m* where *m* is the absolute saturated value. A perfect square wave has a distribution that is an impulse at *m* and at -*m* as described by

$$s_X(x)=\delta(x-m)+\delta(x+m) \tag{3.11}$$

where *m* is the saturation level of the signal and *x* is the random variable. Microphone saturation does not create a perfect square wave, and in addition, the signal used was filtered by a low-pass filter to be downsampled to 4410 Hz. The square wave has a Fourier series:

$$f(x)=\frac{4}{\pi}\sum_{n=1,3,5...}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi x}{L}\right) \tag{3.12}$$

where *x* is the independent variable, *n* is the number of the harmonic, and *L* is the length of a half period of the square wave. The Fourier series indicates that the spectrum of a square wave is made up of the odd harmonics and has a decreasing amplitude as the frequency increases. The low-pass filter implemented by the tracheal wall and digital filter removes the higher frequency formants. If the low-pass filter removes all but the primary harmonic of the square wave it becomes a sinusoidal signal as shown in Figure 3.10. Thus the second distribution model used for vocalization is the sinusoidal distribution with the equation [14]

$$f_X(x)=\begin{cases}\dfrac{1}{\pi\sqrt{(a^2-x^2)}}, & for\,|x|<a\\ 0, & otherwise\end{cases} \tag{3.13}$$

where *a* is the amplitude of the sinusoid and *x* is the random variable.

In both saturated and nonsaturated cases of the vocalization models the distribu-

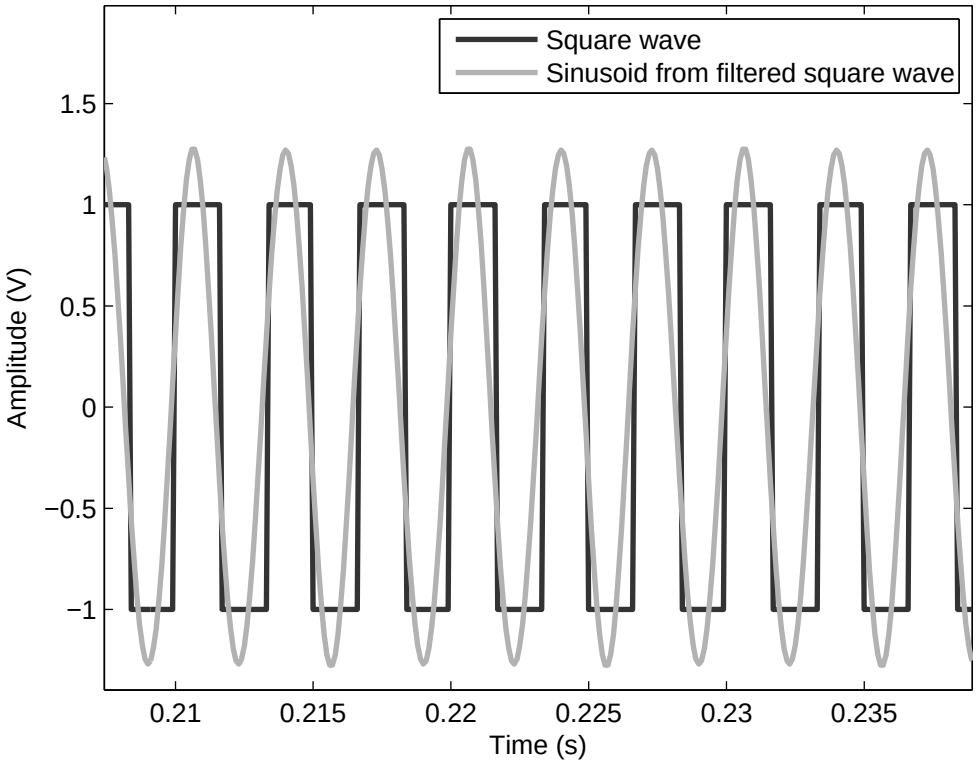


Figure 3.10 Square wave and resultant sine wave after application of a low-pass filter.

tion seemed unchanged when either performing the amplitude normalization method or segmentation. In cases of saturation, this can be attributed to the amplitude of the audio envelope being uniform.

In both cases of the vocalization models, the signal is expected to be periodic because of the repetitive nature of a vocalized sound. This further differentiates vocalization from an obstructive snore sound.

3.2.3 Classification using the sound source models

3.2.3.1 Distribution comparison models

The method of histogram classification pursued here compared the histogram of a sound to several distribution models. The models given as a Gaussian distribution for normal breathing, equation 3.10 for snoring, and a Gaussian distribution or equation 3.13 for vocalization were used to create distribution mixture models, as will be described next.

Twenty-one distribution models were made using two mixture models. The first was a mixture of Laplace and Gaussian distributions. These two distributions were used to differentiate between normal breath sounds and snoring breath sounds. This mixture's equation is

$$P(x, p_h) = p_h \times \frac{1}{\sqrt{(2\pi)}} e^{\left(\frac{-x^2}{2}\right)} + (1 - p_h) \times \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{3.14}$$

where p_h is the probability of the Gaussian distribution and x is the random variable. This mixture was used because each breath sound was made up of portions of different kinds of sounds independently in time. For example a nonobstructive snore typically begins as a normal breath and then changes to a snore sound as the pharynx closes.

Assuming that the snore sound is significantly louder than the normal breath sound, these two sounds can be considered sequential. Using the values of x and p_h, a matrix M_{GL} was created with dimensions of 11×101. To create this matrix, p_h was incremented from 0 to 1 in steps of 0.1 and x was varied from -3 to 3 in steps of 0.06. Each of the 11 rows corresponded to a different distribution of the mixture for each value of p_h. An overlay of the eleven models is shown in Figure 3.11.

The second mixture model was the mixture of the Laplace distribution, $f_X(x)=1/(2b)e^{-|x|/b}$, with a sinusoidal distribution, $f_{XS}(x)=\frac{1}{\pi\sqrt{(1-x^2)}}rect(\frac{x}{2})$. This model was useful in the detection of vocalization versus snoring sounds. Vocalization is typically sequentially spaced with another sound source such as a normal breath sound, so the distribution of the sinusoidal random variable was mixed with that of the Laplace distribution. This produces the mixture

$$P(x, p_S)=p_S\times\frac{\sqrt{2}}{2}e^{-\sqrt{2}|x|}+(1-p_S)\times\frac{1}{\pi\sqrt{(1-x^2)}}rect(\frac{x}{2}) \qquad (3.15)$$

where p_S is the probability of the Laplace distribution, that was incremented from 0 to 0.9 in increments of 0.1, and x is the random variable, that was incremented from -3 to 3 in steps of 0.06. This created a matrix M_{LS} with dimensions of 10×101. The overlay of the 10 models can be seen in Figure 3.12.

Concatenating matrix M_{LS} with matrix M_{GL} created a matrix M_{GLS} with dimensions of 21×101, where each of the 21 rows represented a different mixture model.

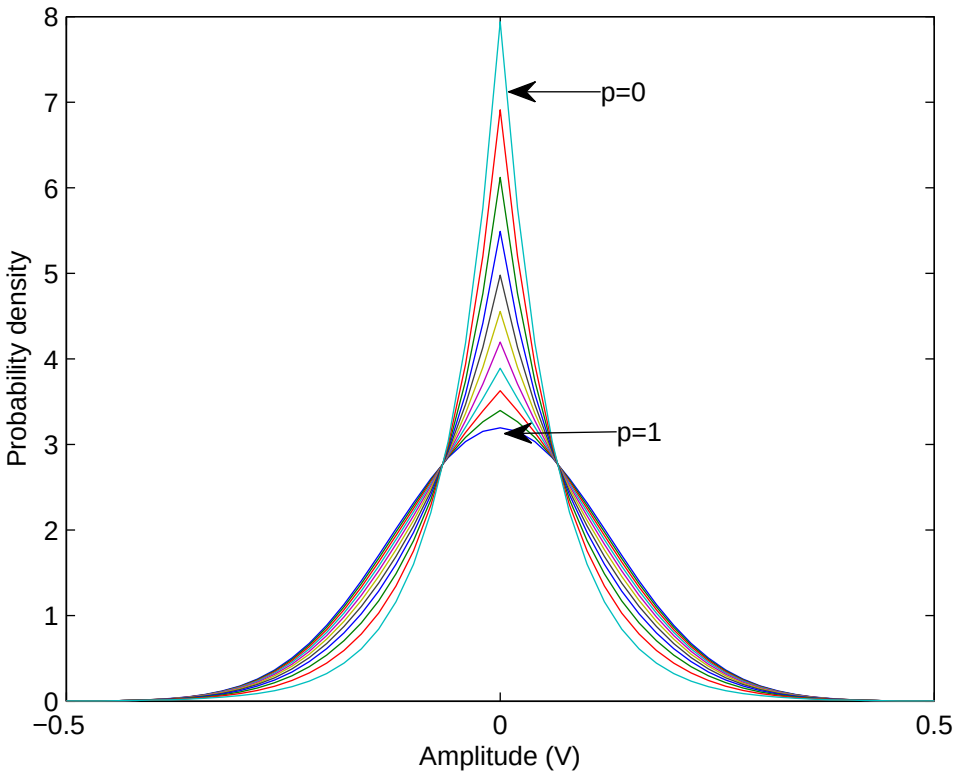


Figure 3.11 Overlay of 11 distribution models from the described Gaussian-Laplace mixtures of equation 3.12.

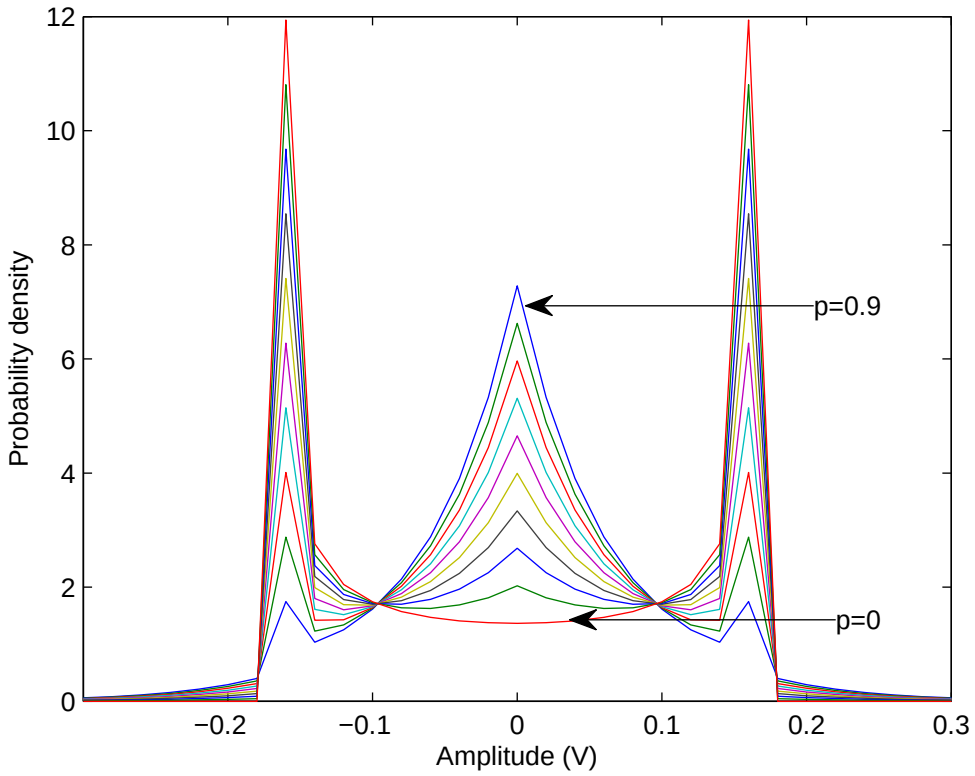


Figure 3.12 Overlay of 10 distribution models from the Laplace-sinusoid mixtures of equation 3.13.

3.2.3.2 Sound Segmentation

To ensure that the no pauses between breath sounds were missed and that the loudest portion of each breath sound was used for classification, each segmented sound found by the breath detection algorithm was further subdivided by removing the quarter of the samples with the lowest amplitude of the envelope of the segment. This was done by finding a quarter of the samples of the envelope of the signal that were the lowest in amplitude, and removing the corresponding samples of the signal from the sound to be classified. If there were any additional pauses that resulted from this removal of the samples, the resulting segments were considered independent sounds. The resulting segment or segments will be referred to as subsegments. As mentioned in Section 3.1.4 the sounds from inspiration to expiration can vary from an obstructed sound to a normal breath sound. This additional segmentation was performed to isolate different sounds to be classified. An example of a segmented audio envelope is shown in Figure 3.13a. The subsegments were required to be greater than 0.3 seconds in length or they would be discarded. This was done because it was observed that pauses between inspiration and expiration were missed by the breath detection algorithm if the inspiratory pause was too short. An example of this is shown in Figure 3.13b. This also ensured that the loudest portion of the segment was used to classify the signal and exclude any pause sounds. Differentiation between individual sounds is important to breath classification because the sound during inspiration can be significantly different from the sound during expiration in the case of snoring or vocalization, and if classified together could result in a misclassification of a period of normal breathing for that of respiratory obstruction. Although differentiation between inspiration and expiration was not performed, this step

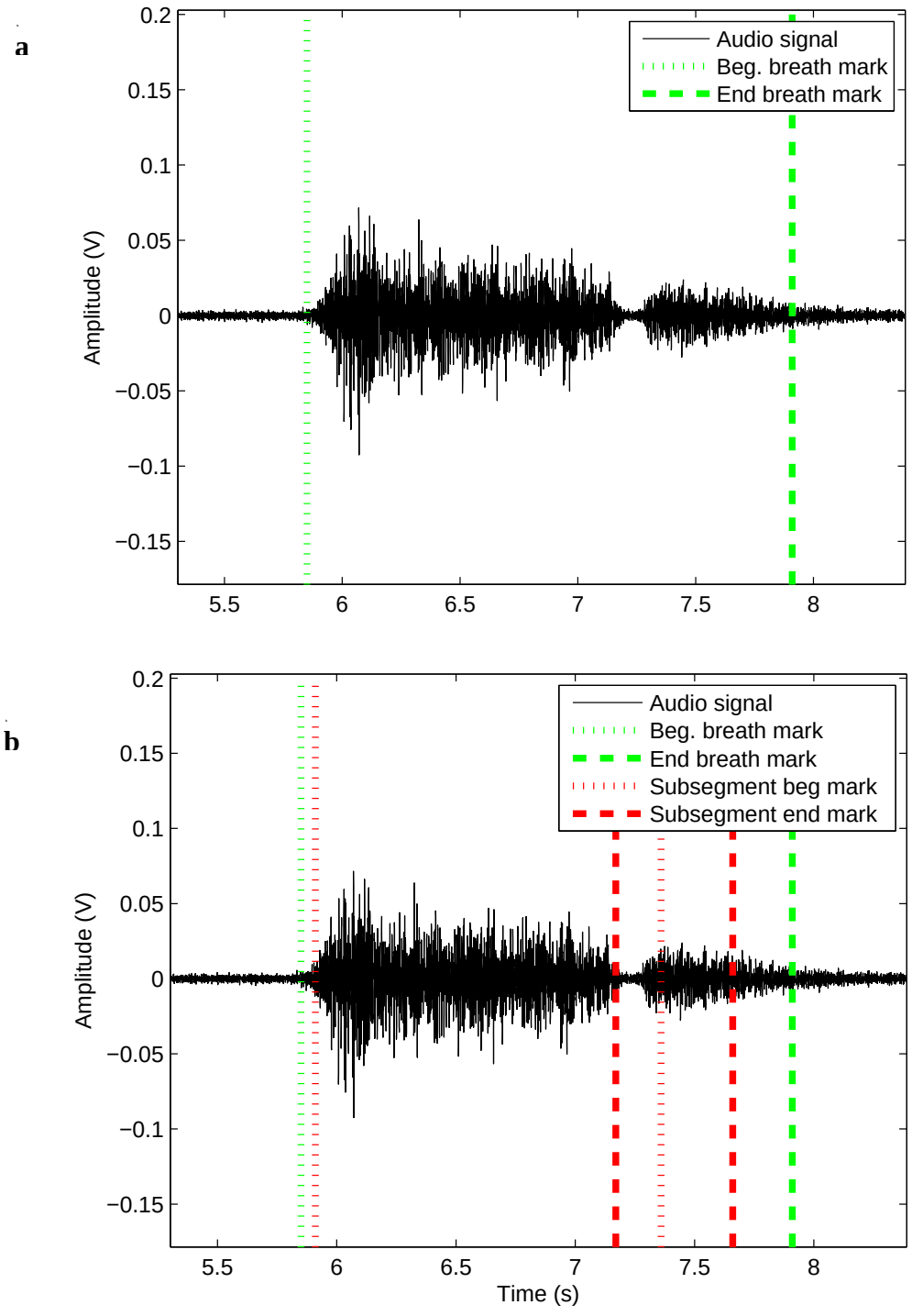


Figure 3.13 Filtered audio showing segmentation markers (a) Segmented breath sound. (b) Segmented breath sound with subsegments.

c

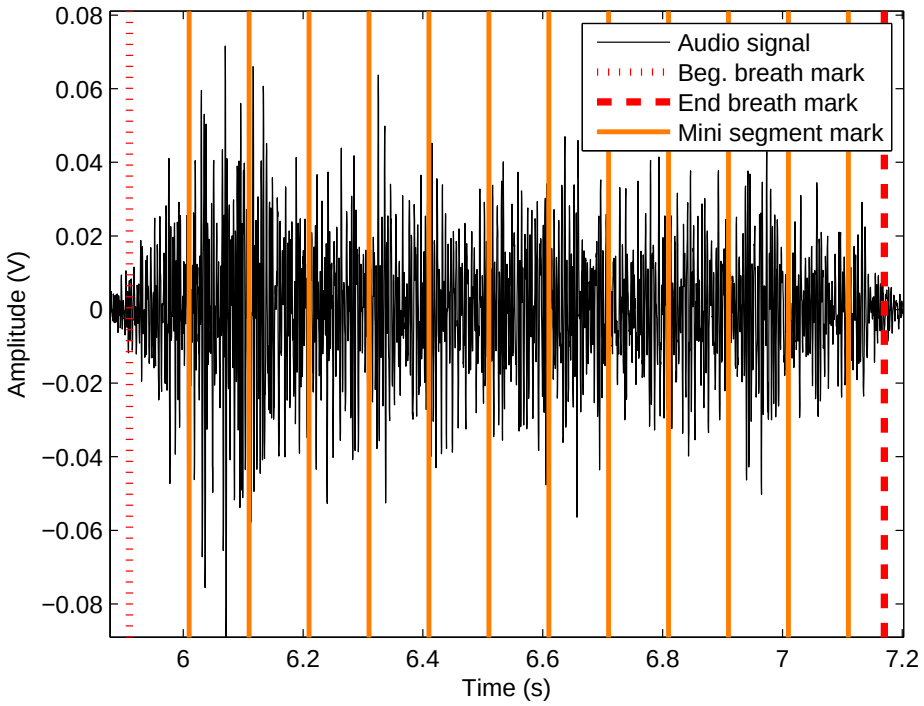


Figure 3.13 cont'd (c) Mini segmentation of a subsegmented breath sound.

helped insure that inspiratory sounds and expiratroy sounds were separated for classification purposes. Chapter 2 did not attempt to differentiate inspiratory sounds from expiratory sounds, instead breaths were detected in order to find periods of apnea. Because of this some detected breaths included both the inspiratory and expiratory sound.

3.2.3.3 Histogram classification procedure

The signal used to classify the sound was the filtered signal that was described in Chapter 2 with the example shown in Figure 2.4b. To summarize this process, the input signal of 22 kHz was first filtered by a digital low pass filter with a cutoff frequency of 1500 Hz and then down-sampled to 4410 Hz. This signal was then filtered with a high-pass filter with corner frequency of 75 Hz to remove heartbeat sounds and other noise.

Each sound subsegment was further divided into uniform 0.1-second segments consisting of 441 samples that were called mini segments, as shown in Figure 3.13c. The mean and standard deviation for each mini segment were calculated. As stated in Section 3.2.2.2 the smaller segments were used instead of the amplitude normalization method. Both were successful in removing the variation of the envelope from the signal. In addition the mini segments provided a uniform-length signal to classify. The mean was subtracted from the mini segment signal and the result was divided by the standard deviation to normalize the sound segment. The histogram of the normalized mini segment signal was then calculated using 101 bins ranging in value from -3 to 3 in steps of 0.06. The value in each bin was then divided by 441 and divided by the bin width of 0.06 to normalize the area of the histogram to unity. The histogram was then compared to the 21 model distributions described above. The comparison was made by summing the absolute difference between each of the 101 points of the model and the histogram of the

audio segment. The model that had the lowest summed difference was classified as the closest model.

As a preliminary test a subjective test was used to determine if this classifier could differentiate between snore sounds and normal breath sounds. The 843 segmented breath sounds were classified by the author. The segments were classified using the audio signal alone. He listened to each segment and classified it as either related to snoring or related to normal breathing. He classified 263 of these sounds as related to snoring. Each of the sound segments was then classified using the Laplace-Gaussian mixture only. The comparison of the classifiers is shown in Figure 3.14 where the classification of the author is shown as a black bar graph for normal breath sounds and a light-gray bar graph for snore sounds. When a decision threshold of $p_h=0.7$ was selected as the decision threshold, the classifier performed with an overall error rate of 7.24%. This method of testing had a subjective testing standard. This test was only meant to gain confidence in the validity of this classifier. The final classifier used for this project was classified against human classifications based on flow and chest movement, not on sounds classified manually. Although this preliminary test only involved the classification of snoring to normal breathing and did not include other sounds such as vocalization, the results will be used in the final classifier to differentiate between some breath sounds.

3.2.3.4 Periodicity classifier

Histogram classification has been shown to be able to classify the difference in most cases of a snore sound when compared to a normal breath sound. The histogram classifier does not, however, differentiate between productive snores and obstructive snores. One shortcoming of a histogram classifier is that it removes all dependence on

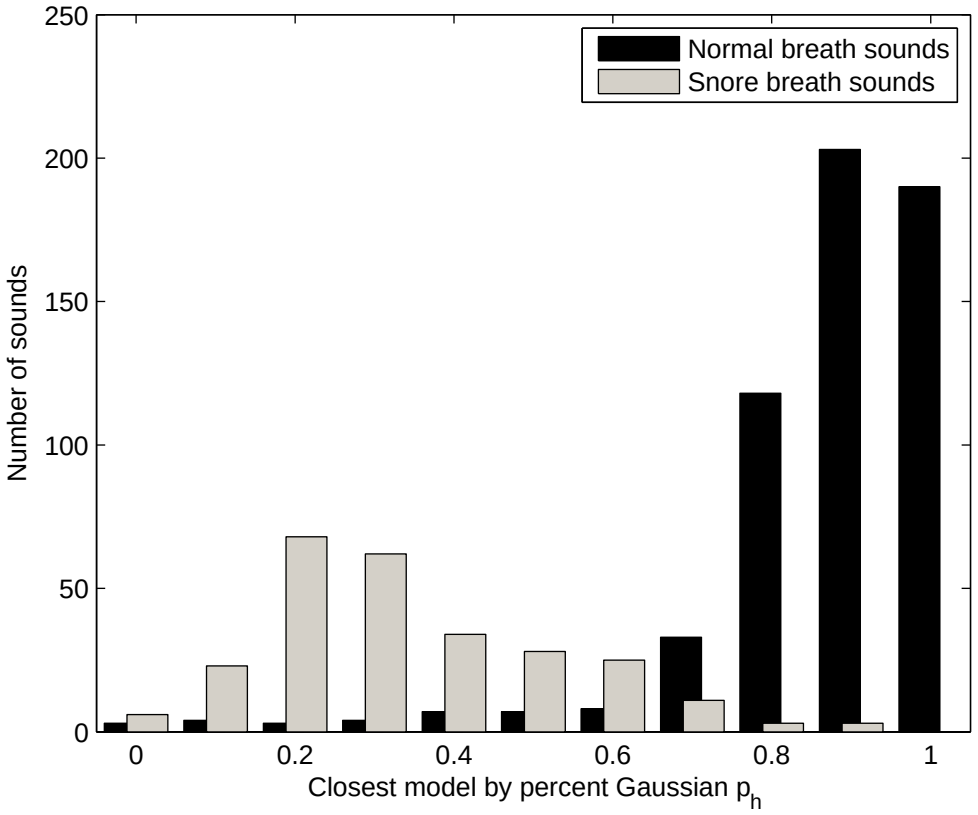


Figure 3.14 Classification of sounds by closest distribution when compared to human standard.

the sequence of the signal. It was hypothesized that the overall outcome of classification could be improved if a second classifier was used. Because of the lack of dependence on time of the histogram classifier, methods of classification using the time domain as an indicator were considered.

Comparing the models of the productive snore sound, the obstructive snore sound and the normal breath sound, one feature stood out. In a normal breath sound there was no repetitive sound that was distinguishable as with the obstructive snore sound. In a productive snore sound, the slap sound was observed to repeat. To take advantage of this difference in the sounds, a novel approach to find the repetitive nature of a sound was formulated. A tonal signal repeats itself and can be spectrally identified by a spike at the frequency of the fundamental component. However, if the repeating signal is not tonal the signal may have many spikes throughout the spectrum of the signal. However, a nontonal periodic signal can also be easily identified using the circular autocorrelation of the signal.

Circular autocorrelation is a specific form of circular convolution. In discrete circular convolution two signals of the same length, $x_1(n)$ and $x_2(n)$, are convolved in the following manner. Given that both $x_1(n)$ and $x_2(n)$ have N samples, the output of the circular convolution is $x_3(n)=circconv(x_1(n),x_2(n))$, where $circonv()$ denotes the circular convolution of the two values input. This operation is performed by [16]:

$$x_c(n)=\sum_{l=0}^{N-1} x(l)x((n-l)_N) \tag{3.16}$$

where $x((n)_N)$ denotes a periodic repeating signal with period N. Circular autocorrelation uses the same equation as stated above with the special case that $x_1(n)=x_2(N-n)$. Circular autocorrelation is an advantage over linear autocorrelation because it makes use of the

entire length of $x_1(n)$ rather than padding the input vectors with zeros as in linear autocorrelation.

Circular autocorrelation was performed on the mini segments described, where $x_C(n)=circconv(x_1(n),x_1(N-n))$. The standard deviation σ_C of $x_C(n)$ was calculated and the ratio of σ_C to the absolute maximum of $x_C(n)$ was calculated. Some assumptions of the input signal $x_1(n)$ were that it had a mean of zero and that it was real valued. The maximum used for this value in this dissertation is $\frac{\sqrt{2}}{2}$ because a perfectly correlated periodic signal such as a sine wave produces a circular autocorrelation that is a perfect sine wave and no signal was observed to have a value greater than this. The standard deviation of a sine wave is $\sigma_{sine}=\frac{\sqrt{2}}{2}\times a_{sine}$ where a_{sine} is the maximum amplitude of $x_C(n)$. The theoretical minimum for this ratio for a zero-mean segment of 441 samples is 0.0476, which happens when $x_C(n)$ is a vector of 440 zeros with one value of one. This excludes the possibility that this is a vector of 440 zeros, in which case the ratio would be undefined because the standard deviation would be zero and the absolute maximum would be zero. Because the segments were taken from detected sounds with noise, the scenario of having all zeros was not considered a possibility.

The value calculated by this ratio is a measure of how periodic the input signal is and was called the periodicity index. A productive snoring signal would exhibit a more periodic value and would have a value closer to the maximum of the periodicity index. A nonperiodic signal such as a breath sound would have a value that is closer to the theoretical minimum. The threshold chosen for this classifier was a ratio of 0.25. This value

was chosen by observation of the values produced by snore sounds versus the values produced by a normal sound in the training set of data. This threshold was identified as sounds were played and listened to by the author as the periodicity index was displayed.

3.2.3.5 Two-dimensional classifier

Using the values returned from both the periodicity classifier and histogram classifier, a series of points was created. Because each subsegment to be classified contained multiple mini segments, the classification of the subsegment had to be classified from multiple mini segment classifications. The x-axis on Figure 3.15 corresponds to the index of the closest histogram to the mini segment classified. The index has 21 possibilities of the closest distribution mixture described in Section 3.2.3.1. The y-axis is the periodicity index as described in Section 3.2.3.4. The five regions shown in Figure 3.15 correspond to five classifications of sound. If the classification of the mini segment falls into region I, the mini segment has a predominantly Gaussian distribution and is predominantly nonperiodic. If the classification of the mini segment falls into region II, the mini segment has a predominantly Gaussian distribution and is predominantly periodic. If the classification of the mini segment falls into region III, the mini segment has a predominantly Laplace distribution and is predominantly non periodic. If the classification of the mini segment falls into region IV, the mini segment has a predominantly Laplace distribution and is predominantly periodic. If the mini segment falls into region V, the mini segment has a partially sinusoidal distribution because the distribution of this mini segment is correlated with a distribution model that is a mixture of a sinusoid and Laplace distribution.

Examples of sounds classified in the regions described were found by the author

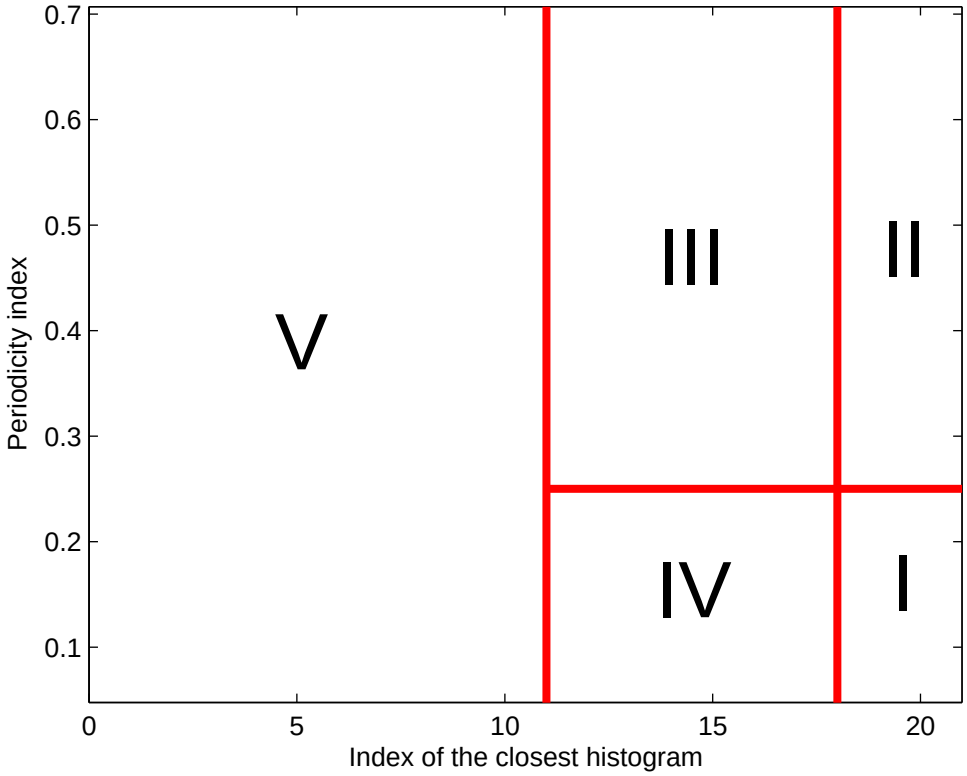


Figure 3.15 Regions of the two-dimensional classifier.

from the training data set. The classification of the mini segments by the two-dimensional classifier is shown for each region for each given sound. The author described the sounds as a normal breath, nonsaturated vocalization, saturated vocalization, non-obstructive snoring, and obstructive snoring. Each sound described here was classified by the author by listening to the audio signal and viewing flow and RIP data from the training set. Listening to the audio signal was used to determine if the signal was a normal breath, vocalization, or snore sound. The RIP and flow data were used to determine if the snore sounds were obstructed snores or non-obstructive snores. The decision boundaries of Figure 3.15 were found in the following manner: The rightmost vertical boundary was chosen based on the preliminary study of Section 3.2.3.3. The vertical boundary separating Region V from Region IV and Region III was chosen so that if the sound showed saturation by having a histogram that most closely matched any of the distributions of equation 3.15 it fell into Region V. The horizontal boundary was chosen based on the observations from listening to the audio sound while viewing the flow and RIP data in this manner: The author listened to segmented sounds from the training set. If the sound was determined to be a snoring sound by the author, the flow and RIP signals were then observed. Using these data the author decided whether the sound provided the flow volume to clear the 180 mL respiratory deadspace. From the observations of several sounds from the training set, the author set a boundary separating the obstructive sound from the nonobstructive sound as seen in Figure 3.15. Region I is consistent with a normal breath because it has a predominantly Gaussian classified distribution and very little periodicity of the signal. An example of the classifications of the mini segments for a normal breath sound is seen in Figure 3.16 in dark-blue markers. It is noted that not all

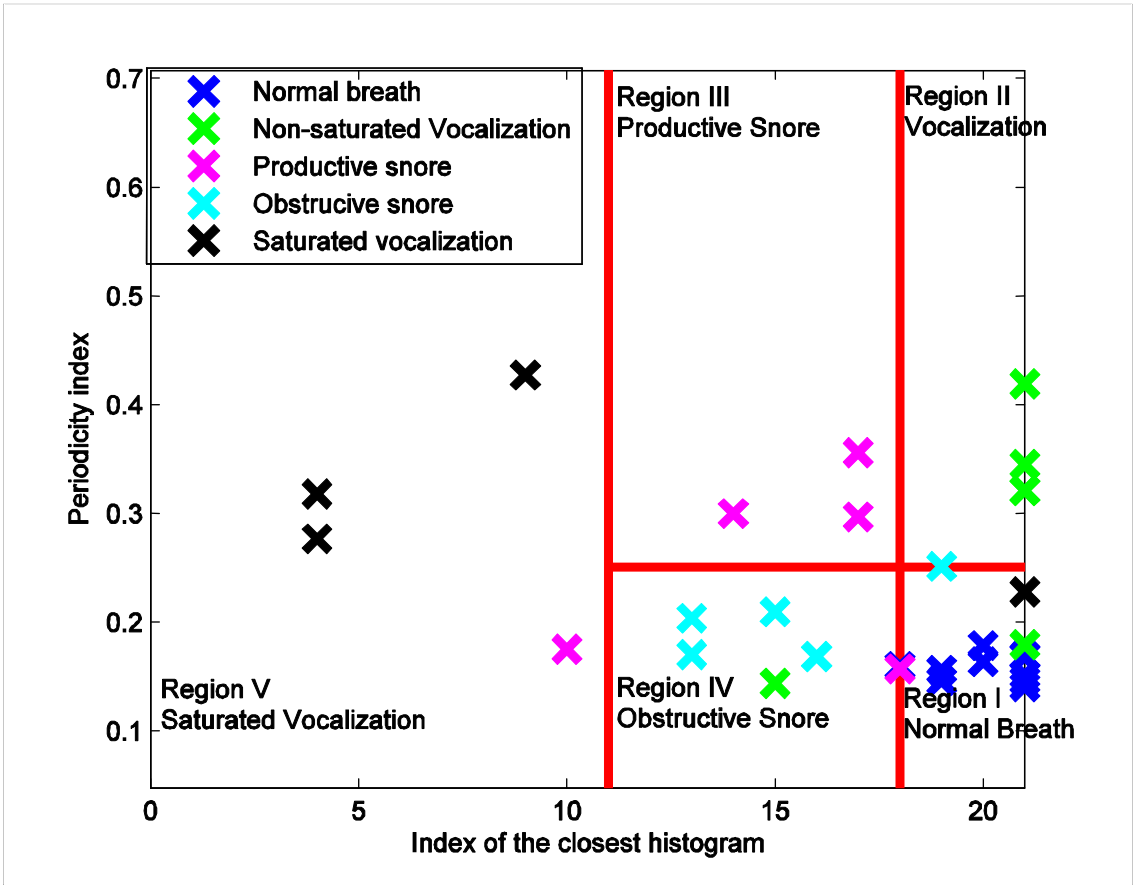


Figure 3.16 Examples of classifications of mini segments of five different types of sounds by the two-dimensional classifier. Dark-blue markers are from mini segments of a normal breath sound. Green markers are from mini segments of a non-saturating vocalization sound. Magenta markers are from mini segments of a productive snore sound. Light-blue markers are from mini segments of an obstructive snore sound. Black markers are from mini segments of a saturated vocalization sound. The x-axis represents the index of the closest matching histogram from a series of mixtures. The y-axis represents the periodicity index of each mini segment classified.

of the classifications of the mini segments are within the boundary set for this sound. The majority of the signals, however, do lie within this region, just as with the rest of the examples. Region II is consistent with vocalization because it has high periodicity but is more correlated with a Gaussian distribution than a Laplace distribution. An example of the classifications of the mini segments for a vocalized sound is seen in Figure 3.16 in green markers. Region III is consistent with a productive snoring sound. The sound in this case has a predominantly Laplace distribution and is predominantly periodic. An example of the classifications of the mini segments for a productive snore sound is seen in Figure 3.16 in magenta markers. Region IV is consistent with an obstructive snore sound. The sound has a heavy-tailed distribution and is not periodic, meaning that the airway closed and opened randomly during the mini segment. Previously it was believed that the periodicity would differentiate this sound from a normal breath, but the lack of periodicity differentiates it from the productive snoring sound. It is believed that the difference comes from a steady flow during a nonobstructive snore versus a varying rate flow during partial obstruction. An example of the classifications of the mini segments for an obstructive snore sound is seen in Figure 3.16 in light-blue markers. Region V is consistent with saturated vocalization. An example of the classifications of mini segments for a saturated vocalization are shown in Figure 3.16 in black markers.

The method of finding the decision boundaries for this classification method was performed for several sounds deemed to belong to each kind of sound by the author. These sounds were then viewed to see what features the sound had. This observation method led to the selection of the decision boundaries for the sounds described. This was not performed for all sounds in the training set because of the number of sounds recorded.

Instead sounds were selected that preceded the periods of apnea detected by the breath detection method and also other periods that were known to have vocalization or productive snoring.

If over half of the classified mini segments from the sub-segment were found to be in Region I, II, III, and V the sub-segment was classified as a normal state. This is because the sounds associated with these regions are consistent with a patient that is breathing with an open airway or is awake. If over half of the classified mini segments fell into region IV the sound was deemed to be related to obstruction and the markers segmenting the breath segment were removed from the breath marker set. Using the minimum of half of the markers in the region was chosen based upon the observations of the groupings of the classification markers. As shown in Figure 3.16, the classification markers for a single sound were not always consistent with one region. This prompted the use of the majority of the markers in a combination of Regions I, II, III, and V or the majority of the markers in Region IV as the classification of the sound. The apnea detection algorithm then found any additional periods of apnea due to segmentation markers being removed. If an existing period of apnea became longer, or a new period of apnea was found due to the removal of obstruction sounds, the period was classified as RO; otherwise the existing period was classified as a RD.

3.3 Results

Using the data from Chapter 2, the periods of apnea were divided into three categories using the two-dimensional automated acoustic classifier and the classification standard as described in Section 3.2.2. The three classification categories were RD, RO and Normal. The comparison between the two classifiers creates nine possible categories

and is shown in Figure 3.17 for the training set and 3.18 for the testing set. Figure 3.18 does not show all nine categories because not all possible combinations had a value greater than zero. These categories also produce a three by three truth table as shown in Table 3.1 for the training set and Table 3.2 for the testing set.

3.4 Discussion

In the project proposal of this project the goal of 95% sensitivity and specificity in differentiating RO from RD was set. The automated acoustic classification method did not perform as well as hoped. The results in Tables 3.1 and 3.2 show that the automated acoustic classifier misclassified 47.4% of the periods identified as RO by the classification standard as RD in the training set. It also did not classify any periods identified as RO by the classification standard as RO in the testing set. Only one of the areas was classified within the goal of the project proposal. This was the periods classified as RD in the testing set.

Despite the large error in the differentiation between types of apnea, the classification of the breath sounds does not decrease the overall reliability of detecting apnea. The automated acoustic classifier did not mis-classify periods of normal breathing as obstructive periods, which would make the overall apnea detection less reliable. Since normal breathing sounds were the most common sound recorded, the majority of the sounds classified were classified in a way that did not degrade the performance of the apnea detection algorithm described in Chapter 2.

The classifier failed most commonly when differentiating RO as identified by the classification standard from other signals. The total number of periods of apnea classified as RO by the standard was a small percentage of the total periods of apnea. The

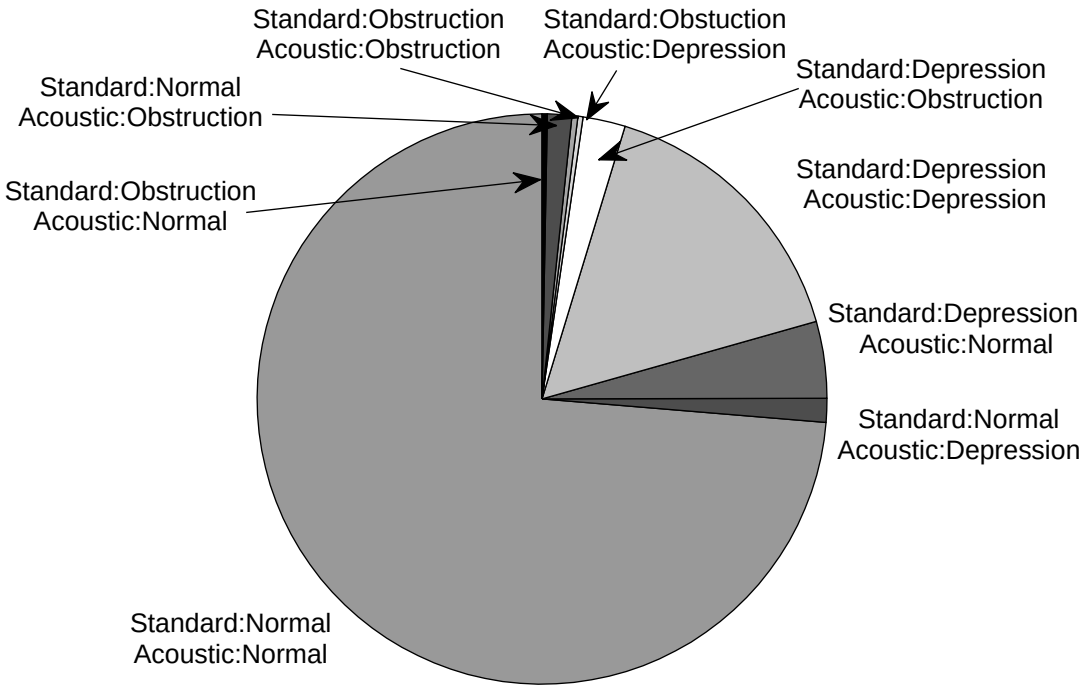


Figure 3.17 Classification results of the training data set.

.

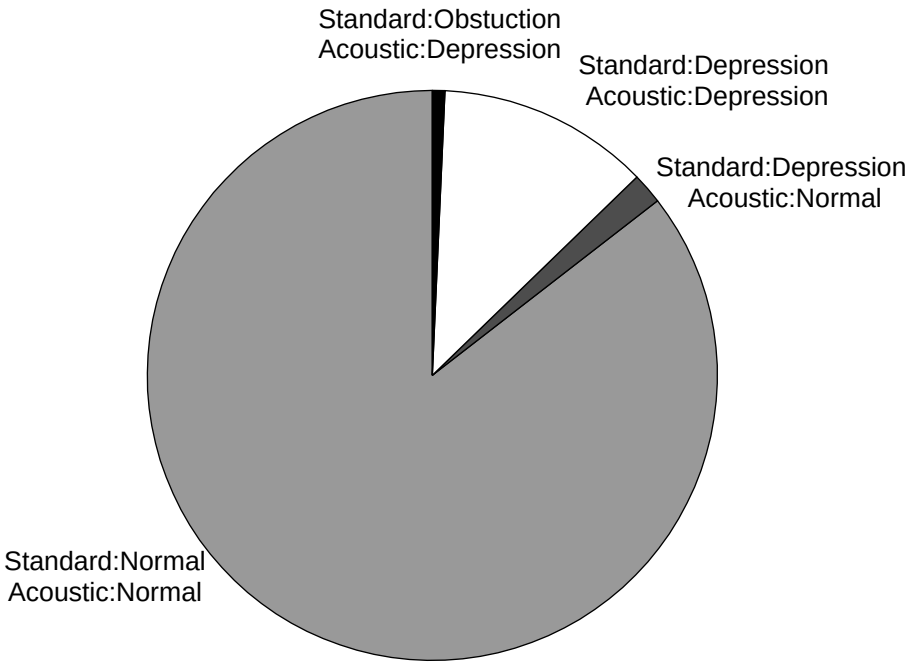


Figure 3.18 Classification results of the testing data set.

Table 3.1 Training set apnea classification data comparing the classification standard to automated acoustic classification.

		Classification Standard		
		Normal	***RD***	***RO***
Automated acoustic classification	***Normal***	2238	42	9
	RD	135	483	9
	RO	41	73	10

Table 3.2 Testing set apnea classification data comparing the classification standard to automated acoustic classification.

		Classification Standard		
		Normal	***RD***	***RO***
Automated acoustic classification	***Normal***	3167	0	0
	RD	65	446	26
	RO	0	0	0

small sample size of RO periods makes it difficult to accurately create a functional algorithm and be able to test its validity.

One reason the classifier may have failed in differentiating RO from RD was the procedure during an obstructive period. Subjects were allowed to obstruct until their blood oxygen saturation dropped to 85%. Obstructive sounds were heard by the author leading up to these periods of RO. The anesthesiologist present alleviated the RO by performing a jaw thrust maneuver. The subject would then breathe normally again. The anesthesiologist would then attend to other duties and allow the subject to obstruct again. The sudden drop of the jaw did not allow for periods of partial obstruction and no sounds associated with obstruction were recorded. The classification standard would have defined these periods as obstruction because of the chest and abdomen movements, but the automated acoustic method would have only heard normal breathing followed by a period of silence. Obviously the anesthesiologist was aware of the source of apnea at this point, and the mis-classification would have caused little confusion in a clinical setting. However, this would lead to a high overall classification error rate.

Considering the mis-classifications and the procedures used to record the data, classification is unlikely to be performed using the acoustic signal alone. In addition the reliance upon only the audio signal for apnea detection is very risky because sounds from the trachea are not the only sounds that can be detected. Examples of sounds that are not generated at the trachea are ambient talking, music, or machine noises that are loud enough to be detected at the microphone within the stethoscope cup.

Compared to the other approaches presented in Section 3.1.3 this approach uses a combination of the primary forms of classifying individual breaths and classifying peri-

ods of apnea. It also uses data that were recorded from sedated subjects rather than patients undergoing PSG in a non-sedated state.

3.5 References

[1] A.K. Ng and T.S. Koh, “Using psychoacoustics of snoring sounds to screen for obstructive sleep apnea,” *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2008*, 2008, pp. 1647-1650.

[2] R. Jane, J. Sola-Soler, J.A. Fiz, and J. Morera, “Automatic detection of snoring signals: validation with simple snorers and OSAS patients,” *Proc. 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2000, pp. 3129-3131.

[3] H. Hara, N. Murakami, Y. Miyauchi, and H. Yamashita, “Acoustic analysis of snoring sounds by a multidimensional voice program.,” *Laryngoscope*, vol. 116, Mar. 2006, pp. 379-381.

[4] H. Nakano, M. Hayashi, E. Ohshima, N. Nishikata, and T. Shinohara, “Validation of a new system of tracheal sound analysis for the diagnosis of sleep apnea-hypopnea syndrome.,” *Sleep*, vol. 27, 2004, pp. 951-957.

[5] D.W. Hudgel, “The role of upper airway anatomy and physiology in obstructive sleep apnea.,” *Clin Chest Med*, vol. 13, Oct. 1992, pp. 383-398.

[6] P. Steiropoulos, N. Papanas, E. Nena, E. Maltezos, and D. Bouros, “Continuous positive airway pressure treatment in patients with sleep apnoea: does it really improve glucose metabolism?,” *Current Diabetes Reviews*, Apr. 2010.

[7] “Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force,” *Sleep*, vol. 22, Aug. 1999, pp. 667-689.

[8] J.E. Remmers, W.J. deGroot, E.K. Sauerland, and A.M. Anch, “Pathogenesis of upper airway occlusion during sleep,” *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*, vol. 44, Jun. 1978, pp. 931-938.

[9] H. Heinze and W. Eichler, “Measurements of functional residual capacity during intensive care treatment: the technical aspects and its possible clinical applications,” *Acta Anaesthesiologica Scandinavica*, vol. 53, Oct. 2009, pp. 1121-1130.

[10] R.D. Yates and D.J. Goodman, *Probability and stochastic processes: a friendly introduction for electrical and computer engineers*, John Wiley & Sons, 2005.

[11] A.L.L. Egnel and M. Paiva, *Gas Mixing and distribution in the lung*, Dekker, 1985.

[12] V.P. Harper, H. Pasterkamp, H. Kiyokawa, and G.R. Wodicka, “Modeling and measurement of flow effects on tracheal sounds,” *IEEE Journal of Biomedical*

Engineering, vol. 50, 2003, pp. 1-10.

[13] M. Kompis and E.W. Russi, “Computer-based lung sound simulation,” *Medical & Biological Engineering & Computing*, vol. 35, May. 1997, pp. 231-238.

[14] K. G.D. Cain, K.V. Lever and A.Yardim, “Probability density functions of amplitude-modulated random signals,” *IEEE Electronics Letters*, vol. 34, 1998, pp. 1560-1561.

[15] S. Gazor and Wei Zhang, “Speech probability distribution,” *Signal Processing Letters, IEEE*, vol. 10, 2003, pp. 204-207.

[16] J.G. Proakis and D.G. Manolakis, *Digital signal processing*, Pearson Prentice Hall, 2007.

CHAPTER 4

ADAPTIVE NOISE CANCELLATION

4.1 Introduction

4.1.1 Goal of adaptive noise cancellation

The third goal of this project was to find the validity of using noise cancellation to reduce ambient sounds recorded within the precordial stethoscope. Noises generated outside of the stethoscope that are not wide sense stationary (WSS) can cause signals to be recorded within the stethoscope that are loud enough to be counted as a detected breath. If these sounds are recorded during a period of apnea the acoustic signal will not be able to detect it as a period of apnea and it is classified as a period missed by the acoustic apnea detection algorithm.

For this reason a method to reduce disturbances caused by ambient sounds was explored. The method used was an adaptive filter requiring a secondary microphone to record ambient sounds.

4.1.2 Use of precordial stethoscope to limit ambient noise

The stethoscope cup used was a heavy precordial cup shown in Figure 4.1. The stethoscope cup was designed to amplify signals detected within the cup. Physiologically the skin within the stethoscope cup creates a diaphragm that acts like a loudspeaker for the vibrations on the skin. The metal stethoscope attenuates external signals from

Figure 4.1 Metal stethoscope cup used to attenuate ambient sounds and amplify tracheal sounds.

entering the cup. Acoustics of a stethoscope cup attenuation and amplification can be characterized through experimentation [1], but this can change depending on the placement of the stethoscope cup on the skin, the tightness of the skin within the cup, and the placement of the stethoscope cup on the trachea. Observations from Chapter 2 have shown that sounds such as talking or machine alarms can be loud enough to be detected as breath sounds.

4.1.3 Overview of adaptive noise cancellation

4.1.3.1 Adaptive noise cancellation

Adaptive noise cancellation is a method of signal processing that uses multiple sources of signal to produce a desired signal. A simple adaptive filter has two inputs, with block diagram as shown in Figure 4.2. Signal d is the primary input and signal x is the reference input. The reference input is then filtered by the weights $\boldsymbol{w}$ of the adaptive filter and an estimate of d is created called y. The difference of the filtered signal y and desired signal d is then calculated to create the error signal $e=d-y$. The error signal is then used as an input to the adaptive filter to update the filtering weights $\boldsymbol{w}$ as will be described the next section [2].

4.1.3.2 Least-mean-square adaptive filtering

The adaptive filter algorithm chosen to adapt the weights $\boldsymbol{w}$ can be calculated using the least-mean-square (LMS) algorithm. The LMS algorithm is the most widely used adaptive filter algorithm due to its stability, robustness and simplicity [2]. This filter has many variants, including the simplified LMS algorithm and the normalized LMS algorithm.

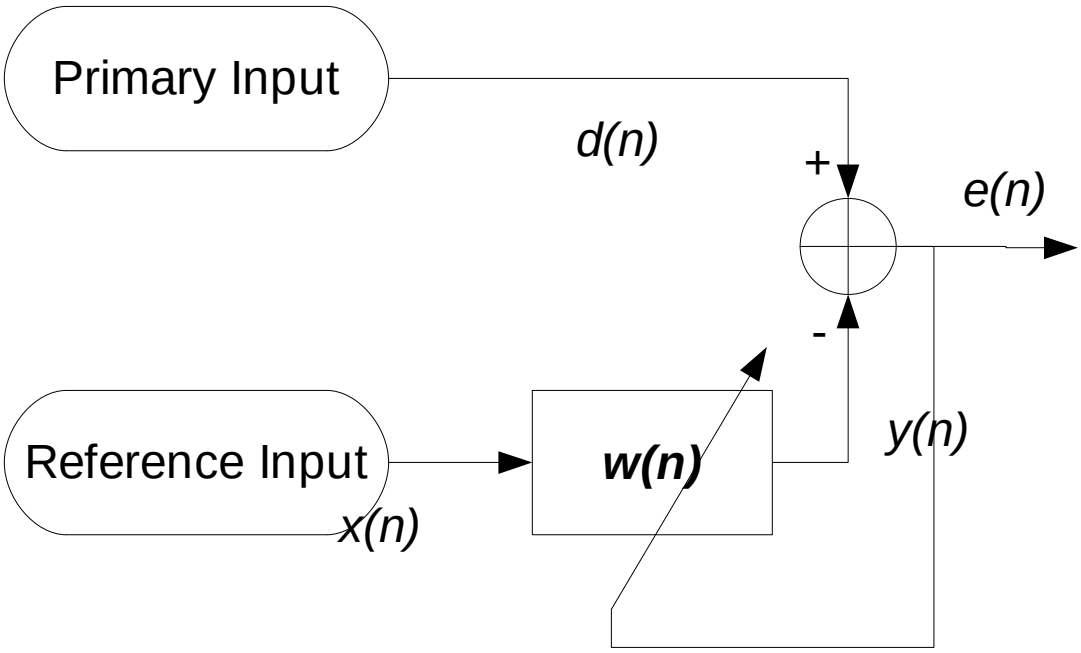


Figure 4.2 Block diagram of a two-input adaptive filter.

The classic LMS algorithm updates the weights $\boldsymbol{w}$ by

$$\boldsymbol{w}(n+1)=\boldsymbol{w}(n)+2\mu e(n)\boldsymbol{x}(n) \tag{4.1}$$

where $\boldsymbol{w}(n+1)$ is the new vector of filtering weights with length N, $\boldsymbol{w}(n)$ is the current vector of filtering weights with length N, μ is the step size parameter, $e(n)$ is the error signal, and $\boldsymbol{x}(n)$ is a vector of recorded signals with length of N.

The simplified LMS algorithm has three forms and is similar to the classic LMS algorithm. The three forms shown here are the sign, signed-regressor, and sign-sign algorithms respectively:

$$\boldsymbol{w}(n+1)=\boldsymbol{w}(n)+2\mu\ sign(e(n))\boldsymbol{x}(n) \tag{4.2a}$$

$$\boldsymbol{w}(n+1)=\boldsymbol{w}(n)+2\mu\ e(n)sign(\boldsymbol{x}(n)) \tag{4.2b}$$

$$\boldsymbol{w}(n+1)=\boldsymbol{w}(n)+2\mu\ sign(en)sign(\boldsymbol{x}(n)) \tag{4.2a}$$

The signed-regressor algorithm is favored for its ability to adapt similarly to the classic LMS but requiring fewer computations. The sign algorithm and the sign-sign algorithm do not converge as quickly [2] but are not much less computationally expensive than the signed-regressor algorithm.

The normalized LMS (NLMS) algorithm adds some complexity in order to improve stability. The general form of the NLMS algorithm is

$$\boldsymbol{w}(n+1)=\boldsymbol{w}(n)+\frac{\tilde{\mu}}{\boldsymbol{x}^{T}(n)\,\boldsymbol{x}(n)+\psi}e(n)\boldsymbol{x}(n) \tag{4.3}$$

where $\tilde{\mu}$ is the step size parameter and ψ is a small value to ensure that the denominator of the equation is never zero. The NLMS algorithm improves stability of the adaptation at the expense of computation. The step size parameter is normalized to the values of the input ensuring that if $\tilde{\mu}$ is properly chosen the output will never become unstable.

4.1.3.3 Stability of adaptive filters

The step size parameter μ in the LMS algorithm is the value that controls how quickly the filter adapts. If this value is too high the filter can become unstable and the output of the filter becomes useless. If its value is too low the filter will not adapt quickly to changes in the input signals. The maximum value that μ can be while maintaining the stability of the filter can be calculated as follows [2]:

$$\mu_{MAX} = \frac{1}{3\mathrm{tr}[\boldsymbol{R}]} \tag{4.4}$$

where $\boldsymbol{R}$ is the autocorrelation of the input calculated by:

$$\boldsymbol{R} = E[\boldsymbol{x}(n)\boldsymbol{x}(n)^T] \tag{4.5}$$

and tr[] is the trace of a matrix, defined as the sum of the diagonal of the matrix [2]. Although this can be calculated for every sample of the signal being filtered it is computationally expensive.

4.1.4 Use of adaptive noise cancellation in stethoscopes

Adaptive noise cancellation in stethoscopes has been performed for very noisy environments. Patel et al. [3] used an adaptive filtering algorithm to filter helicopter noise from cardiac and breathing sounds through a diaphragm stethoscope cup with a second microphone to record ambient sounds. Data were recorded on a subject in a soundproof room using this stethoscope and also using a pneumotachometer to measure respiratory flow rate. Sounds simulating being inside a helicopter were played inside the soundproof room. A real-time adaptive filter was used to monitor the progress of the filter. Postprocessing was performed using both an LMS algorithm with N=40 and μ=0.02 and an NLMS algorithm with N=40 and $\tilde{\mu} = 1.2$. Patel found that the NLMS

algorithm provided a significant improvement over both non-filtered data and the LMS filtered data.

Although the work done in [3] is closely related to the noise cancellation performed in this project, there are significant differences in the methods performed. The comparison of the respiratory rate of the acoustic data to that of the respiratory flow data was not discussed by Patel. In addition the adaptive filter was not applied when the reference signal was in a quiet environment. In a quiet enough setting the signal detected by the reference input can be uncorrelated with the additive noise on the primary input. When the two input signals of the adaptive filter are uncorrelated the reference input can increase the noise on the output when compared with the primary input. This dissertation defines this phenomenon as contamination. Another difference between the research performed in [3] and the current research was that Patel only used one kind of additive noise at one amplitude. The helicopter noise also did not change drastically and may be considered WSS. In this dissertation several different types of additive noise were used at several different additive gain amplitudes. An increase in the number of different kinds of sources and the amplitudes of these sources allows for the production of a more robust adaptive filter. Finally, this research uses an automatic breath detection algorithm to determine the validity of the adaptive filtering algorithm rather than a subjective argument.

4.2 Methods

4.2.1 Data set

4.2.1.1 Respiratory data set

Ideally the data used to test the adaptive filter algorithm would be recorded at the trachea on a subject as he/she lies in an anechoic chamber, as different sounds are played and recorded in the stethoscope cup while the subject performs breath-hold periods. This method was not pursued due to necessity of obtaining IRB approval, the lack of time to receive that approval, and budget constraints. Instead, data recorded from the data set as described in Chapter 2 were used to test the adaptive filter algorithms. From the periods of apnea identified by the respiratory flow meter and confirmed as true positives by the acoustic method, forty periods of apnea were selected on a manual basis. The periods were selected if the signal fifteen seconds prior to the apnea showed normal breathing on the flow rate signal. The data used included the respiratory flow rate from the CosmoII+ (Respironics, Wallingford, CT) and the raw acoustic signal recorded at the trachea in the stethoscope cup described.

Each data segment was marked as the breathing period and the apnea period. This distinction was made so that during breath detection the number of breaths detected could be counted toward a period of breathing or a period of apnea.

The standard deviation of the noise σ_{NF} was calculated for each breath sound segment using the method described in Chapter 2. The standard deviation of the noise served as a reference for the amplitude of the additive noise sound that will be described in the next section.

4.2.1.2 Additive noise signal data

The audio data in the 40 sets described in section 4.2.1.1 were assumed to be free of any major ambient artifact such as talking, machine sounds, or music. These data were also recorded without the use of a second reference microphone to record the disturbances coming from outside of the stethoscope cup. For this reason additive noise was recorded at a different time using two microphones and a phantom material representing human tissue. The additive noise was added to the acoustic breathing signal after the recordings were finished. This method allowed for the original signal to be known before noise was added and also allowed for different amplitude levels of additive noise to be tested without the need of multiple recordings. This method is also similar to a common image processing technique where an image assumed to be noise free is corrupted with noise, and filtering algorithms are tested by comparing the filtered image to the original image.

Two channels were recorded for each additive noise signal. One channel was the from the microphone within the stethoscope cup and the other channel was recorded from the reference microphone. Eight sound segments were used consisting of simulated Gaussian noise, talking, and several kinds of music. These sounds were chosen because they are common in an operating room environment.

A microphone (WM-56A103 Panasonic) was placed inside the stethoscope cup (Wenger #00-390-C, AINCA, San Marcos, CA) as was done during the recording of the breathing sounds. The cup was affixed by a double-stick disk (#2181 3M, MN) to a gelatin phantom made of edible gelatin formed inside a latex balloon. Gelatin was chosen as phantom that has similar properties to human soft-tissue. The balloon was

suspended above a table with the second microphone resting outside the stethoscope cup. Speakers were placed on a second table approximately 25 cm from the microphone assembly. The balloon suspension and use of two separate tables was an attempt to minimize mechanical coupling of the loudspeakers to the microphones. The speakers played the eight noise segments described above. The amplitude of the sound played by the speakers was adjusted to ensure that the external microphone was not saturated and that the signal was detected by the microphone inside the stethoscope cup. The data recorded at the microphones were digitized via an audio soundcard (SoundBlaster Audigy, Creative, Singapore) at a sample rate of 22 kHz directly to a computer hard-drive. A diagram of this setup is shown in Figure 4.3a.

4.2.1.3 Mixing the additive noise

The standard deviation σ_{IN} of the additive noise segment with signal $A_{IN}(n)$ recorded inside the stethoscope cup was calculated for each of the eight segments. The corresponding reference input will be called $A_{OUT}(n)$. The segment $A_{IN}(n)$ and the segment $A_{OUT}(n)$ were divided by σ_{IN} to normalize the sounds recorded inside the stethoscope. The signal $A_{IN}(n)$ was additionally divided by the standard deviation of the noise of the breathing signal σ_{NF} described above to normalize the amplitude of the additive noise signal to the amplitude of the noise of the breathing signal.

The normalized sound $A_{IN}(n)$ was added to the breathing sound $B(n)$, resulting in $X(n)$ as follows:

$$X(n)=B(n)+G_i \times A_{IN}(n) \tag{4.6}$$

where G_i is the gain applied to the additive noise. The gain G_i was not applied to the reference input $A_{OUT}(n)$ so that the reference signal was a consistent input to the adaptive

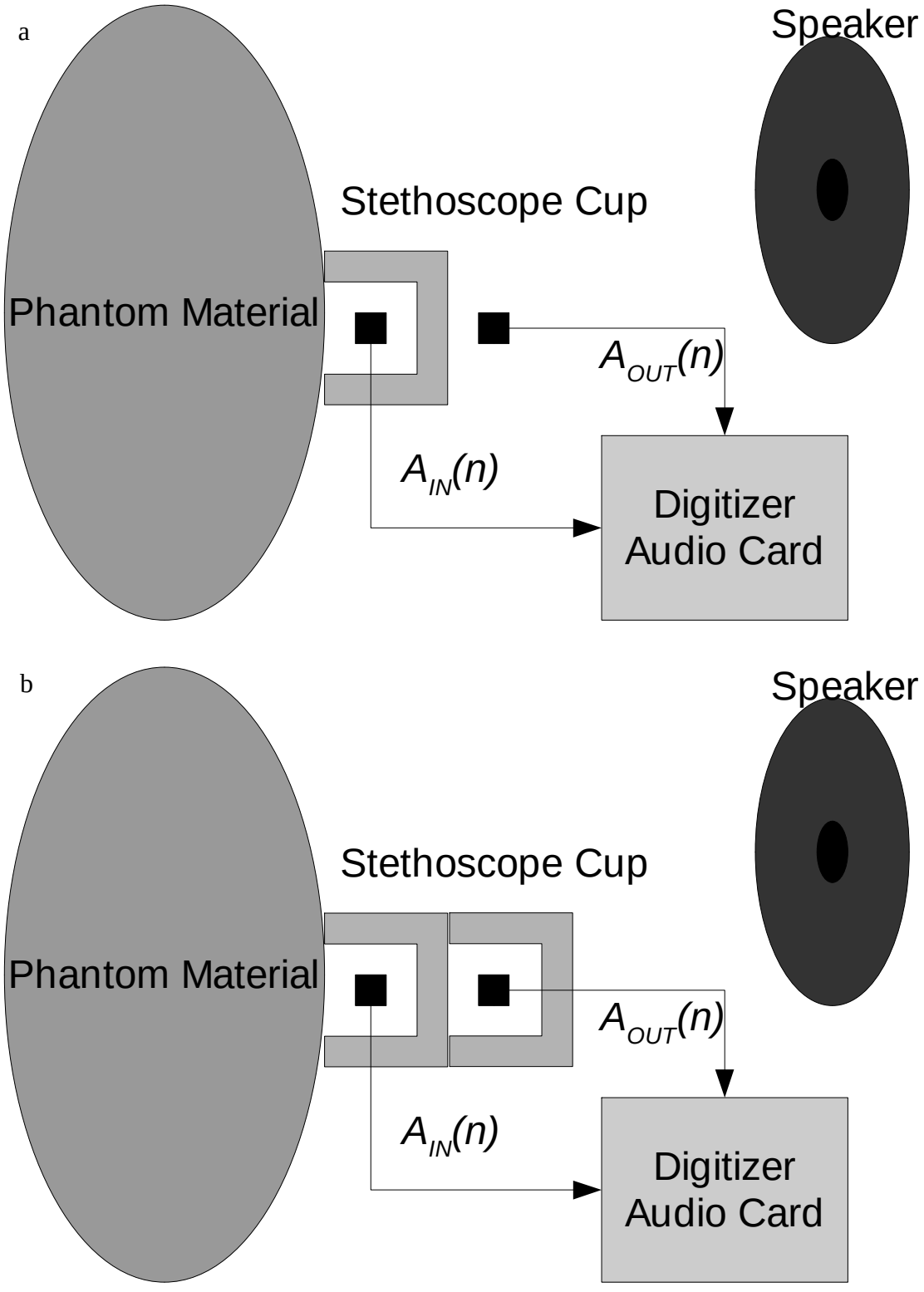


Figure 4.3 Diagram of the setup used to record the additive noise. (a) Reference microphone in open-air environment. (b) Reference microphone in stethoscope cup.

filter. The adaptive filter was processed with G_i having values of 0, 2, 4, 6, and 8. The value of 0 was chosen to test the effect of having an uncorrelated signal as the reference input. The other values of G_i were chosen to add a range of sound amplitudes that would be detected by the breath detection algorithm.

4.2.1.4 Filtering the mixed signals

The adaptive filter used will be described in Section 4.2.2. After performing the adaptation using this filter on the data described, it was noticed that the type of additive noise affected how well the filter performed. The sound sources that performed better were Gaussian noise, talking, and symphonic music. The method for determining how well the filters performed will be described in Section 4.2.3. Rock music caused some problems because of contamination of the signal even when the gain G_i was zero. After looking directly at the resulting waveforms and listening to them, the sounds that were not able to be filtered were strong impulses related to a drum beat or similar high-frequency sound. It was concluded that the impulse disturbance had a high enough frequency that the adaptive filter could not adapt quickly enough to remove the sound.

An additional problem that was noticed was that the signal of the reference microphone contained much higher frequency signals than the microphone inside the stethoscope cup. It was also hypothesized that the stethoscope cup attenuates signals in a nonlinear manner. This is a problem because the filter used to match the attenuation of the stethoscope cup is a linear filter. Matching a nonlinear filter such as the stethoscope cup has considerable challenges. Although this is possible, a single-order linear filter such as the NLMS adaptive filter algorithm described would not be sufficient. A nonlinear adaptive filter such as an adaptive polynomial filter could be a solution to this

problem [4], but it was decided that physically filtering the reference signal with a similar stethoscope cup would be a simpler solution.

The additive noise was recorded again as described above, with the one difference. The difference was that the reference microphone was placed inside an identical stethoscope cup as contains the primary microphone and affixed to the back of the primary microphone cup via a double-stick disk. A diagram of this setup is shown in Figure 4.3b. This setup did not test the amount of desired sound signal such as breathing that would pass through the primary stethoscope cup to the reference microphone. It is assumed that the reference microphone and cup can be sufficiently isolated to eliminate any of the sounds recorded on the trachea. The most important features of the reference microphone in this situation are that it is in proximity of the primary microphone, physically filtered similarly to the primary microphone, and the same kind of microphone as the primary microphone. This is because the additive noise is easiest to filter from the primary source when the reference source is as close as possible to the additive noise signal.

4.2.2 Adaptive filter

An NLMS adaptive filter was used to filter additive noise from the signal $X(n)$. The NLMS algorithm was chosen over the others to ensure stability of the filter. Using the adaptive filter shown in Figure 4.2, the signal $X(n)$ was used as the primary input d and the signal $A_{OUT}(n)$ was used as the reference input x. The output signal of the adaptive filter was the error signal e. This configuration of the adaptive filter works for the following reasons. If the reference signal $A_{OUT}(n)$ was perfectly filtered to match the external noise signal within the stethoscope cup $A_{IN}(n)$, the error signal between y and the

input signal $X(n)$ would result in just the breathing signal, $B(n)$.

The value of $\tilde{\mu}$ was chosen experimentally by performing adaptation on the signals of $A_{IN}(n)$ and $A_{OUT}(n)$ before adding the additive noise to the respiratory sound. The signal $A_{IN}(n)$ was used as the primary signal d and the signal $A_{OUT}(n)$ was used as the reference signal x. The output signal used was the error signal e.

The filtration process was calculated using each of the eight noise signals and varying the value of $\tilde{\mu}$ from 0 to 2 in steps of 0.01. The standard deviation of the filtered signal e was calculated. The value of $\tilde{\mu}$ corresponding to the minimum standard deviation of the output was chosen as the optimal value. The optimized value of $\tilde{\mu}$ ranged from 0.3 to 0.8 for the eight types of additive noise. The value of 0.65 was chosen to be the optimal value for the step parameter based on the mode of the optimized values for the eight types of additive noise. The length N of the filter was chosen to be 51 from experimental observations. Figure 4.4 shows an example of the original signal $A_{IN}(n)$ and the output signal $e(n)$ after filtering.

After the samples were filtered they were compared to the original sample. The output sample $e(n)$ was compared to the input sample $X(n)$. If $|X(n)|<|e(n)|$, then the value of the sample from input stream $X(n)$ was chosen as the output sample rather than the filtered signal $e(n)$. This was done to minimize contamination of the output signal with the reference signal. Contamination is clearly evident when the gain G_i is 0 and the output signal $e(n)$ is not the same as the input signal $X(n)$. For the purposes of breath detection this is a problem only when the amplitude of the contamination increases the absolute amplitude of the signal. Using this technique works well for breath detection, but because of the sudden jumps due to switching between the value of $X(n)$ as the output

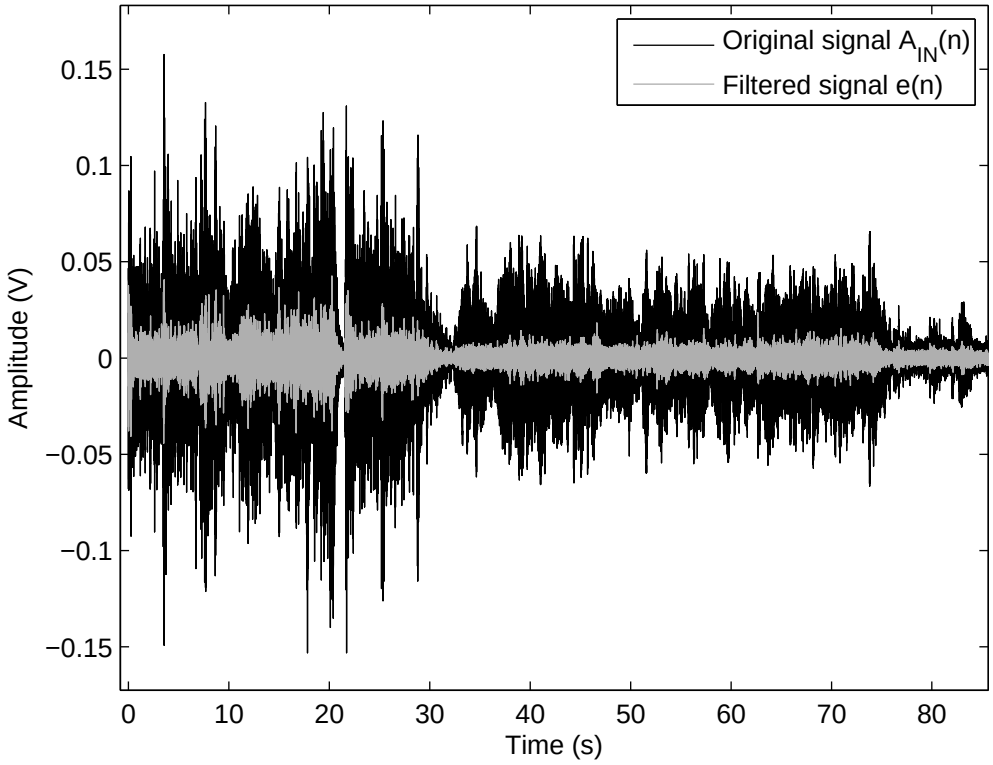


Figure 4.4 Example of noise cancellation of the adaptive filter.

to *e(n)* as the output, this signal is not optimal for classification or for human hearing.

4.2.3 Breath detection

Breath detection was performed exactly as described in Chapter 2. The standard deviation of the noise σ_{NF} and the standard deviation of the detected signal σ_S were calculated using the parameter estimation algorithm. The breath detection threshold τ was calculated from the two signals σ_{NF} and σ_S. An audio envelope was calculated for the original audio signal *X(n)*, and also for the filtered signal *e* for each audio segment, gain G_i, and additive noise signal. Sounds were detected as breaths when they had an envelope that rose above the noise threshold for 0.3 seconds or more. The breath detection calculated using the original respiratory signal *X(n)* was used as a reference for the number of sounds detected in the breathing period and apnea period before noise was added and the adaptive filter algorithm was performed. The number of sounds detected were counted during the breathing period of each segment, and the number of sounds detected were counted during the apnea section of each segment. The number of sounds detected in each period was compared to the number of sounds detected when it was not filtered, when it was filtered using an open-air reference signal, and when using a reference signal inside a stethoscope cup.

4.3 Results

The number of detected signals in each period of apnea for each gain and additive noise filter were counted and calculated as a percentage of missed detections of apnea. The percentages for unfiltered, filtered with an open-air microphone and filtered with a microphone in a stethoscope cup for a given gain and type of additive noise signal

are shown in Table 4.1. The improvement of percentage of the two filtering techniques over the unfiltered apnea detection is shown in Table 4.2.

The number of detected signals in each segment for the breathing period was also calculated. If there were no sounds detected the period was considered a period of apnea. The percent of breathing periods detected as apnea by the acoustic method are shown in Table 4.3.

4.4 Discussion

The data in Tables 4.1 to 4.3 show that not only does the cupped reference microphone adaptation improve or maintain the apnea detection for all forms of additive noise at all additive gain levels, but it also improves or maintains the breath detection for periods of breathing when compared to the unfiltered breath detection. The open-air microphone does improve or maintain the apnea detection and breath detection in most cases, but not in all cases. In the case of zero gain the open-air microphone degrades the performance of the apnea detection. This can be attributed to the dissimilarities in the additive noise signal and the open-air signal causing contamination when the adaptive filter was performed, as well as the high frequency noise picked up by the open-air microphone. The open-air microphone also degraded apnea detection for higher gains using the Gaussian additive noise signal.

Both the open-air microphone and the cupped microphone detected all periods of breathing as breathing periods as compared to the unfiltered approach. The unfiltered data did not detect sounds during the additive high gain Gaussian noise due to the model used to detect the breath sounds. The additive Gaussian had a high enough standard deviation that the Gaussian-Laplace mixture became simply a Gaussian signal and no

Table 4.1 Percent measure of missing an apnea period due to additive noise.

A No filtering performed

		Type of additive noise							
		Gaussian	Talking	Symphonic	Jazz	Jazz	Vocal	Rock	Rock
Gain G_i	0	0	0	0	0	0	0	0	0
	2	0	52.5	7.5	37.5	60	100	95	92.5
	4	0	82.5	32.5	75	92.5	100	97.5	100
	6	0	92.5	47.5	85	100	100	97.5	100
	8	0	92.5	50	90	100	100	100	100

B Filtering with reference microphone in open air

		Type of additive noise							
		Gaussian	Talking	Symphonic	Jazz	Jazz	Vocal	Rock	Rock
Gain G_i	0	17.5	2.5	20	42.5	100	90	22.5	92.5
	2	17.5	2.5	20	42.5	100	90	22.5	92.5
	4	25	2.5	20	42.5	100	90	22.5	92.5
	6	67.5	2.5	20	42.5	100	90	22.5	92.5
	8	82.5	2.5	20	42.5	100	90	22.5	92.5

C Filtering with reference microphone in stethoscope cup

		Type of additive noise							
		Gaussian	Talking	Symphonic	Jazz	Jazz	Vocal	Rock	Rock
Gain G_i	0	0	0	0	0	0	0	0	0
	2	0	2.5	0	0	0	0	0	0
	4	0	15	5	2.5	2.5	10	2.5	10
	6	0	40	27.5	17.5	30	30	2.5	25
	8	0	57.5	32.5	32.5	45	42.5	10	37.5

Table 4.2 Improvement of percentage using adaptive filter when compared to non-filtered signal.

A Improvement when filtering with reference microphone in open air

		Type of additive noise							
		Gaussian	**Talking**	**Symphonic**	**Jazz**	**Jazz**	**Vocal**	**Rock**	**Rock**
Gain G_i	**0**	-17.5	-2.5	-20	-42.5	-100	-90	-22.5	-92.5
	2	-17.5	50	-12.5	-5	-40	10	72.5	0
	4	-25	80	12.5	32.5	-7.5	10	75	7.5
	6	-67.5	90	27.5	42.5	0	10	75	7.5
	8	-82.5	90	30	47.5	0	10	77.5	7.5

B Improvement when filtering with reference microphone in stethoscope cup

		Type of additive noise							
		Gaussian	**Talking**	**Symphonic**	**Jazz**	**Jazz**	**Vocal**	**Rock**	**Rock**
Gain G_i	**0**	0	0	0	0	0	0	0	0
	2	0	50	7.5	37.5	60	100	95	92.5
	4	0	67.5	27.5	72.5	90	90	95	90
	6	0	52.5	20	67.5	70	70	95	75
	8	0	35	17.5	57.5	55	57.5	90	62.5

Table 4.3 Percent error in detecting a period of breathing as apnea.

A No filtering performed

		Type of additive noise							
		Gaussian	Talking	Symphonic	Jazz	Jazz	Vocal	Rock	Rock
Gain G_i	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0
	4	2.5	0	0	2.5	0	0	0	0
	6	20	0	0	0	0.05	0	0	0
	8	32.5	0	0	0	0.1	0	0	0

B Filtering with reference microphone in open air

		Type of additive noise							
		Gaussian	Talking	Symphonic	Jazz	Jazz	Vocal	Rock	Rock
Gain G_i	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0

C Filtering with reference microphone in stethoscope cup

		Type of additive noise							
		Gaussian	Talking	Symphonic	Jazz	Jazz	Vocal	Rock	Rock
Gain G_i	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0

breath sounds were detected. The other additive noise signals are assumed to have a non-Gaussian distribution and therefore the additive signal improves the signal-to-noise ratio although it be with corruptive noise.

Because of these results it is believed that the best option to reduce the effect of external sounds is an adaptive filter using a reference microphone in a similar stethoscope cup to the one detecting the breath signals. The reference microphone in such a situation would record ambient signals as close as possible to the additive noise on the primary microphone. The NLMS adaptive filter also allows for the filtering to be stable.

The selection of audio output where the output value was chosen as the minimum of $|X(n)|$ and $|e(n)|$ as described in Section 4.2.2 is important for automatic breath detection because it reduces the effect of contamination. This does not produce a consistent sound because the final output can jump between being a sample from the unfiltered data to being a sample of the filtered data. This implementation works well for the automated breath detection but not for human listening because it adds high frequency noise.

This work is also an improvement over what was performed by Patel [3]. The results are consistent with what Patel found, but introduces multiple variables to improve the robustness of the filter. It also uses an automatic breath detection algorithm to test the validity of the adaptive filter in this application. In addition the approach of putting the reference microphone in a stethoscope similar to the one recording the desired signal improved the results over simply using an open-air microphone.

4.6 References

[1] P.Y. Ertel, M. Lawrence, R.K. Brown, and A.M. Stern, “Stethoscope acoustics. II. Transmission and filtration patterns.,” *Circulation*, vol. 34, 1966, pp. 899-909.

[2] B. Farhang-Boroujeny, *Adaptive filters*, Wiley, 1998.

[3] S.B. Patel, T.F. Callahan, M.G. Callahan, J.T. Jones, G.P. Graber, K.S. Foster, K. Glifort, and G.R. Wodicka, “An adaptive noise reduction stethoscope for auscultation in high noise environments,” *The Journal of the Acoustical Society of America*, vol. 103, May. 1998, pp. 2483-2491.

[4] V.J. Mathews and G.L. Sicuranza, *Polynomial signal processing*, Wiley, 2000.

CHAPTER 5

SUMMARY, CONCLUSIONS AND FUTURE WORK

5.1 Summary

5.1.1 Data set

The data set was recorded from 24 volunteers in an IRB approved study. Each subject was sedated by a licensed anesthesiologist. During the sedation, data were collected from each subject using direct flow respiratory measurement with the CosmoII+ monitor, Respitrace (RIP) bands, and a microphone in a stethoscope cup placed on the trachea. The data were recorded directly to computer hard drives. During the sedation all subjects developed respiratory depression (RD) or respiratory obstruction (RO).

5.1.2 Apnea detection

Detecting periods of apnea using acoustic sounds was one goal of this study. The apnea detection was performed by creating a distribution model of the histogram of the sounds. The model was a Laplace-Gaussian mixture distribution where the Guassian signal represented the noise portion of the signal and the Laplace distribution represented the breathing signal.

The model was fitted to the audio data by using the Gaussian estimation (GE) algorithm. The GE algorithm estimated the values of the standard deviation of the Laplace and Gaussian portion of the signal. Using the standard deviation of the two

portions of the signals a threshold was calculated using the minimax algorithm. Limits were put on the minimax threshold to ensure that the threshold would not be lower than the standard deviation of the noise.

An audio envelope was calculated from the audio signal by a series of filters and signal processing manipulations. This envelope, was divided into breathing segments and periods of pause using the threshold described. The length of each pause was measured and pauses with length greater than 15 seconds were defined to be periods of apnea.

The gold standard based on the direct flow measurement identified periods of apnea. The flow measurement was used to determine periods of flow below a certain threshold for a period longer than 15 seconds. This threshold was determined from physiologic calculations of average adult airway deadspaces. These periods of apnea were used as the standard to test against the acoustic apnea detection algorithm.

5.1.2 Breath sound classification

In an attempt to determine the cause of apnea, a breath sound classifier was created. Apnea can be caused by either RD or RO. Each breath sound was classified to identify the source of apnea and additionally any periods of apnea missed by the acoustic method due to the sounds produced during partial obstruction.

Each sound was classified using a two-dimensional classifier. One dimension compared the histogram of each sound detected to 21 different distribution models. The models included a Laplace-Gaussian mixture and a Laplace-sinusoidal mixture.

The second dimension of the classifier was a test of the segment's periodicity. The periodicity of each signal was calculated by finding the ratio of the maximum of the circular autocorrelation to the standard deviation of the circular autocorrelation. The ratio

ranged from $\frac{\sqrt{2}}{2}$ being completely periodic to 0.0476 being non-periodic based on the fact that the segments were 441 samples in length. Sound distributions were classified against a Laplace-Gaussian distribution mixture with varying percentages of Laplace and Gaussian components. Segments with distributions that were closer in value to the Gaussian distribution that are nonperiodic were classified as normal breath sounds. Segments with distributions that were closer in value to the Laplace distribution that were non-periodic were classified as obstructive sounds and removed as segmented breath sounds. Segments with distributions that were closer in value to a sinusoidal distribution were classified as vocalization and considered normal. Segments with distributions that were close to either Laplace or Gaussian distributions that were periodic were either vocalization or light snoring and considered normal breaths. The segmentation markers of sounds that were classified as obstructive were removed from the data set, and periods of apnea were recalculated given the removed breath detection markers. In addition the periods of apnea found using the detection algorithm described were classified as either RO or RD based on the classification of the sounds preceding the apnea. If any sound in the prior five seconds or within the period of apnea were classified as obstruction, the period of apnea was classified as RO. The period of apnea was classified as RD otherwise.

5.1.3 Adaptive filtering of sounds from the stethoscope cup

One of the causes of periods of apnea being missed by the acoustic algorithm was ambient sounds such as talking being detected through the stethoscope cup. An adaptive filter was constructed to test the validity of using noise cancellation to remove sounds

created from external sources.

The data set described was not recorded using a reference microphone that would be needed to perform the noise cancellation. For this reason an alternative method to test the validity of using a noise-cancellation technique was devised. Forty of the periods of apnea detected by the flow meter and confirmed by the acoustic method were selected. Each period of apnea was accompanied by the period of breathing prior to the apnea. Additive noise was used to simulate external sounds. The additive noise was added to each breathing and apnea period at five different amplitudes. The different sounds used as additive noise were white noise, talking, vocal music, rock music and jazz music. The additive noise was recorded with the primary microphone in a stethoscope cup attached to a phantom material and the reference microphone in proximity. The standard deviation of the noise floor of the breathing data was calculated. The additive noise was normalized by its standard deviation and added to the breathing sound by proportions of 0, 2, 4, 6 and 8 times the standard deviation of the noise floor of the breathing sounds. Each of the 40 segments was combined with the additive noise at each gain level. The adaptive filter was then applied to produce a filtered output.

The breath detection algorithm was applied to all 40 segments using the 5 different gains and the different types of additive noise for both the unfiltered and filtered output of the adaptive filter. If sounds were detected during the period of apnea, the apnea was counted as a false negative, otherwise it was counted as a true positive. If there were sounds detected during breathing segments the period was counted as a true negative, otherwise it was counted as a false positive. The performance of the filtered output was compared to the unfiltered signal. The filtered signal did not perform better

than the unfiltered data in some circumstances. For this reason the additive noise was re-recorded, placing the reference microphone in a second stethoscope cup to attenuate the reference signal similar to the additive signal. Placing the reference microphone inside a stethoscope cup coupled to the back of the primary stethoscope cup improved the performance of the adaptive filter so that it performed the same or better than the unfiltered data in all cases.

5.2 Summary of Observations

5.2.1 Apnea detection

The model used to perform the sound detection for apnea detection proved to be accurate. It worked well enough that the methods described could be used for applications such as a voice activity detector (VAD). A VAD is a detection algorithm used to detect voice so that only voice is transmitted on a communication line and not silence. The histogram model worked mainly because of the known distribution of the noise. Although the Laplace distribution may not be a perfect fit for the breathing signal, the measure of the standard deviation of this signal is enough to differentiate the noise signal from the breathing signal.

The minimax algorithm used to determine the noise threshold was not novel but the application of this was useful for varying noise levels and breath signal amplitudes. The wide range of amplitudes of the detected signal made the threshold more specific to each type of audio data.

5.2.2 Classification

Personal observation from the classification of sounds showed that the histogram classifier performed very well on normal breath sounds. Sounds that did not display a Gaussian distribution were not as reliably classified. It was observed that if a breath sound distribution was classified as Gaussian with low periodicity, it was reliably classified as normal breath sound. The only other classification that was extremely reliable was the sinusoidal distribution reliably predicting vocalization.

5.2.3 Adaptive filtering

Using the adaptive filter to reduce ambient noise proved useful in the testing method described. Whether it will work in a real-world application has yet to be tested. The most important finding of the adaptive filter was the importance of the placement of the reference microphone. In this case the best position of the reference microphone was within a second stethoscope. This attenuated the ambient noise to the reference microphone similar to the ambient noise received at the primary microphone.

5.3 Concurrent Work

While this work was underway I learned of similar work being performed by Andromed Corporation. Andromed used a piezo-electric film as a transducer on the trachea. Adnromed was subsequently purchased by Masimo Corporation. Masimo has now produced an FDA-approved device that uses the piezo-electric film transducer as a respiratory monitor [1]. This device is used in conjunction with a Rainbow SET® pulse-oximeter. This device shows the need of an acoustic respiratory monitor. The Masimo

device has also established FDA approval on an acoustic respiratory monitor that allows for a precedent in approving future acoustic respiratory monitors.

The Masimo product uses a piezo-electric film rather than a conventional microphone inside a sthethoscope cup. This makes it subject to lower frequency vibrations. From my observations using piezo-electric films the signal-to-noise ratio is not as high using a piezo-electric film as compared to a conventional microphone inside a stethoscope cup. From the observations of colleagues at demonstrations of the Masimo product it was noted that it did not work well in an acoustically noisy environment. It does have the advantage of being easily applied and there is no need for a thick metal stethoscope. Unfortunately it does not appear to have a noise-cancellation feature. The Masimo device also might rely heavily on the pulse-oximeter to detect apnea. The device also has no mention of its ability to differentiate RO from RD.

5.4 Suggestions for Future Work

5.4.1 Create a synchronized data collection system

One of the problems described in the current work was the lack of automatic synchronization between the flow data and the acoustic signals. The solution to this is to devise a way to record the audio signal with a time stamp from the computer recording the flow data, or to have all of the data recorded on the same computer and have a time stamp attached to all samples taken.

5.4.2 Create a standard for apnea classification

Because there is no simple standard or measurement tool that determines the source of apnea, the construction of a standard would be a useful tool for future research. Since the current standard of care for identifying obstruction is polysomnography, this would have to be performed by a licensed professional on either the data recorded or on a new data set. An automated classification standard would then have to be constructed and tested against the physician classification.

I believe that using both the flow meter data and the RIP data, a reliable automated apnea detection and classification algorithm can be constructed. This could be constructed using the existing data set as the training set and then testing it on a secondary data set recorded in a patient care center or from a paid subject study. When the flow meter shows no valid flow and the calibrated RIP volume shows no valid volume change then the patient is in RD. When the flow meter shows no valid flow and the calibrated RIP shows volumes less than approximately 180 mL but are detected as breaths by the RIP, these periods would be considered RO. Although this algorithm seems simple, there are technical issues that arise from both the RIP monitor and the flow meter that require special attention. Some of these were addressed in Chapter 3.

5.4.3 Manual audio classification testing

The primary difficulty in classification of breath sounds was the wide variation of sounds that can be present for any given state. It was concluded that RO and RD cannot be reliably discerned by sounds alone. This can be confirmed by classifying periods of apnea based on the sounds preceding them by a trained researcher and comparing them to

the already classified periods of apnea described in Chapter 3. This method is subjective and the outcome is not scientifically based, but it would support claims made about the methods of recording the data for this project.

5.4.4 Use of tomography and acoustics to differentiate RO and RD

From my observation of the data from the RIP, flow, and acoustic monitors used, I believe that given the data from any two of the three monitors, apnea and its source can be identified reliably. It has already been described how the combination of the RIP and the flow measurement can be used to identify the difference between RO and RD. I believe that using the acoustic signal and the flow meter, RO can be identified when the breath sounds are loud and the flow shows little to no flow. When the flow signal and breath sounds are correlated prior to a period of apnea the period would be RD. Similarly RO would be identified using the RIP and acoustic measurement when the RIP signal showed breaths but the acoustic flow detected no breaths. RD would be detected when both the RIP and acoustic measurement showed no flow.

Because RIP bands are uncomfortable and difficult to place, an alternative method to measure chest and abdomen excursions has been hypothesized. Griffith et al. [2] performed electrical impedance tomography on the pharynx using an array of sixteen hydrogel electrodes. The tomography was able to image the pharynx as subjects swallowed. This is related to performing chest impedance plethysmography. I believe that using two or three electrodes on the trachea near the stethoscope cup, changes in the trachea due to changing pressures can be measured and would be similar in nature to the RIP measurement described. In addition, during obstruction the tracheal diameter would

change even more due to the increase in pressure. This could be tested further by inserting a tracheal pressure probe past the pharynx while comparing the pressures exhibited to the change in tracheal electrical impedance.

5.4.5 Further testing of adaptive filter to remove ambient sounds

The adaptive filter was tested using additive noise that was recorded separately from the breathing data. There are many acoustic differences between this method and a method that would record the additive noise while recording breathing sounds. Two methods are suggested to perform further tests.

The first method would be to create a simulated trachea that is capable of creating breath sounds through a phantom material. The sounds recorded in the stethoscope cup and the reference microphone would then be used to test the adaptive filter against additive noise created by external sources. The second method would be to get an IRB approved study and have subjects perform breath holds as recorded noise is played in the room.

5.4.6 Create algorithm to improve adaptive filter

One problem that occurred during the adaptive filtering research was that the reference microphone could corrupt the output of the adaptive filter if the reference signal was uncorrelated with the primary input and if the frequency was high enough. In order to reduce this it is proposed that the adaptive filter is only used when needed. The filter is only needed when there is a signal detected above the noise floor on both the reference microphone and the primary microphone. In order to do this the breath detection

algorithm needs to be performed on the reference input and only periods that are detected on both inputs will be filtered using the adaptive filter. This will not only reduce the amount of corruption from the reference microphone, but it will reduce the computing power needed to perform the algorithms described.

5.4.7 Testing of different stethoscope cups

The use of a thick metal stethoscope cup helped in the attenuation of the ambient sounds, but the heavy stethoscope is bulky. Two stethoscope cups attached to provide for the adaptive filtration is long and heavy and will likely not be held by a single double-stick disk. The weight and awkwardness of the stacked cups is not ideal in real-world applications. In order to alleviate this, smaller and thinner stethoscope cups made of both metal and plastic need to be tested using the adaptive filter. This will allow for a more practical apparatus that can be held to the trachea using a double stick-disk.

5.4.8 Recording data from patient volunteers

The best test of the apnea monitor is the use of a version of the software on patient volunteers undergoing sedation procedures. In order to do this IRB approval would have to be obtained and the data recording process would have to be perfected. Use of this technology in a patient application could also be a step towards FDA approval for an acoustic respiratory monitor. The data could be performed on patients undergoing sedative procedures such as a colonoscopies.

5.4.9 Implementation of algorithms into a integrated device

Up until this point the data have been recorded and processed offline using modern computers and programming software such as MATLAB. The ultimate goal of this project would be to implement the algorithms that have been described as well as the algorithms that will be developed into a microcontroller and process the data in real time. This will be a final step toward getting the device ready to be a marketable medical device.

5.5 References

[1] “Masimo - Acoustic Respiration Rate (RRa).” Available http://www.masimo.com/rra/index.htm. [Accessed June 15, 2010].

[2] P. Liu, H. Griffiths, C.M. Wiles, K.M. Nathadwarawala, and W. Stewart, “Measurement of pharyngeal transit time by electrical impedance tomography.,” *Clin Phys Physiol Meas*, vol. 13 Suppl A, 1992, pp. 197-200.

APPENDIX A

HISTOGRAM OF A MODULATED GAUSSIAN SIGNAL

Mathematically the distribution of a Gaussian signal modulated by a sinusoid can be calculated as shown in Cain et al. [1]. They modulated five common random signals (sinusoid, uniform distribution, Gaussian distribution, Rayleigh distribution, and Laplace distribution) by a sinusoid. They started by giving the probability density function (PDF) of the sinusoidal process:

$$f_X(x) = \frac{1}{\pi\sqrt{1-x^2}}\, rect\left(\frac{x}{2}\right) \tag{A.1}$$

where x is the random variable. The common method of finding the distribution of the multiplication of two random variables with known distributions is performed using the equation [1]:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_X\left(\frac{z}{y}\right) f_Y(y)\, dy \tag{A.2}$$

where $f_X\left(\frac{z}{y}\right)$ is the distribution of the first random variable with a modification to the input variable, and $f_Y(y)$ is the distribution of the second random variable. Because one of the distributions is the Gaussian distribution and contains the term of e^{x^2}, the integral cannot be solved directly. Instead of solving this directly, Cain et al. [1] used the Fourier transform of the sinusoidal distribution to find the resulting distribution. First, they perform the Fourier transform of equation A.1 to get

$$\Phi_X(\beta) = J_0(2\pi\beta) \tag{A.3}$$

where J_0 is the zeroth-order Bessel function of the first kind, β is the frequency, and Φ_X represents the Fourier transform of equation A.1. They then used a lesser known relationship to find the Fourier transform of the modulated signal, which has the equation

$$\Phi_Z(\beta) = \int_{-\infty}^{\infty} \Phi_X(\nu\beta)\, f_Y(\nu)\, d\nu \tag{A.4}$$

where $f_Y(\nu)$ is the distribution of the second random variable, in this case the Gaussian distribution. They then substituted equation A.3 into equation A.4 directly and substituted the Gaussian distribution for $f_Y(\nu)$. The result of this was

$$\Phi_Z(\beta) = e^{-\pi^2\beta^2} I_0(\pi^2\beta^2) \tag{A.5}$$

where I_0 is the zeroth-order modified Bessel function of the first kind. Using the Campbell and Foster Fourier transform and adding a scaling factor, the modulation of a Gaussian signal was shown to be [1]

$$f_Z(z) = \frac{1}{\pi\sqrt{(2\pi)}}[e^{\frac{-z^2}{4}} K_0(\frac{z^2}{4})], \tag{A.5}$$

where z is the random variable, and K_0 is the order zero modified Bessel function of the second kind. They go further and compare a simulation to the distribution of the data. The simulation was created by taking the dot product of a vector of a Gaussian distributed signal and a vector of a full-wave sinusoid of the same length, then finding the normalized histogram of this signal. A comparison of this model to simulated data is shown in Figure A.1.

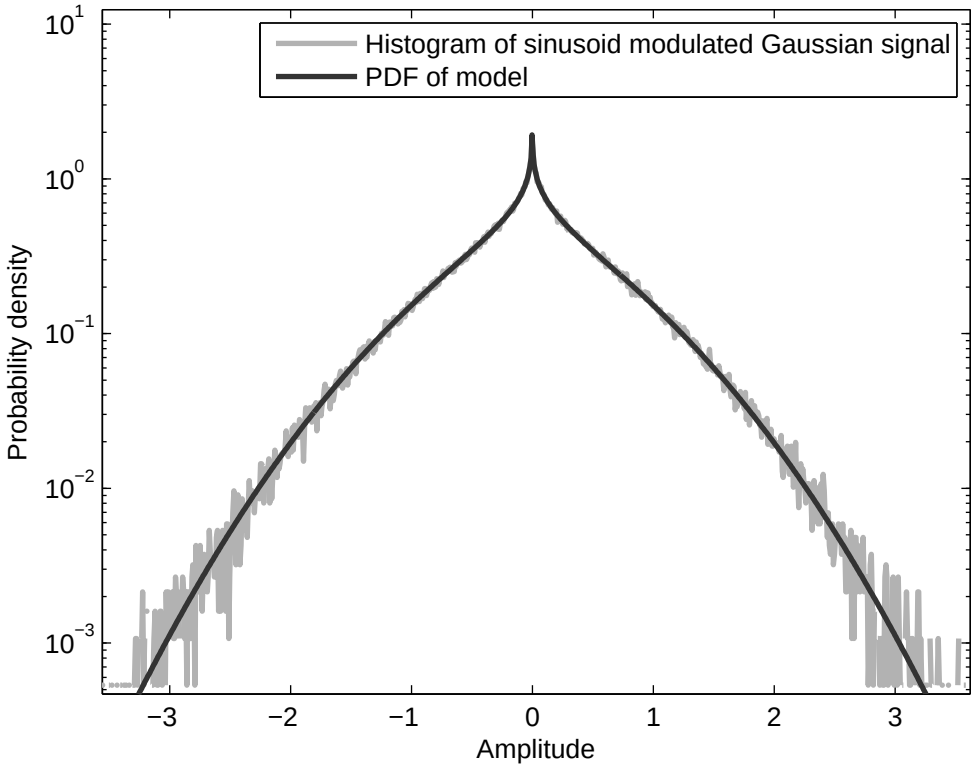


Figure A.1 Normalized histogram of a sinusoid modulated Gaussian random signal overlaid with the PDF for the signal.

References

[1] K. G.D. Cain, K.V. Lever and A.Yardim, “Probability density functions of amplitude-modulated random signals,” *IEEE Electronics Letters*, vol. 34, 1998, pp. 1560-1561.

APPENDIX B

SNORE SLAP DISTRIBUTION CALCULATIONS

In an attempt to understand what caused the distribution displayed by a snoring sound, a snore model was created. This model was a simplistic model of an individual snore slap and has the following equation:

$$y_{db}(t) = c e^{\frac{-t}{d}} \cdot a \sin(2\pi f_n t + \phi_n) \tag{B.1}$$

where t is time, c is the amplitude of the individual slap, d is the decay of the slap, f_n is the frequency of the slap, a is the amplitude of the sinusoid, and Φ_n is a random phase of the frequency of that slap. The random phase is not present in the model presented in Chapter 3. In a physical setting, a phase of Φ being anything but zero would not be possible because it would require an instantaneous jump from zero to the first value of the signal described by equation B.1. This term was added for the sake of calculating the distribution of the slap. It ensures that the distribution model for a sinusoid can be used for the models described. Although this is different from the actual histogram produced by the sound, the discrepancy is believed to be small when comparing the distributions. The phase is a single term added for each individual snore slap and is uniformly distributed in the range of $0<\Phi<\pi$. Using this model, the distribution for a series of snore slaps with uniform amplitude and length can be calculated. The distribution is calculated by finding the distribution of the multiplication of two random variables. The first variable

is the exponential decay y_{dbe}:

$$y_{dbe}(t) = c \times e^{\frac{-t}{d}} \tag{B.2}$$

where t is time, c is the initial amplitude of the exponential decay, and d is the decay time constant of the signal. The distribution of equation B.2 can be found by the Jacobian transformation to find the PDF of a function. The Jacobian transformation is performed by taking the absolute value of the derivative of the inverse of the function. Solving the inverse of equation B.2 and substituting y for $y_{dbe}(x)$, yields

$$t = -d \times ln(y/c), \tag{B.3}$$

and calculating the derivative of this function yields

$$t' = cd/y. \tag{B.4}$$

In addition, this equation is unbounded as y approaches zero. To avoid an unbounded distribution, the length of the slap was limited to $0 \le t \le l$. The limitation on t in equation B.2 means that the amplitude of y in equation B.4 is limited to $ce^{-l/d} \le y \le c$. Thus, the distribution of an exponential decay after normalization is

$$f_{ZE}(z) = \begin{cases} \dfrac{d}{l\,z}, & for\ c\,e^{-l/d} \le z \le c \\ 0, & otherwise \end{cases} \tag{B.5}$$

The second part of the model is the modulating sinusoid y_{dbs}:

$$y_{dbs}(t) = a\ sin(2\pi f_n t + \Phi_n). \tag{B.6}$$

where t is time, a is the amplitude of the sinusoid, f_n is the frequency of the modulating sinusoid and Φ_n is a random phase of the modulating sinusoid. The distribution of equation B.6 is given in [1] as:

$$f_{YS}(y)=\begin{cases}\dfrac{1}{\pi\sqrt{(a^2-y^2)}} & for -a<y<a \\ 0 & otherwise\end{cases} \tag{B.7}$$

Now that the distributions of the two parts of the snore slap model are known they will be used to find the distribution of the snore slap. The distribution of the multiplication of two variables is well known and using the variables of *y* and *z* as examples is given in [1] as

$$f_X(x)=\int_{-\infty}^{\infty}\frac{1}{|y|}f_Z(\frac{x}{y})f_Y(y)dy. \tag{B.8}$$

However, this is only valid for distributions that extend across all real values. Glen et al. [2] describe a method to calculate the distribution of the multiplication of two bounded random variables. Glen et al. state that, for the two distributions, the first distribution, *f(x)*, with limits of r and s where $0<r<s<\infty$, and the second distribution, *g(x)*, with limits of *t* and *u* where $0<t<u<\infty$, the distribution of the multiplication of the two distributions can be found. In this case the first distribution is the sinusoid distribution and does not meet these criteria. In order to fit the criteria, the symmetry of the sinusoid distribution was used. The random phase of each snore slap ensures the symmetry of the distribution and the multiplication of the exponential decay distribution ensures that the end distribution will be symmetrical. For this reason, the distribution of equation B.7 was modified to the positive half of the distribution with equation:

$$f_{YSH}(y)=\begin{cases}\dfrac{2}{\pi\sqrt{(a^2-y^2)}} & for\ 0\le y<a \\ 0 & otherwise\end{cases}. \tag{B.9}$$

Although the lower limit of this distribution is now zero, it can be assumed to be very close to zero to satisfy the criteria set by Glen. Once the distribution of the positive half

of the model is found, the negative half can be found through symmetry. Glen et al. [2] then give three cases for the possible distributions computed in a piece-wise style. The first is if *ru*<*st*, the second is if *ru*=*st*, and the third is if *ru*>*st*. In this case, since the lower limit of the sinusoid distribution denoted by *r* is very close to zero when compared to *u*, *s*, and *t*, the first case is used for this calculation. The distribution is calculated by:

$$h(v)=\begin{cases}\int_{r}^{v/t} g\left(\frac{v}{x}\right) f(x)\frac{1}{x}dx, & for\ rs<v<ru\\ \int_{v/u}^{v/t} g\left(\frac{v}{x}\right) f(x)\frac{1}{x}dx, & for\ ru<v<st.\\ \int_{v/u}^{s} g\left(\frac{v}{x}\right) f(x)\frac{1}{x}dx, & for\ st<v<su\end{cases} \tag{B.10}$$

where *h(v)* is the resultant distribution and *v* is the resultant random variable. Because the value for *r* is assumed to be very close to zero the first case spans a range that is very close to zero and has been ignored thus leaving two cases. In fact *r* can be set to zero with no adverse effects to the distribution when compared to the simulation as will be shown. Applying equations B.5 as *g(x)* and B.9 as *f(x)* to equation B.10 and using arbitrary limits results in

$$h(v)=\int_{n}^{m}\frac{x}{v}\frac{d}{l}\frac{2}{\pi\sqrt{(a^2-y^2)}}\frac{1}{x}dx \tag{B.11}$$

where *m* and *n* are arbitrary limits for the integral. This equation then simplifies to

$$h(v)=\frac{2\mathrm{d}}{\pi lv}\int_{n}^{m}\frac{1}{\sqrt{(a^2-y^2)}}dx. \tag{B.12}$$

The general form of the integral can be solved as

$$h(v)=\frac{2\mathrm{d}}{\pi lv}\left(\sin^{-1}\left(\frac{m}{a}\right)-\sin^{-1}\left(\frac{n}{a}\right)\right). \tag{B.13}$$

Applying the limits of equation B.10 two cases are shown:

$$h(v)=\begin{cases}\dfrac{2\mathrm{d}}{\pi lv}\left(\sin^{-1}\left(\dfrac{v}{ace^{-l/d}}\right)-\sin^{-1}\left(\dfrac{v}{ac}\right)\right), & for\ 0<v<ace^{-l/d}\\ \dfrac{2\mathrm{d}}{\pi lv}\left(\dfrac{\pi}{2}-\sin^{-1}\left(\dfrac{v}{ac}\right)\right), & for\ ace^{-l/d}<v<ac\\ 0, & otherwise\end{cases} \tag{B.14}$$

where *a, c, d,* and *l* have already been defined. Since this is a symmetrical distribution, the model for the values from $-ac<v<ac$ is:

$$h(v)=\begin{cases}\dfrac{d}{\pi lv}\left(\dfrac{-\pi}{2}-\sin^{-1}\left(\dfrac{v}{ac}\right)\right), & for\ \ -ac<v<-ace^{-l/d}\\ \dfrac{d}{\pi lv}\left(\sin^{-1}\left(\dfrac{v}{ace^{-l/d}}\right)-\sin^{-1}\left(\dfrac{v}{ac}\right)\right), & for\ \ -ace^{-l/d}<v<ace^{-l/d}\\ \dfrac{d}{\pi lv}\left(\dfrac{\pi}{2}-\sin^{-1}\left(\dfrac{v}{ac}\right)\right), & for\ \ ace^{-l/d}<v<ac\\ 0, & otherwise\end{cases} \tag{B.15}$$

Using the fact that the real value of the function of $sin^{-1}()$ for values greater than 1 is $\pi/2$ and less than -1 is $-\pi/2$ the distribution can be simplified to one single case:

$$h(v)=\begin{cases}\dfrac{d}{\pi lv}\left(real\left(\sin^{-1}\left(\dfrac{v}{ace^{-l/d}}\right)\right)-\sin^{-1}\left(\dfrac{v}{ac}\right)\right), & for\ \ -ac<v<ac\\ 0, & otherwise\end{cases} \tag{B.16}$$

where *real()* denotes the real values of the function.

A simulation using a vector of 100 simulated snore slap signals and *a=0.3, c=0.5, d=0.1,* and *l=0.1* at a sample rate of 22050 Hz was created and the normalized histogram was calculated from that signal. The values were applied to the distribution model shown in equation B.11 and compared to the histogram of the simulation as shown in Figure B.1a and B.1b. It can be seen that the distribution matches the simulation with very little error. Similarly, the distributions and normalized simulation histograms for the values of *a=0.3, c=0.5, d=1,* and *l=0.1,* and *a=0.3, c=0.5, d=0.015,* and *l=0.1* are shown in

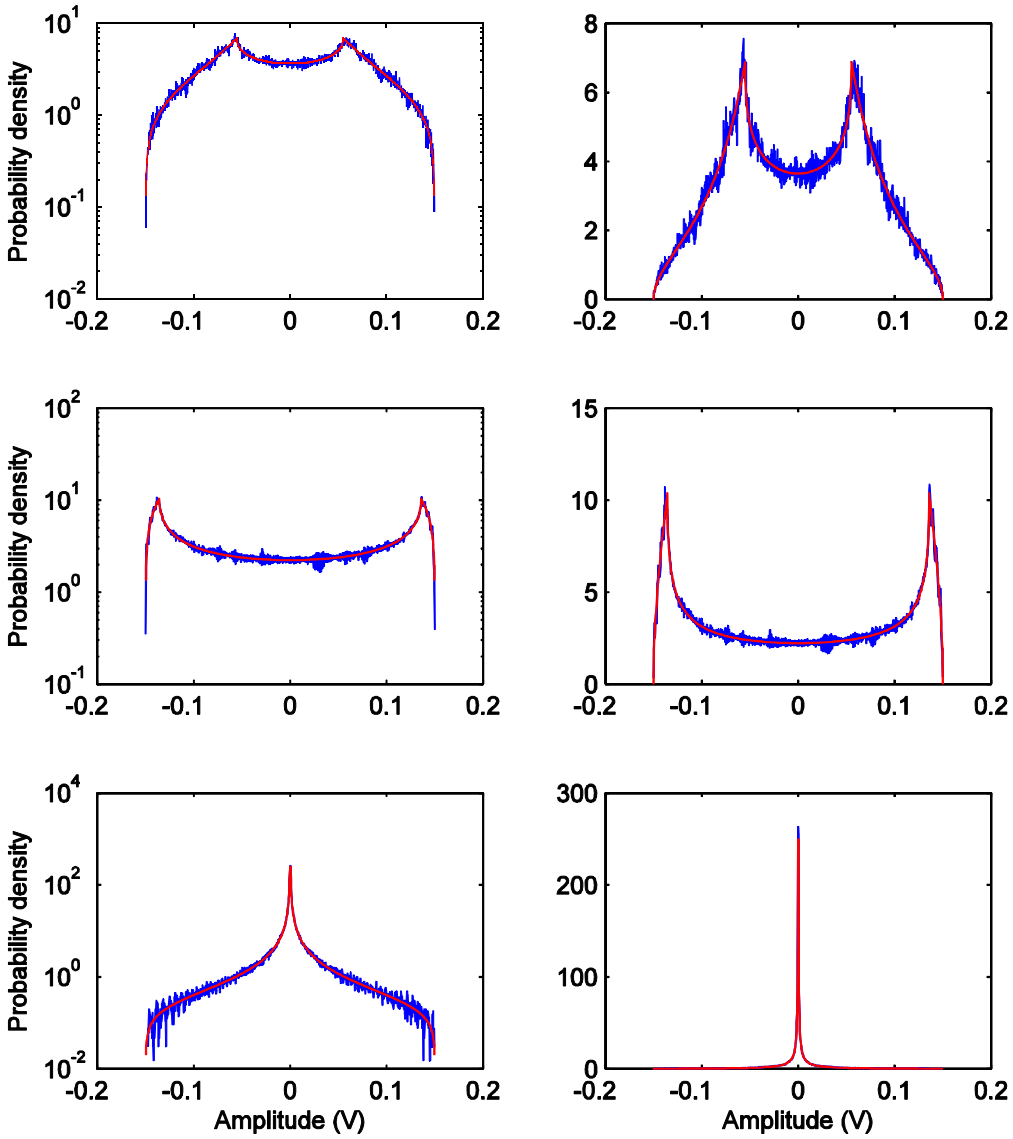


Figure B.1 Normalized histograms of a repeated slap sound with the PDF model for (a) semilog plot with *a=0.3, c=0.5, d=0.1,* and *l=0.1* (b) linear plot with *a=0.3, c=0.5, d=0.1,* and *l=0.1* (c) semilog plot with *a=0.3, c=0.5, d=1,* and *l=0.1* (d) linear plot with *a=0.3, c=0.5, d=1,* and *l=0.1* (e) semilog plot with *a=0.3, c=0.5, d=0.015,* and *l=0.1* (f) linear plot with *a=0.3, c=0.5, d=0.015,* and *l=0.1.*

Figures B.1c through B.1f to show the extreme values for the ratio of *l* to *d.*

References

[1] K. G.D. Cain and A.Yardim, “Probability density functions of amplitude-modulated random signals,” *IEEE Electronics Letters*, vol. 34, 1998, pp. 1560-1561.

[2] A.G. Glen, L.M. Leemis, and J.H. Drew, “Computing the Distribution of the Product of Two Continuous Random Variables,” 2003.