Neural Network Signal Processing Techniques for Adventitious Lung Sound Classification

By

Kevin Edward Forkheim

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of

Master of Science
in
Computer Science

Department of Computer Science
University of Manitoba
Winnipeg, Manitoba, Canada

ISBN 0-612-13124-6

Name ______________________

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Computer Science — 0984 — U·M·I

SUBJECT TERM — SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
- Architecture ... 0729
- Art History ... 0377
- Cinema ... 0900
- Dance ... 0378
- Fine Arts ... 0357
- Information Science ... 0723
- Journalism ... 0391
- Library Science ... 0399
- Mass Communications ... 0708
- Music ... 0413
- Speech Communication ... 0459
- Theater ... 0465

EDUCATION
- General ... 0515
- Administration ... 0514
- Adult and Continuing ... 0516
- Agricultural ... 0517
- Art ... 0273
- Bilingual and Multicultural ... 0282
- Business ... 0688
- Community College ... 0275
- Curriculum and Instruction ... 0727
- Early Childhood ... 0518
- Elementary ... 0524
- Finance ... 0277
- Guidance and Counseling ... 0519
- Health ... 0680
- Higher ... 0745
- History of ... 0520
- Home Economics ... 0278
- Industrial ... 0521
- Language and Literature ... 0279
- Mathematics ... 0280
- Music ... 0522
- Philosophy of ... 0998
- Physical ... 0523
- Psychology ... 0525
- Reading ... 0535
- Religious ... 0527
- Sciences ... 0714
- Secondary ... 0533
- Social Sciences ... 0534
- Sociology of ... 0340
- Special ... 0529
- Teacher Training ... 0530
- Technology ... 0710
- Tests and Measurements ... 0288
- Vocational ... 0747

LANGUAGE, LITERATURE AND LINGUISTICS
- Language
  - General ... 0679
  - Ancient ... 0289
  - Linguistics ... 0290
  - Modern ... 0291
- Literature
  - General ... 0401
  - Classical ... 0294
  - Comparative ... 0295
  - Medieval ... 0297
  - Modern ... 0298
  - African ... 0316
  - American ... 0591
  - Asian ... 0305
  - Canadian (English) ... 0352
  - Canadian (French) ... 0355
  - English ... 0593
  - Germanic ... 0311
  - Latin American ... 0312
  - Middle Eastern ... 0315
  - Romance ... 0313
  - Slavic and East European ... 0314

PHILOSOPHY, RELIGION AND THEOLOGY
- Philosophy ... 0422
- Religion
  - General ... 0318
  - Biblical Studies ... 0321
  - Clergy ... 0319
  - History of ... 0320
  - Philosophy of ... 0322
- Theology ... 0469

SOCIAL SCIENCES
- American Studies ... 0323
- Anthropology
  - Archaeology ... 0324
  - Cultural ... 0326
  - Physical ... 0327
- Business Administration
  - General ... 0310
  - Accounting ... 0272
  - Banking ... 0770
  - Management ... 0454
  - Marketing ... 0338
- Canadian Studies ... 0385
- Economics
  - General ... 0501
  - Agricultural ... 0503
  - Commerce-Business ... 0505
  - Finance ... 0508
  - History ... 0509
  - Labor ... 0510
  - Theory ... 0511
- Folklore ... 0358
- Geography ... 0366
- Gerontology ... 0351
- History
  - General ... 0578
  - Ancient ... 0579
  - Medieval ... 0581
  - Modern ... 0582
  - Black ... 0328
  - African ... 0331
  - Asia, Australia and Oceania 0332
  - Canadian ... 0334
  - European ... 0335
  - Latin American ... 0336
  - Middle Eastern ... 0333
  - United States ... 0337
- History of Science ... 0585
- Law ... 0398
- Political Science
  - General ... 0615
  - International Law and Relations ... 0616
  - Public Administration ... 0617
- Recreation ... 0814
- Social Work ... 0452
- Sociology
  - General ... 0626
  - Criminology and Penology ... 0627
  - Demography ... 0938
  - Ethnic and Racial Studies ... 0631
  - Individual and Family Studies ... 0628
  - Industrial and Labor Relations ... 0629
  - Public and Social Welfare ... 0630
  - Social Structure and Development ... 0700
  - Theory and Methods ... 0344
- Transportation ... 0709
- Urban and Regional Planning ... 0999
- Women's Studies ... 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
- Agriculture
  - General ... 0473
  - Agronomy ... 0285
  - Animal Culture and Nutrition ... 0475
  - Animal Pathology ... 0476
  - Food Science and Technology ... 0359
  - Forestry and Wildlife ... 0478
  - Plant Culture ... 0479
  - Plant Pathology ... 0480
  - Plant Physiology ... 0817
  - Range Management ... 0777
  - Wood Technology ... 0746
- Biology
  - General ... 0306
  - Anatomy ... 0287
  - Biostatistics ... 0308
  - Botany ... 0309
  - Cell ... 0379
  - Ecology ... 0329
  - Entomology ... 0353
  - Genetics ... 0369
  - Limnology ... 0793
  - Microbiology ... 0410
  - Molecular ... 0307
  - Neuroscience ... 0317
  - Oceanography ... 0416
  - Physiology ... 0433
  - Radiation ... 0821
  - Veterinary Science ... 0778
  - Zoology ... 0472
- Biophysics
  - General ... 0786
  - Medical ... 0760

EARTH SCIENCES
- Biogeochemistry ... 0425
- Geochemistry ... 0996
- Geodesy ... 0370
- Geology ... 0372
- Geophysics ... 0373
- Hydrology ... 0388
- Mineralogy ... 0411
- Paleobotany ... 0345
- Paleoecology ... 0426
- Paleontology ... 0418
- Paleozoology ... 0985
- Palynology ... 0427
- Physical Geography ... 0368
- Physical Oceanography ... 0415

HEALTH AND ENVIRONMENTAL SCIENCES
- Environmental Sciences ... 0768
- Health Sciences
  - General ... 0566
  - Audiology ... 0300
  - Chemotherapy ... 0992
  - Dentistry ... 0567
  - Education ... 0350
  - Hospital Management ... 0769
  - Human Development ... 0758
  - Immunology ... 0982
  - Medicine and Surgery ... 0564
  - Mental Health ... 0347
  - Nursing ... 0569
  - Nutrition ... 0570
  - Obstetrics and Gynecology ... 0380
  - Occupational Health and Therapy ... 0354
  - Ophthalmology ... 0381
  - Pathology ... 0571
  - Pharmacology ... 0419
  - Pharmacy ... 0572
  - Physical Therapy ... 0382
  - Public Health ... 0573
  - Radiology ... 0574
  - Recreation ... 0575
  - Speech Pathology ... 0460
  - Toxicology ... 0383
- Home Economics ... 0386

PHYSICAL SCIENCES

Pure Sciences
- Chemistry
  - General ... 0485
  - Agricultural ... 0749
  - Analytical ... 0486
  - Biochemistry ... 0487
  - Inorganic ... 0488
  - Nuclear ... 0738
  - Organic ... 0490
  - Pharmaceutical ... 0491
  - Physical ... 0494
  - Polymer ... 0495
  - Radiation ... 0754
- Mathematics ... 0405
- Physics
  - General ... 0605
  - Acoustics ... 0986
  - Astronomy and Astrophysics ... 0606
  - Atmospheric Science ... 0608
  - Atomic ... 0748
  - Electronics and Electricity ... 0607
  - Elementary Particles and High Energy ... 0798
  - Fluid and Plasma ... 0759
  - Molecular ... 0609
  - Nuclear ... 0610
  - Optics ... 0752
  - Radiation ... 0756
  - Solid State ... 0611
- Statistics ... 0463

Applied Sciences
- Applied Mechanics ... 0346
- Computer Science ... 0984
- Engineering
  - General ... 0537
  - Aerospace ... 0538
  - Agricultural ... 0539
  - Automotive ... 0540
  - Biomedical ... 0541
  - Chemical ... 0542
  - Civil ... 0543
  - Electronics and Electrical ... 0544
  - Heat and Thermodynamics ... 0348
  - Hydraulic ... 0545
  - Industrial ... 0546
  - Marine ... 0547
  - Materials Science ... 0794
  - Mechanical ... 0548
  - Metallurgy ... 0743
  - Mining ... 0551
  - Nuclear ... 0552
  - Packaging ... 0549
  - Petroleum ... 0765
  - Sanitary and Municipal ... 0554
  - System Science ... 0790
- Geotechnology ... 0428
- Operations Research ... 0796
- Plastics Technology ... 0795
- Textile Technology ... 0994

PSYCHOLOGY
- General ... 0621
- Behavioral ... 0384
- Clinical ... 0622
- Developmental ... 0620
- Experimental ... 0623
- Industrial ... 0624
- Personality ... 0625
- Physiological ... 0989
- Psychobiology ... 0349
- Psychometrics ... 0632
- Social ... 0451

Nom ____

Dissertation Abstracts International est organisé en catégories de sujets. Veuillez s.v.p. choisir le sujet qui décrit le mieux votre thèse et inscrivez le code numérique approprié dans l'espace réservé ci-dessous.

______________________________________ SUJET

[ ][ ][ ][ ] CODE DE SUJET

U·M·I

Catégories par sujets

HUMANITÉS ET SCIENCES SOCIALES

COMMUNICATIONS ET LES ARTS
- Architecture 0729
- Beaux-arts 0357
- Bibliothéconomie 0399
- Cinéma 0900
- Communication verbale 0459
- Communications 0708
- Danse 0378
- Histoire de l'art 0377
- Journalisme 0391
- Musique 0413
- Sciences de l'information 0723
- Théâtre 0465

ÉDUCATION
- Généralités 515
- Administration 0514
- Art 0273
- Collèges communautaires 0275
- Commerce 0688
- Économie domestique 0278
- Éducation permanente 0516
- Éducation préscolaire 0518
- Éducation sanitaire 0680
- Enseignement agricole 0517
- Enseignement bilingue et multiculturel 0282
- Enseignement industriel 0521
- Enseignement primaire. 0524
- Enseignement professionnel 0747
- Enseignement religieux 0527
- Enseignement secondaire 0533
- Enseignement spécial 0529
- Enseignement supérieur 0745
- Évaluation 0288
- Finances 0277
- Formation des enseignants 0530
- Histoire de l'éducation 0520
- Langues et littérature 0279
- Lecture 0535
- Mathématiques 0280
- Musique 0522
- Orientation et consultation 0519
- Philosophie de l'éducation 0998
- Physique 0523
- Programmes d'études et enseignement 0727
- Psychologie 0525
- Sciences 0714
- Sciences sociales 0534
- Sociologie de l'éducation 0340
- Technologie 0710

LANGUE, LITTÉRATURE ET LINGUISTIQUE
- Langues
  - Généralités 0679
  - Anciennes 0289
  - Linguistique 0290
  - Modernes 0291
- Littérature
  - Généralités 0401
  - Anciennes 0294
  - Comparée 0295
  - Mediévale 0297
  - Moderne 0298
  - Africaine 0316
  - Américaine 0591
  - Anglaise 0593
  - Asiatique 0305
  - Canadienne (Anglaise) 0352
  - Canadienne (Française) 0355
  - Germanique 0311
  - Latino-américaine 0312
  - Moyen-orientale 0315
  - Romane 0313
  - Slave et est-européenne 0314

PHILOSOPHIE, RELIGION ET THÉOLOGIE
- Philosophie 0422
- Religion
  - Généralités 0318
  - Clergé 0319
  - Études bibliques 0321
  - Histoire des religions 0320
  - Philosophie de la religion 0322
- Théologie 0469

SCIENCES SOCIALES
- Anthropologie
  - Archéologie 0324
  - Culturelle 0326
  - Physique 0327
- Droit 0398
- Économie
  - Généralités 0501
  - Commerce-Affaires 0505
  - Économie agricole 0503
  - Économie du travail 0510
  - Finances 0508
  - Histoire 0509
  - Théorie 0511
- Études américaines 0323
- Études canadiennes 0385
- Études féministes 0453
- Folklore 0358
- Géographie 0366
- Gérontologie 0351
- Gestion des affaires
  - Généralités 0310
  - Administration 0454
  - Banques 0770
  - Comptabilité 0272
  - Marketing 0338
- Histoire
  - Histoire générale 0578
  - Ancienne 0579
  - Médiévale 0581
  - Moderne 0582
  - Histoire des noirs 0328
  - Africaine 0331
  - Canadienne 0334
  - États-Unis 0337
  - Européenne 0335
  - Moyen-orientale 0333
  - Latino-américaine 0336
  - Asie, Australie et Océanie 0332
- Histoire des sciences 0585
- Loisirs 0814
- Planification urbaine et régionale 0999
- Science politique
  - Généralités 0615
  - Administration publique 0617
  - Droit et relations internationales 0616
- Sociologie
  - Généralités 0626
  - Aide et bien-àtre social 0630
  - Criminologie et établissements pénitentiaires 0627
  - Démographie 0938
  - Études de l' individu et de la famille 0628
  - Études des relations interethniques et des relations raciales 0631
  - Structure et développement social 0700
  - Théorie et méthodes. 0344
  - Travail et relations industrielles 0629
- Transports 0709
- Travail social 0452

SCIENCES ET INGÉNIERIE

SCIENCES BIOLOGIQUES
- Agriculture
  - Généralités 0473
  - Agronomie. 0285
  - Alimentation et technologie alimentaire 0359
  - Culture 0479
  - Élevage et alimentation 0475
  - Exploitation des péturages 0777
  - Pathologie animale 0476
  - Pathologie végétale 0480
  - Physiologie végétale 0817
  - Sylviculture et faune 0478
  - Technologie du bois 0746
- Biologie
  - Généralités 0306
  - Anatomie 0287
  - Biologie (Statistiques) 0308
  - Biologie moléculaire 0307
  - Botanique 0309
  - Cellule 0379
  - Écologie 0329
  - Entomologie 0353
  - Génétique 0369
  - Limnologie 0793
  - Microbiologie 0410
  - Neurologie 0317
  - Océanographie 0416
  - Physiologie 0433
  - Radiation 0821
  - Science vétérinaire 0778
  - Zoologie 0472
- Biophysique
  - Généralités 0786
  - Medicale 0760

SCIENCES DE LA TERRE
- Biogéochimie 0425
- Géochimie 0996
- Géodésie 0370
- Géographie physique 0368
- Géologie 0372
- Géophysique 0373
- Hydrologie 0388
- Minéralogie 0411
- Océanographie physique 0415
- Paléobotanique 0345
- Paléoécologie 0426
- Paléontologie 0418
- Paléozoologie 0985
- Palynologie 0427

SCIENCES DE LA SANTÉ ET DE L'ENVIRONNEMENT
- Économie domestique 0386
- Sciences de l'environnement 0768
- Sciences de la santé
  - Généralités 0566
  - Administration des hipitaux 0769
  - Alimentation et nutrition 0570
  - Audiologie 0300
  - Chimiothérapie 0992
  - Dentisterie 0567
  - Développement humain 0758
  - Enseignement 0350
  - Immunologie 0982
  - Loisirs 0575
  - Médecine du travail et thérapie 0354
  - Médecine et chirurgie 0564
  - Obstétrique et gynécologie 0380
  - Ophtalmologie 0381
  - Orthophonie 0460
  - Pathologie 0571
  - Pharmacie 0572
  - Pharmacologie 0419
  - Physiothérapie 0382
  - Radiologie 0574
  - Santé mentale 0347
  - Santé publique 0573
  - Soins infirmiers 0569
  - Toxicologie 0383

SCIENCES PHYSIQUES

Sciences Pures
- Chimie
  - Genéralités 0485
  - Biochimie 487
  - Chimie agricole 0749
  - Chimie analytique 0486
  - Chimie minérale 0488
  - Chimie nucléaire 0738
  - Chimie organique 0490
  - Chimie pharmaceutique 0491
  - Physique 0494
  - PolymÇres 0495
  - Radiation 0754
- Mathématiques 0405
- Physique
  - Généralités 0605
  - Acoustique 0986
  - Astronomie et astrophysique 0606
  - Electronique et électricité 0607
  - Fluides et plasma 0759
  - Météorologie 0608
  - Optique 0752
  - Particules (Physique nucléaire) 0798
  - Physique atomique 0748
  - Physique de l'état solide 0611
  - Physique moléculaire 0609
  - Physique nucléoire 0610
  - Radiation 0756
- Statistiques 0463

Sciences Appliqués Et Technologie
- Informatique 0984
- Ingénierie
  - Généralités 0537
  - Agricole 0539
  - Automobile 0540
  - Biomédicale 0541
  - Chaleur et ther modynamique 0348
  - Conditionnement (Emballage) 0549
  - Génie aérospatial 0538
  - Génie chimique 0542
  - Génie civil 0543
  - Génie électronique et électrique 0544
  - Génie industriel 0546
  - Génie mécanique 0548
  - Génie nucléaire 0552
  - Ingénierie des systòmes 0790
  - Mécanique navale 0547
  - Métallurgie 0743
  - Science des matériaux 0794
  - Technique du pétrole 0765
  - Technique minière 0551
  - Techniques sanitaires et municipales 0554
  - Technologie hydraulique 0545
- Mécanique appliquée 0346
- Géotechnologie 0428
- Matières plastiques (Technologie) 0795
- Recherche opérationnelle 0796
- Textiles et tissus (Technologie) 0794

PSYCHOLOGIE
- Généralités 0621
- Personnalité 0625
- Psychobiologie 0349
- Psychologie clinique 0622
- Psychologie du comportement 0384
- Psychologie du développement 0620
- Psychologie expérimentale 0623
- Psychologie industrielle 0624
- Psychologie physiologique 0989
- Psychologie sociale 0451
- Psychométrie 0632

NEURAL NETWORK SIGNAL PROCESSING TECHNIQUES FOR ADVENTITIOUS LUNG SOUND CLASSIFICATION

BY

KEVIN EDWARD FORKHEIM

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Abstract

This thesis compares different neural network techniques for classifying adventitious lung sounds, with an emphasis on wheeze detection. A neural network is a powerful tool of Artificial Intelligence that is capable of performing pattern recognition tasks, such as signal classification, quite well. Adventitious lung sounds are abnormal physiological sounds produced within the lungs and its associated airways. The abnormal sounds often indicate the existence of a disease. Auscultation, the process of listening to sounds produced within the body, is subjective. Existing computerized techniques for wheeze detection have been found to have a poor performance when detecting wheezes of low to moderate intensity. **The purpose of this research is to compare the performance of different representations of lung sounds and different types of neural network models at classifying wheezes. The result of combining multiple neural networks is also investigated.**

Lung sounds were recorded from the thorax region and divided into segments of 0.2 seconds. These segments were then classified as either containing a wheeze or not containing a wheeze by a trained clinician. A neural network was trained and tested using the raw signal, filtered, and Fourier Transform representations so that the best representation for the neural network could be found. Once the best representation was found, five different types of neural networks--Backpropagation (BP), Self Organizing Maps (SOM), Learning Vector Quantization (LVQ), Probabilistic Networks (PNN), and Radial Basis Functions (RBF)--were compared to find the best neural network for classifying wheezes. Experiments were also performed with different post-processing techniques such as thresholding and combining multiple classifiers.

The results of the research indicates that the Fourier Transform spectrum of the lung sounds provides the best representation for neural network wheeze detection. It was also found that LVQ and RBF networks provide the best single neural network classifier. In addition, combining the results of three neural networks--LVQ, RBF, and BP--provided the best performance with only one percent of the data being misclassified.

I dedicate this thesis
to my mother, Vera Forkheim and
my father, Edward Forkheim. They
taught me the value of hard word,
perseverance, and dedication.

And also to my grandmother,
Anna Forkheim, who passed away
during the final stage of this thesis.
She showed me the value of inner strength
and the fortune found in
family and friends.

"To make a great dream come true, the first requirement is a great capacity to dream; the second is persistence--a faith in the dream."

Hans Selye, M.D.

Acknowledgments

On completion of this thesis and other requirements of a Masters of Computer Science I would like to extend a warm thank you to the people who have helped and encouraged me. I realize that without these people and good fortune, I would have never achieved this goal.

I would like to express my sincere gratitude to David Scuse for his encouragement, enthusiasm, and countless hours of instruction, and to Hans Pasterkamp for his time, insight, and guidance. I would also like to thank Mark Evans for his advice and help for the duration of the research.

I wish to gratefully acknowledge the help and instruction that Yuns Oh, Dereck Meek, and Epiphany Vera provided me. I would also like to thank PixSoft Corporation for lending me an audio board for the duration of the research and Hans Pasterkamp for allowing me to use the R.A.L.E. © software.

I would also like to thank the graduate students, academic and support staff from the department of Computer Science for their support and patience over the duration of the thesis. I am particularly indebted to Tom Dubinski and Gilbert Detillieux for their technical assistance and Lynne Romuld for her academic advice.

Finally, I would also like to thank Donal McCarthy and Dean Kriellaars for their comments on the thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Θ	Big-O notation
σ	width of categories for PNN and RBF
"	inch
ε	learning rate of the neural network
A_f	Average frequency threshold constant
BP	Backpropagation
2CD	Two Cycle Duration
FEV_1	Forced Expiratory Volume in one second
FFT	Fast Fourier Transform
f_p	frequency of maximum power
FT	Fourier Transform
Hz	Hertz
i	$\sqrt{-1}$
IDW	Initial Deflection Width
$Input_{ij}$	Input to neural network unit in row i and column j
kHz	kiloHertz
LVQ	Learning Vector Quantization
msec	millisecond
m	meter
NN	Neural Network
O_a	calculated Output value
O_c	Correct Output value
p (pp)	page (pages)
Pavg	Average Power
PNN	Probabilistic Neural Network
RBF	Radial Basis Function
RMS	Root Mean Square
sec	second
SOM	Self Organizing Map
V	Volt
W	Watt
w_{ij}	weight of neural network unit in row i and column j

Chapter 1

Introduction

1.1 Statement of the Problem

Auscultation, the art of listening to the sounds produced within the body with a stethoscope, is more than 170 years old. It is one of the commonest procedures performed during a physical examination and is the most important technique for assessing air flow in the tracheobronchial tree and the lungs (Bates et al., pp 244-245). Little has changed since 1816 when Laennec used his first stethoscope to auscultate a young woman with heart disease. However, in some hospital laboratories the stethoscope is being replaced by a small microphone that records data for a computer that is capable of performing complex analytical techniques on the lung sounds. Computer analysis of lung sounds is an emerging discipline that is still in its infancy. It is hoped that one day, it will provide physicians with a reliable non-invasive tool for diagnosing lung abnormalities such as asthma and cystic fibrosis. This thesis outlines how neural networks can be used to classify lung sounds.

For the most part, this thesis deals with detecting wheeze lung sounds. Computerized wheeze detection can be very important to the medical profession for many reasons. Conventional tests for pulmonary obstructive diseases often involve spirometry which requires the patient's cooperation. However, young infants or patients who are too sick to perform these procedures are difficult to assess because they are unable to perform a spirometric test. An objective wheeze classification system could be used to assess these patients. In addition, conventional auscultation is subjective and often involves much interobserver and intraobserver variability which may lead to an uncertain diagnosis

(Pasterkamp et al., March 1987, pp 376-381). Also, having extra clinical information regarding a patient's condition may prejudice a clinician's assessment of the lung sounds, and may cause the clinician to overlook certain symptoms (Ono et al., p 290). A computerized wheeze detection algorithm is very objective and would not have these biases. Computerized wheeze detection algorithms can be utilized for continuous monitoring of a patient instead of having a medical practitioner continually monitoring the patient. This would free the practitioner for other duties. Finally, computerized analysis of lung sounds provides a non-invasive monitoring device that does not disturb the patient's physiological state. This is important for diseases such as nocturnal asthma which requires that the patient's sleep stage be related to airflow obstruction. Under these circumstances, waking up the patient to perform a spirogram is impractical (Loudon 1993, p 585).

The goal of this research is to evaluate neural networks as a tool for wheeze detection and to construct a suitable representation and neural network system that is able to classify lung sounds with a high degree of accuracy and minimize the amount of data that is misclassified. A comparison of neural network classifiers and other existing algorithmic techniques for classifying wheezes is also made. In the past, the existing algorithmic techniques for wheeze detection have been found to be inadequate for measuring wheezes of low to moderate intensity.

A review of lung sound nomenclature and existing lung sound analysis techniques is also made in this thesis so that the reader can better understand some of the problems that exist in lung sound classification. An introduction to neural networks and signal processing is presented so that the reader may comprehend the results of the research presented in this thesis.

1.2 Research Issues

The detection methods proposed in this research utilize neural networks. Neural networks are powerful tools of Artificial Intelligence that are very good at pattern classification. Each type of neural network has its own architecture and characteristics that must be optimized to obtain the best possible performance. This thesis examines five different types of neural networks--Backpropagation, Self Organizing Maps, Learning Vector Quantization, Probabilistic Networks, and Radial Basis Functions--and compares the performance of each type of neural network at classifying asthmatic lung sounds. Different parameters such as the size of the hidden layer and learning rates were also altered to optimize performance of each network.

In addition to using different types of neural networks, different pre-processing and post-processing techniques were used. Signals were pre-processed by using different mathematical operators such as the Fourier Transform, first derivative, and filtering to increase classification performance. The results of neural network processing were also post-processed to improve performance. Both thresholding and consensus diagnosis were used as post-processing techniques.

1.3 Thesis Overview

Chapter 2 presents a brief summary of lung sound nomenclature. The chapter begins with a short review of respiration and the anatomy of the respiratory system. A short history of auscultation and lung sound nomenclature is also presented. A major emphasis has been placed upon the rather poor performance of medical practitioners in the field of lung sound classification. The chapter ends with a discussion of the lung sound nomenclature used in the thesis.

Chapter 3 provides an introduction to sound and signal processing. The basics of lung sounds, including frequency, intensity, duration, and quality are outlined. The chapter also summarizes signal processing by discussing issues involved in signal sampling, frequency representation, and the Fast Fourier Transform algorithm. A discussion of how the Fast Fourier Transform algorithm was used in the research is also provided.

Chapter 4 provides an overview of some of the previous research performed into lung sound analysis and the method that was used to record lung sounds for this research. A very brief discussion of crackle detection is provided. Two wheeze detection algorithms are discussed and the results of testing these algorithms on different wheezes are also presented. The chapter ends with a brief discussion on how the lung sounds were recorded for use in the thesis and some uses of an automated wheeze detection device.

Chapter 5 provides a brief introduction to neural networks. Five types of networks--Backpropagation, Self Organizing Maps, Learning Vector Quantization, Probabilistic Networks, and Radial Basis Functions--are introduced in this chapter. Each neural network is explained at a level that is required to understand how it was utilized in the research. The biological basis and comparison with other programming techniques are also provided to give the reader a better understanding of the neural network field. The chapter ends with a brief discussion of other successful applications of neural networks.

Chapter 6 presents some of the research findings that investigate how the classification results were affected by signal representation. A backpropagation neural network is used to find the most suitable representation of the lung sounds for the neural network classifiers. The results of using the raw signal, filtered, and Fourier Transform representations are compared.

A comparison of the five different types of neural network processors is made in Chapter 7. Different network parameters are modified in an attempt to optimize network performance. Different post-processing techniques such as thresholding and combining multiple classifiers are also discussed in this chapter.

Chapter 8 summarizes the research presented in the thesis and highlights the more important results. It also discusses some of the areas that require further investigation in the future. Finally, the chapter ends with a brief summary of other applications of the research.

Two appendices are also included in this thesis. Appendix A contains the Fast Fourier Transform algorithm which is described in Chapter 3. Appendix B presents the definitions for specificity, sensitivity, overall performance, and other performance metrics used in this thesis.

It is the author's intent to provide a thorough and understandable review of the information acquired during the duration of the research project so that others may continue the research into this field.

Chapter 2

Respiration and its Associated Sounds

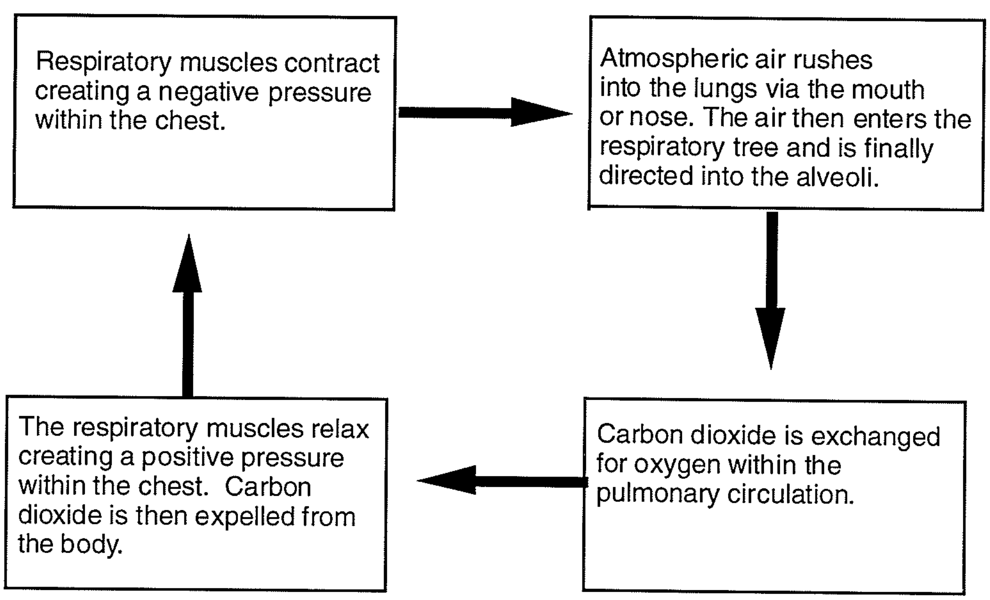
2.1 Respiration

Every living cell within the human body requires a fresh supply of oxygen and the removal of carbon dioxide in order to survive. The lungs play an integral role in the complex system which allows respiration to occur. The term respiration actually refers to several important processes, including:

> 1) ventilation, movement of air into and out of the respiratory system;
>
> 2) transport of oxygen and carbon dioxide within the blood;
>
> 3) gas exchange between blood and tissues; and
>
> 4) cellular metabolism which is an oxygen and carbohydrate requiring process that creates carbon dioxide, water, and energy in the form of high energy molecules.
>
> (Seeley and Stephens, pp 695-697)

The respiratory system works in a continuous cycle as described in Figure 2.1.

Figure 2.1: Breathing Cycle



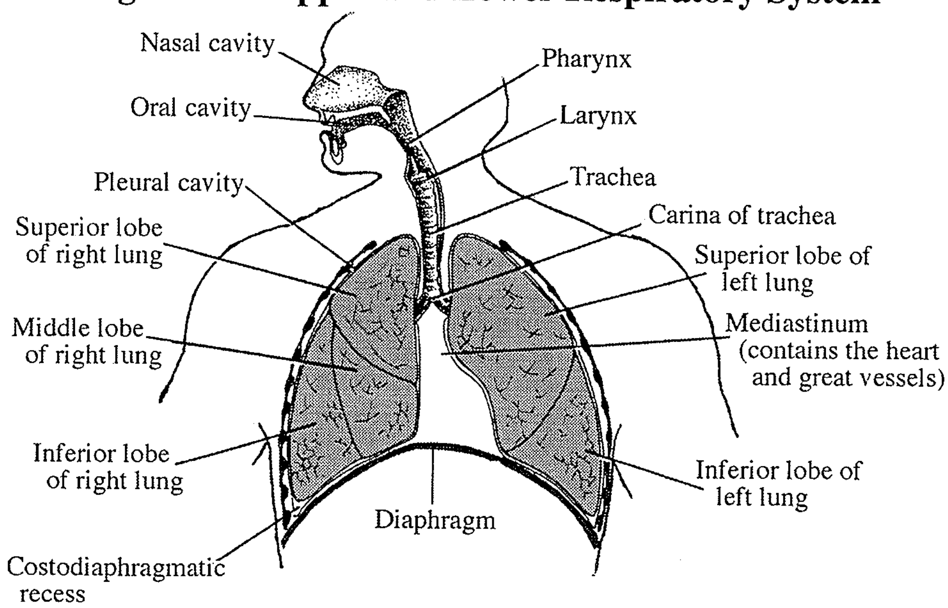
The anatomy of the respiratory system provides a superior environment for gas exchange and for moving air in and out of the lungs.

2.2 Anatomy of the Respiratory System

The respiratory system can be divided into two sections: (1) the upper respiratory system consisting of the nose, mouth, pharynx, larynx, and trachea; and (2) the lower respiratory system, consisting of the bronchi, bronchioles, and alveoli.

Air enters the respiratory system via the nose or mouth where it is warmed to body temperature and large particles of dust are removed. The inhaled air then proceeds to the pharynx, which is the common opening for the digestive and respiratory systems, and to the larynx which contains the vocal chords. The trachea or windpipe, which is a membranous tube, divides into two smaller airways, the primary bronchi. The primary bronchi give rise to secondary bronchi which also divide further (Seeley and Stephens, pp 695-705). The respiratory system has five lung lobes, three on the right and two on the left. The upper and lower respiratory systems are depicted in Figure 2.2.

Figure 2.2: Upper and Lower Respiratory System



(Figure 2.2 adapted from Moore, p 49.)

The bronchi continue to divide until they reach the bronchiole stage, which are smaller airways devoid of cartilage. Bronchioles are 1 mm or less in diameter and contain smooth muscle which can contract as may occur during an asthmatic attack. Each respiratory bronchiole further divides into alveolar ducts that end as clusters of airsacs, called alveoli. There are about eight million alveoli, each about 0.1 to 0.5 mm in diameter, within the lungs of an average adult. This large concentration of alveoli provides a large surface area so that oxygen and carbon dioxide can easily be exchanged within the lungs (Kryger, p 11).

2.3 Auscultation

> In 1816 I was consulted by a young woman presenting general symptoms of disease of the heart. Owing to her stoutness, little information could be gathered by application of the hand and percussion. The patient's age and sex did not permit me to resort to the kind of examination I have just described (ie., direct application of the ear to the chest). I recalled a well known acoustic phenomenon: namely, if you place your ear against one end of a wooden beam, the scratch of a pin at the other extremity is most distinctly audible. It occurred to me that this physical property might serve a useful purpose in the case with which I was then dealing. Taking a sheaf of paper I rolled it into a very tight roll, one end of which I placed over the precordial region, whilst I put my ear to other. I was both surprised and gratified at being able to hear the beating of the heart with much greater clearness and distinctness than I had ever done before by direct application of my ear.
>
> I at once saw that this means might become a useful method for studying, not only the beating of the heart, but likewise all movements capable of producing sound in the thoracic cavity, and that consequently it might serve for the investigation of respiration, the voice, râles, and even possibly the movements of a liquid effused into the pleural cavity or pericardium.

Taken from an English translation of Laennec's book *De l'Auscultation Médiate.*
(Hale-White, p 25)

Auscultation is the most important technique for the clinical assessment of airflow in the tracheobronchial tree and lungs. Auscultation involves: (1) listening to the sounds generated by breathing; (2) listening for adventitious (added) sounds; and (3) listening to the sounds of the patient's spoken or whispered voice as they are transmitted through the chest wall (Bates et al., pp 244-245). Auscultation of the lungs began to decline shortly after chest x-rays became popular because physicians thought that more information was available from x-rays than from auscultation (Fitzgerald, pp 377-382). One of the prominent textbooks reported that:

> The stethoscope is largely a decorative instrument insofar as its value in diagnosis of pulmonary diseases is concerned. Nevertheless, it occupies an important place in the art of medicine. Apprehensive patients with functional complaints are often relieved as soon as they feel the chest piece on their pectoral muscles.
>
> (Rubin and Rubin)

However, several things have happened that have reversed this trend. First, it was recognized that several disease states can be detected only by auscultation. For example, wheezes can be auscultated in asthmatic patients and crackles in patients suffering from interstitial lung disease, but the radiographs may be perfectly normal. Also, new techniques for recording lung sounds have helped to clarify the significance of adventitious lung sounds. Finally, efforts have been made to standardize and simplify lung sound nomenclature. Although there is still some disagreement on the nomenclature, much progress has been made (Murray and Nadel, pp 576-577).

When Laennec developed his first stethoscope, there was no widely accepted terminology for the various lung sounds produced within the diseased lungs and airways. He initially proposed that all abnormal sounds be called "râles" (Hale-White, pp 149-151). He further classified râles into four groups: (1) moist râle or *crepitation*; (2) mucous râle or *gurgling;* (3) dry sonorous râle or *snoring*; (4) dry sibilant râle or *whistling*. A fifth category was later added under the heading of dry crepitant râles, or *crackling* (Forgacs, p 1). He also noted that lung sounds were easier to distinguish than to describe so he clarified the definitions by using examples from everyday life. For example, the dry sonorous râle was said to "... resemble the cooing of a dove" (Hale-White, p 150). However, the interpretation of verbal definitions was still somewhat subjective and depended upon the physician's prior experience with the sounds described.

Other physicians, including Laennec's own pupils, later added their own observations and suggested changes to the original classification and the clarity in Laennec's definitions was soon lost (Forgacs, p 1). The semantics of lung sound classification were further altered when they were translated into other languages. For example, a German translation of the original terminology was later translated into Chinese, and the Chinese characters were later used by Japanese physicians to describe the sounds (Mikami et al., p 342).

Laennec also used the word rhonchus instead of râle in the presence of patients because he felt it might be misinterpreted by the patient to mean the "death rattle." However, in an English translation of Laennec's work, the terms râle and rhonchus were translated as having different meanings. Confusion about the terminology still exists more than 170 years after Laennec's initial work (Loudon and Murphy, p 664).

The confusion about lung sound classification has been studied by many researchers. For example, in Pasterkamp's paper *Subjective Assessment vs. Computer Analysis of Wheezing in Asthma*, he describes the results of an experiment which involved forty health professionals analyzing ten different breath sounds produced by asthmatic patients (Pasterkamp et al., August 1987, pp 376-381). The professionals were asked seven questions which pertained to the classification of asthmatic lung sounds. The study was also repeated with the same health professionals a short time later. The results are outline in Table 2.1.

Table 2.1: Kappa Values of Observer Agreement

Health Care Professional	Intraobserver Agreement	Interobserver Agreement
Residents	0.43 ± 0.23	0.36 ± 0.13
Nurses	0.51 ± 0.24	0.43 ± 0.17
Pediatricians	0.49 ± 0.22	0.36 ± 0.13
Physiotherapists	0.45 ± 0.19	0.40 ± 0.15

Note: Mean kappa values ± standard deviation
Interobserver agreement measures agreement between different observers.
Intraobserver agreement measures agreement between the same individual at two different times.
Kappa statistics takes into account the chance agreement between different examiners.
(Table 2.1 adapted from Pasterkamp et al., August 1987, p 379.)

This study demonstrates that not only do health professionals not agree with each other in regards to how a lung sound should be classified, but also that observers have almost as large a discrepancy when compared against themselves. Other researchers also compared the assessment of physical signs such as wheezes and crackles, of the respiratory system amongst different physicians and found interobserver agreement to "... fall midway between chance and total agreement" (Smyllie et al., p 413).

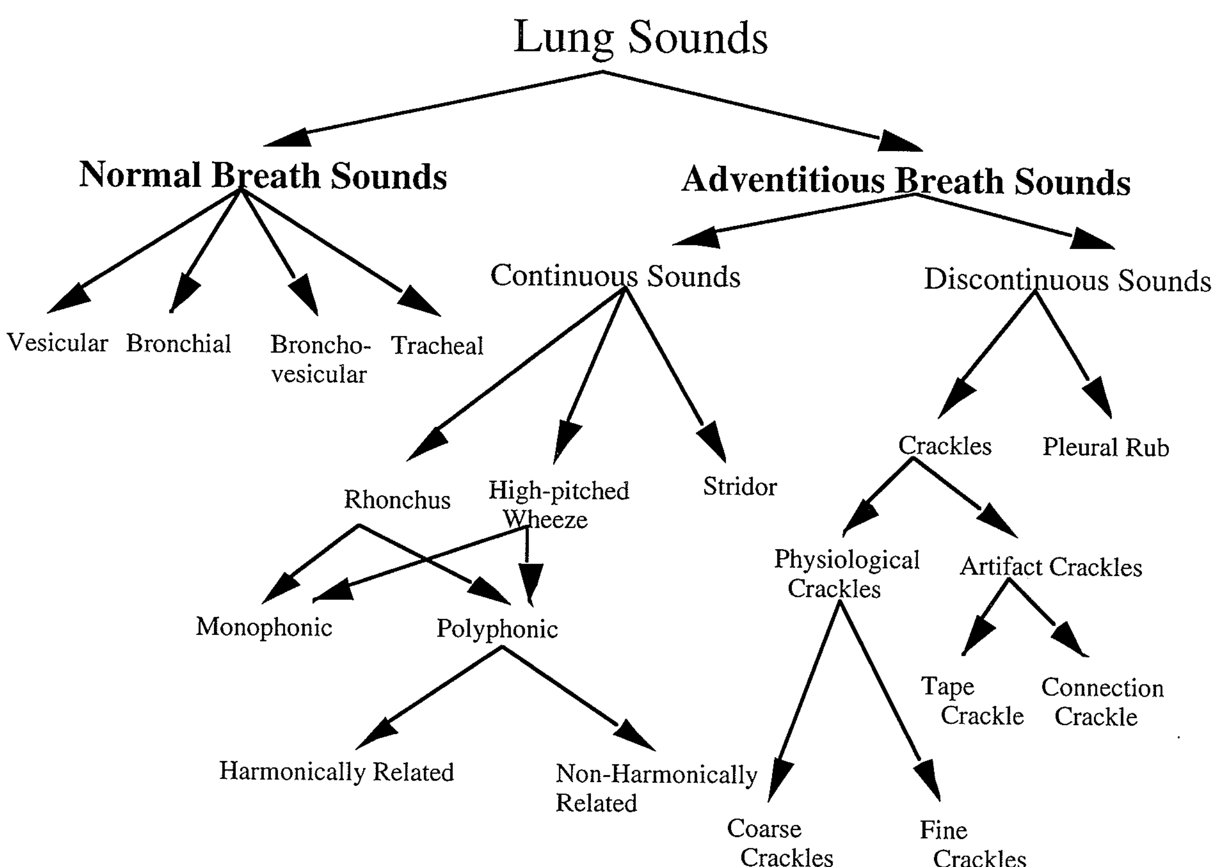
Godfrey et al. conducted a similar study to Pasterkamp and Smyllie's studies and obtained similar results. However, Godfrey et al. also included a training session for some of the physicians to educate them about lung sound nomenclature. This education provided only a slight improvement in classification performance. The author of the study noted that "these results rather suggest that while attention to detail may result in improvement in familiar signs, a certain amount of observer variation will always remain" (Godfrey et al., p 9). Loudon and Murphy also mentioned that lung sounds contain an enormous amount of information "... [that] is so great that it overtaxes the clinician's memory" (Loudon and Murphy, p 671).

Nomenclature used to describe lung sounds still contains much variability. A survey of case reports published in eight different medical journals between 1982 and 1987 found that terminology for adventitious lung sounds is not consistent; however, consistency is improving (Wilkins et al., pp 523-525). The improvement was attributed to the journal editors changing less commonly used terminology to the more accepted terminology. The improvement was also attributed to a set of nomenclature recommended by the American Thoracic Society in a newsletter published in 1977 (American Thoracic Society, 1977) being more widely used.

2.4 Current Lung Sound Nomenclature

The lung sound terminology used in this thesis is described below. Although a few authors, (Fenton et al.) and (Cohen and Landsberg), have attempted to objectively define various lung sounds mathematically, the area of lung sound classification is still subjective and the exact definitions are debatable. The categorization of lung sounds used in this thesis divides the lung sounds into two categories--normal breath sounds and adventitious lung sounds--with the categories being further divided as shown in Figure 2.3.

Figure 2.3: Lung Sound Classification



2.4.1 Normal Breath Sounds

Four different types of breath sounds--vesicular, bronchial, bronchovesicular, and tracheal--can be heard over the chest and lungs of a healthy person without anatomical or physiological abnormality during quiet breathing. The description and symbolic representation of these sounds can be found in Table 2.2.

Table 2.2: Normal Breath Sounds

Description | **Symbolic Representation***

Inspiration | Expiration

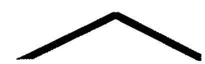
Vesicular Breath Sounds:

Heard over most of the normal chest. A relatively soft and low pitched sound that resembles a sigh. The inspiratory phase is longer by a 3:1 ratio and is much louder than the expiratory phase. There is no pause between inspiration and expiration.

Bronchial Breath Sounds:

Heard over the manubrium and indicates a pathology if heard over the peripheral lung fields. These sounds are relatively loud and high-pitched with a hollow quality. Expiration is louder and longer than inspiration. There is a slight pause between inspiration and expiration.

Bronchovesicular Breath Sounds:



The sounds are normally audible near the first and second intercostal spaces on the chest and between the scapulae on the back. In both locations sounds are more audible on the right than the left. These sounds are intermediate between bronchial and vesicular breath sounds. Inspiration and expiration are of similar quality and duration. There is no pause between inspiration and expiration.

Tracheal Breath Sounds:

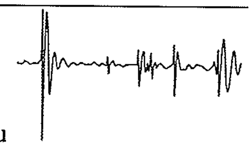
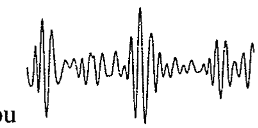
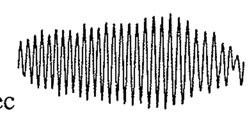
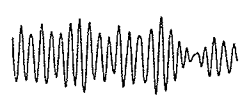
Heard normally over the extrathoracic portion of the trachea. These sounds are very loud and high pitched with a harsh quality. Inspiration and expiration are of similar quality and duration with a slight pause between the two.

* The thickness of the bars indicates intensity; the thicker the bar, the louder the sound. The incline represents pitch; the steeper the incline, the higher the pitch. Symbolic Representation adapted from (Murray and Nadel, p 572).

2.4.2 Adventitious Lung Sounds

The current terminology for adventitious or added lung sounds was first formulated during a meeting of the International Lung Sound Association in 1976 in an attempt to reduce the confusion in the terminology. This terminology was later adopted by the American Thoracic Society in 1977 and is shown in Table 2.3 (American Thoracic Society, 1977).

Table 2.3: Classification of Adventitious Lung Sounds

Acoustic Characteristics	American Thoracic Society	Common Synonyms	Laennec's Original Terms	Time Expanded Waveform*
Discontinuous, interrupted explosive sounds Loud, low in pitch; Average values: IDW = 1.25 msec, 2CD = 9.32 msec	Coarse crackle	Coarse râle Wet crackle**	Râle muquex ou gargouillement	
Discontinuous, interrupted explosive sounds Less loud and of shorter duration than above; higher in pitch than coarse crackles or rales; average values: IDW = 0.92 msec, 2CD = 6.02 msec	Fine crackle	Fine râle Dry crackle**	Râle humide ou crepitation	
Continuous sounds longer than 250 msec, high pitched; dominant frequency of 400 Hz or more, a hissing or whistling sound	Wheeze	Sibilant rhonchus, High-pitched wheeze***	Râle sibilant sec ou sifflement	
Continuous sounds longer than 250 msec, low pitched; dominant frequency about 200 Hz or less, a snoring sound	Rhonchus	Sonorous rhonchus Low-pitched wheeze***	Râle sec sonore ou ronflement	

Definition of abbreviations: IDW = Initial Deflection Width; 2CD = Two Cycle Duration. Refer to Section 2.4.2.1 for a detailed explanation of the measurement of IDW and 2CD.

* Time expanded waveform adapted from Mikami et al., p 344.

** These terms were not included in the original table, although they were are still commonly taught in medical schools (R.A.L.E.).

*** These terms were not included in the original table, although they are commonly used in the British literature (Kryger, p 360).

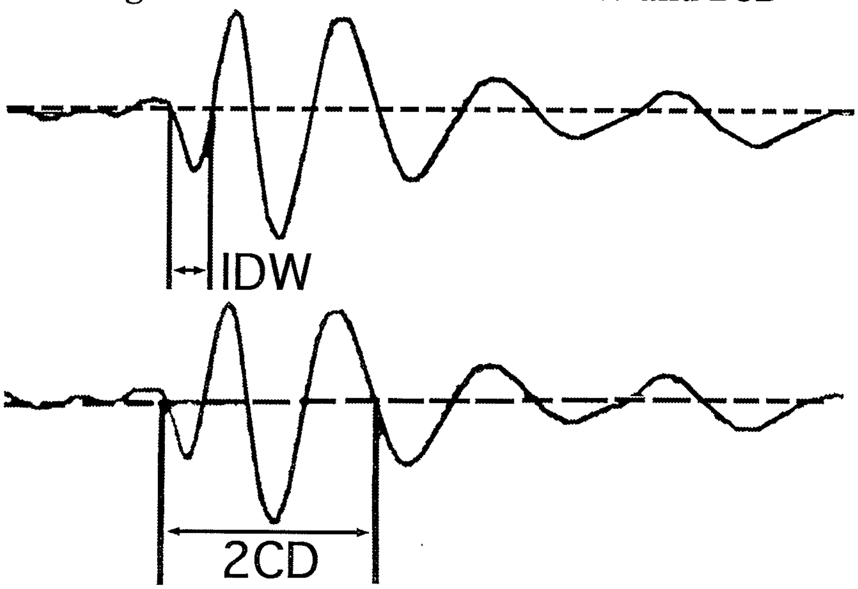
(Table 2.3 adapted from Loudon and Murphy, p 664.)

Adventitious breath sounds are divided into two categories, discontinuous and continuous breath sounds. The former tends to have a duration of 20 msec or less and the latter has a duration of 100 msec or more (Loudon, 1982, p 6). Adventitious sounds that have a duration between the boundaries of the two categories should be classified into one of the two categories by analyzing other properties of the sound such as pitch or waveform appearance.

2.4.2.1 Discontinuous Lung Sounds

Crackles: Crackles are short interrupted sounds with a wide spectrum of frequencies between 200 and 2,000 Hz (Forgacs, Functional Basis, p 399). As stated in Table 2.3, crackles, have been divided into two groups, coarse crackles (•••••) and fine crackles (····) (Bates, p 246). This characterization has been performed by (Holford) by utilizing a large number of features, but mainly the Initial Deflection Width (IDW) and two Cycle Duration (2CD). The IDW is the time in milliseconds between the first deviation, above or below the baseline, to its subsequent return to the baseline. The 2CD is the time in milliseconds required for the wave to complete two S-shaped deviations or four zero crossing (Mikami et al., p 345). The measurement of these values is demonstrated in Figure 2.4.

Figure 2.4: Measurement of IDW and 2CD



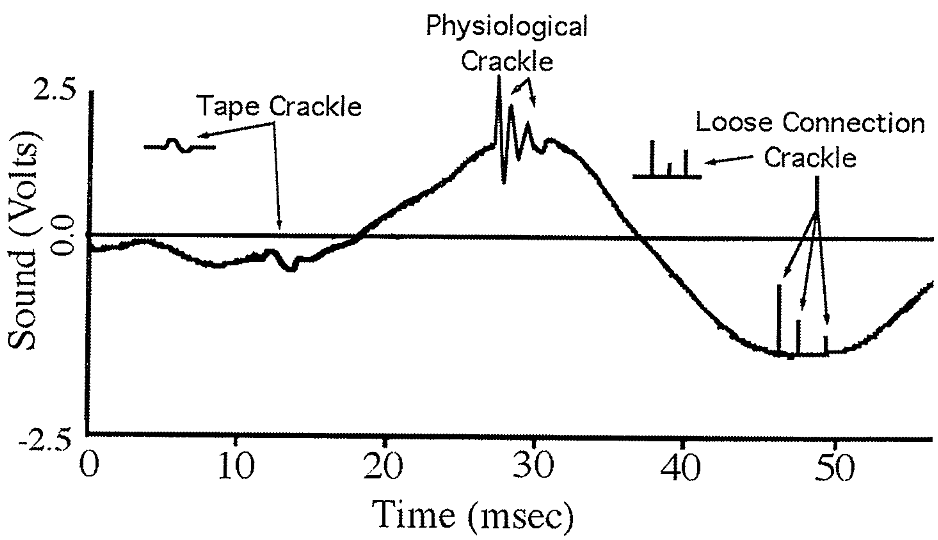
(Figure 2.4 adapted from Murphy, 1985, p 212.)

A paper titled "Measuring Crackles," investigated using the width of the largest deflection instead of the IDW. It is often difficult to define the beginning and end of a crackle, and so this problem can be avoided by using the largest deflection. It was reported that the length of 2CD measured from the largest deflection could be more accurately measured by a human than measuring the first 2CD or the IDW (Hoevers and Loudon, pp 1240-1243).

Crackles are believed to be caused by small airways becoming open or closed. It is believed that crackles may also be caused by air bubbling through excess liquid in the airways (Forgacs, pp 55-56). The quality and position of a crackle in the respiratory cycle can be characteristic of certain diseases. For example, late inspiratory fine crackles are a characteristic sign of fibrosing alveolitis and interstitial pulmonary edema. Early inspiratory crackles at the base of the lungs are characteristic of chronic bronchitis, and late inspiratory crackles are characteristic of left ventricular failure.

Artifact Crackles: Artifact crackles are artificial products introduced during the process of recording lung sounds. The source may be from the adhesive tape used to affix the microphone to the subject's body, as in the case of a tape crackle. It may also be due to a loose wire connecting the microphone to the data collecting device, a wire rubbing against another wire and causing static during the process of recording the lung sounds, or from mechanical vibration. A physiological crackle has been compared to the two types of artifactual crackles in Figure 2.5. Without examining the time or frequency representation of the crackle, it may be very difficult to distinguish between different types of crackles by only listening to the sound.

Figure 2.5: Waveform Representation of the Types of Crackles



Pleural Rub: The pleural rub or pleural crackle is caused by friction between the visceral and parietal pleura of the lungs secondary to an inflammatory or a cancerous process. These sounds, which vary in intensity, are discontinuous and often have a leathery or creaking quality. Rubs can usually be heard during both inspiration and expiration, but are variable and short in duration (Kryger, p 362).

2.4.2.2 Continuous Lung Sounds

Wheeze and Rhonchi: As stated in Table 2.3, high-pitched wheezes () and rhonchi () are continuous musical sounds which are either high or low-pitched, respectively. Wheezes can range in frequency from 60 to 2,000 Hz (Forgacs, p 44). The classification scheme for continuous sounds shown in Table 2.3 requires that continuous sounds be at least 250 msec in duration. In one paper, Loudon defined a wheeze as an abnormal, more or less melodic tone of distinguishable pitch which lasts for at least 100 msec (Loudon, 1982, p 6). In another paper, Loudon defined a wheeze as "a continuous sound lasting more than 20 or 30 msec and of sufficient duration that they have a 'musical' quality, or at least a perceptible pitch" (Loudon, 1993, p 583). However, it often takes about 100 msec for even a pure tone to develop a musical quality (R.A.L.E.). Therefore, the author of the thesis has adopted the former wheeze terminology and the research presented here requires that the duration of a wheeze be at least 100 msec. Lung sounds that have wheeze-like characteristics that are less than 100 msec, but greater than 30 msec, will be termed short wheezes or squawks (R.A.L.E.).

As presently stated, the definitions of continuous lung sounds do not encompass continuous lung sounds with dominant frequencies that are greater than 200 Hz and less than 400 Hz. The author of the thesis therefore has chosen that continuous musical lung sounds with the dominant frequency greater than or equal to 400 Hz be called a high-pitched wheeze. Continuous musical lung sounds with the dominant frequency below 400 Hz should be called a low-pitched wheeze or rhonchus. To summarize, the definition for high-pitched wheeze and rhonchus used in the thesis are stated below.

High-pitched Wheeze:

Continuous musical sound at least 100 msec in duration; dominant frequency is $\geq$ 400Hz.

Rhonchus (or Low-pitched Wheeze):

Continuous musical sound at least 100 msec in duration; dominant frequency is $<$ 400 Hz.

In addition to pitch and frequency, wheezes may also be classified according to complexity (monophonic, polyphonic, or harmonically related), duration (long or short), and timing in the respiratory cycle (inspiratory or expiratory) (Forgacs, p 4). Monophonic wheezes have one frequency or spike in the corresponding Fourier Transform (FT) which is not attributable to noise. Polyphonic wheezes have more than one frequency or spike in the corresponding FT which is not attributable to noise (Forgacs, pp 48-54). Polyphonic wheezes that have spikes that occur at multiples of the lowest frequency spike that is attributable to the wheeze, are termed harmonically related. The FT of monophonic and polyphonic wheezes are shown in Figures 2.6 and 2.7 respectively.

Figure 2.6:
FT of a Monophonic Wheeze

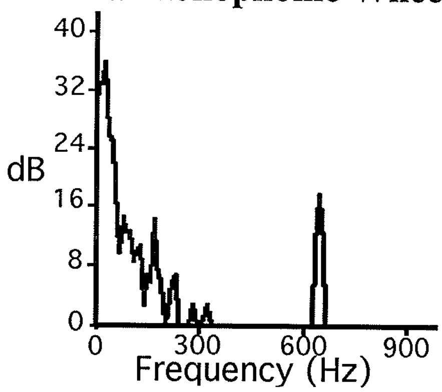
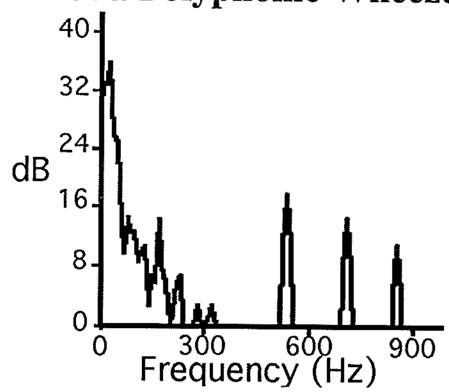


Figure 2.7:
FT of a Polyphonic Wheeze



Wheezes are produced in the intrathoracic airways when the walls are narrowed to a point where the opposing walls nearly touch one another. The rapid oscillation between the open and closed positions produces a single musical note. The pitch of the note is independent of the caliber or length of the airways and the density of the inspired gas. The pitch depends only upon the property of the mass which is oscillating and the velocity of air in the airway ((Forgacs, p 46) and (Murray and Nadel, p 575)). A large and flabby mass will produce a low pitch and a light and stiff mass will produce a higher pitch. This model of wheeze production has been compared to a child's toy trumpet which can sound a single note (Forgacs, p 46). Grotberg and Gavriely provide a good model of how wheezes are believed to be produced in their paper (Grotberg and Gavriely).

Although a wheeze is a highly specific indicator of airflow obstruction (Sanchez et al., p 28), it is important to remember that "all that wheezes is not asthma" (Kryger, p 361). Airflow obstruction may also be caused by many factors as outlined in Table 2.4.

Table 2.4: Factors Causing a Wheeze

Site of Pathology	Factor Causing Wheeze
Endobronchial	Tumor, Foreign body Mucus (bronchitis)
Bronchial wall	Increased mucosal thickness (chronic bronchitis, asthma) Increased bronchial muscle tone (asthma)
Extrabronchial	Compression (tumors, lymph nodes) Loss of elastic recoil of surrounding lung (emphysema)

(Table 2.4 adapted from Kryger, p 361.)

Stridor: Stridor is a high-pitched crowing sound that has its maximal intensity on inspiration and is flow dependent. It is usually caused by an extrathoracic obstruction in the trachea or larynx. (Kryger, p 362).

Squeaks and Squawks: Squeaks and squawks have been defined as short wheezes (of duration less than 100 msec), which are usually produced in association with interstitial lung disease. They are usually preceded by one or more crackles, and are found in late inspiration (R.A.L.E.).

2.5 Summary

In Chapter 2 a brief introduction to Respirology and lung sound nomenclature was presented. A brief review of respiration and the anatomy of the respiratory tract was presented so that the reader may better understand the findings of auscultation. The lung sound nomenclature used in this thesis was described, with adventitious lung sounds being divided into two categories based on the duration of the lung sounds. Both crackles and wheezes were further examined and their definitions were provided.

In the remainder of this thesis, the concentration will be placed on wheeze detection. Although there may be disagreement among health care professionals as to how adventitious lung sounds should be classified, the author has chosen a reasonably simple definition of a wheeze that will allow consistency in the classification of lung sounds. Much work and research is still on-going into the correct nomenclature and definitions for lung sounds.

Chapter 3

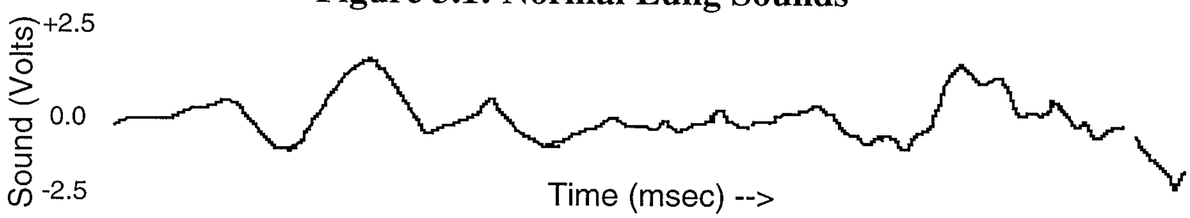
Sound and Signal Processing

3.1 Sound

A sound may be defined as vibrational energy that the human ear can perceive (Berg and Stork, p 9). Therefore, a lung sound is vibrational energy produced within the lungs or its associated airways that the human ear can perceive (Murray and Nadel, p 571). Often this perception requires an acoustical device such as a stethoscope. All sounds, including lung sounds, are characterized by their **frequency**, **intensity**, **duration**, and **quality**.

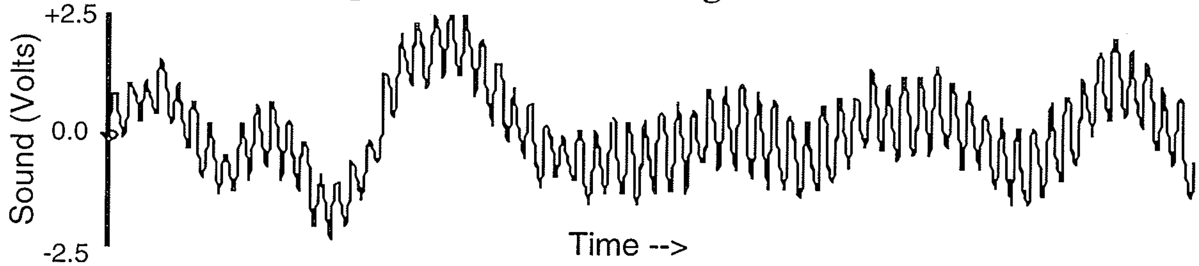
A microphone is a mechanical device which transforms vibrational energy into electrical energy that can be measured or recorded. A microphone was attached to the anterior upper chest of a normal individual and a sample of the recording is shown in Figure 3.1

Figure 3.1: Normal Lung Sounds



The sample contains no adventitious lung sounds and so it appears stochastic and does not follow any particular pattern. However, if the same individual developed asthma and the lung sounds were then recorded during an asthmatic attack, the recording may contain wheeze sounds superimposed on the normal sounds, and may appear as shown in Figure 3.2.

Figure 3.2: Wheeze Lung Sounds



The main difference between the two signals is the repeated oscillatory motion caused by the wheezing. Adventitious lung sounds, such as wheezes, may be superimposed on normal lung sounds, or other adventitious lung sounds, such as crackles.

The oscillatory motion of the wheeze of Figure 3.2 has a **frequency** which denotes the number of vibrations or cycles per unit time. Typically, the frequency is measured in Hertz (Hz) which indicates the number of cycles per second. The frequency may also be measured in kiloHertz (kHz), such that 1 kHz = 1,000 cycles/second. The frequency of a sound determines our perception of pitch. The higher the frequency, the higher the perceived pitch, and the more cycles that occur per unit time. It has been found that the frequency of a wheeze is correlated with the severity of airway obstruction. The higher the pitch of the wheeze, the lower the airflow, and the greater the severity of the asthma (Baughman and Loudon, 1982, p 96).

A large percentage of normal lung sounds are composed of frequencies between 16 and 200 Hz. However, the human ear has a poor sensitivity for detecting these low frequency sounds and so a large percentage of normal lung sounds are generally hard to hear (Murray and Nadel, p 571).

The **intensity** of a lung sound is a function of the amplitude of the vibrations or the height of the waveform from crest to trough. Intensity represents our perception of loudness.

Because there is a wide variety of sound intensities (eg. 10^{-11} W/m^2 = intensity of a normal breath sound, and 10^6 W/m^2 = the intensity of a rocket engine, (Van Heuvelen, p 388)), a value called **Intensity level** has been defined. The intensity level uses the logarithmic scale to measure sound intensity and is defined as follows:

Intensity level = $10 \log_{10}$ Intensity/I_o

where $I_o = 10^{-12}$ W/m^2

This formula compares the intensity of the current sound to the intensity of a sound that is the quietest sound audible to the human ear, or the hearing threshold, called I_o (Van Heuvelen, p 388).

The **duration** of a sound is defined by how long it can be heard. The duration of both normal and adventitious lung sounds is related to the duration of inspiration and expiration. An adventitious sound can be described as either discontinuous or continuous as outlined in Sections 2.4.2.1 and 2.4.2.2, depending on its duration. It has been shown that even a pure sine wave has to be present for a certain length of time for it to develop its musical quality. Pure tones require at least 100 msec to be perceived as musical. If a pure sine wave is played for only 10 msec or less, the musical character disappears and the sound is perceived as a click or a snap. Also a sound with a duration much shorter than 100 msec appears less loud and it becomes increasingly difficult to perceive the pitch of the sound (R.A.L.E.).

The **quality** or timbre of a sound is a characteristic that may be used to distinguish the source of a sound. For example, a violin playing a concert A at 440 Hz does not sound like an oboe or a French horn that is playing a note of equal pitch and at the same intensity. The different qualities of a sound are related to its associated frequency spectra. When a musical instrument is sounded, more than one standing wave is created. For example, if a guitar string is plucked, many standing waves are present simultaneously. As the string vibrates at its whole length, it may also vibrate in segments of 1/2's, 1/3's, 1/4's, etc., of its whole length. The pitch of the sound is the frequency of the lowest standing-wave present in the complex wave. The perceived quality of the sound "depends on the number and relative amplitudes of the other standing-wave frequencies that make up the complex sound waves and the way that this spectrum of tones changes with time" (Van Heuvelen, p 386).

3.2 Signal Processing

The topic of signal processing is quite long and this topic alone could be used to fill several books. The topics of signal sampling, Nyquist's Theorem, and the frequency representation will be briefly discussed in this section. The Fast Fourier Transform algorithm and the way it was used in this thesis will also be discussed.

3.2.1 Sampling the Signal

As was stated previously, a microphone is a mechanical device capable of converting vibrational energy into electrical energy. Figure 3.1 represents a continuous time recording, or analog format, because the voltage measured is recorded over a continuum of time. However, in order for a computer to process the signal, it must be sampled and placed into a discrete time representation, or digital format. Therefore, the signal is recorded only at a designated rate. For example, much of the data utilized in this thesis were sampled at a rate of 5 kHz, which means that the voltage value of the microphone was recorded every 1/5,000th of a second, or 5,000 times per second. However, when a signal is sampled, it must be sampled at a rate which is at least twice the rate of the highest frequency that may be encountered during the recording period. This requirement is called the **Nyquist Theorem** and is required so that the signal is not aliased. The reader is referred to (Cohen, pp 30-31) or (Press et al., pp 427-428) for proof of the theorem or for further reading on this subject. As stated previously in Sections 2.4.2.1 and 2.4.2.2, both crackles and wheezes have a maximum frequency of 2,000 Hz. Therefore, a sampling rate of at least 4,000 Hz should be used when sampling lung sounds.

Because data is sampled in discrete time and broken into segments, a windowing function must be applied before a Fourier Transform is performed so that frequency estimates of the discrete signal match the actual frequency components of the continuous signal. For our study, the Hanning window function was used. The Hanning Window has the following formula:

$$w_j = 1/2[\ 1 - \cos((2\Pi j)/(N-1))\ ]$$

Where: N = number of points in the signal segment;
j = ranges from 1 to N
w_j = weights to multiply corresponding signal segment values
(Press et al., p 468)

Each value in the signal segment is multiplied by its corresponding weight value. For example, w_1 will be multiplied by the first value in the signal segment, w_2 will be multiplied by the second value in the signal segment, and so forth. The resulting signal is then used to compute the FT spectrum.

Several other window functions such as the Hamming and Welch window functions were also implemented; however, the lung sound FT spectrum changed minimally. One text which discusses signal processing stated that "... there is effectively no difference between any of these window functions" (Press et al., p 468). For a more detailed discussion of window functions, the reader is referred to (Press et al., pp 466-470) or (Pasterkamp et al., 1993).

3.2.2 Frequency Representation

Any physical process such as a lung sound can be described mathematically in two ways:

1) **Time Domain**: the value of the function, h, varies as time varies. ie., h(t);
2) **Frequency Domain**: the process is specified by giving its amplitude H (usually a complex number which also indicates phase) as a function of frequency, f. ie., H(f), with $-\infty < f < +\infty$.

(Press et al., p 422)

Both h(t) and H(f) are two different representations of the same process. In continuous time, the Fourier Transform (FT) is used to interchange between the two representations by:

$$H(f) = \int_{-\infty}^{+\infty} h(t)\exp(2\Pi ift)dt$$

$$h(t) = \int_{-\infty}^{+\infty} H(f)\exp(-2\Pi ift)df$$

Note: $i = \sqrt{-1}$ (Press et al., p 422)

The frequency domain representation of the time domain signal can be obtained by performing the FT on the signals and then constructing the Fourier Transform spectrum. This has been computed for Figures 3.1 and 3.2, and the results are displayed in Figures 3.3 and 3.4 respectively.

Figure 3.3 FT Spectrum of the Non-Wheeze Sound in Figure 3.1

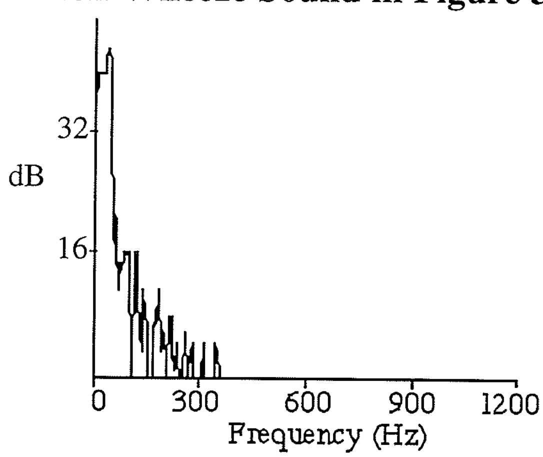
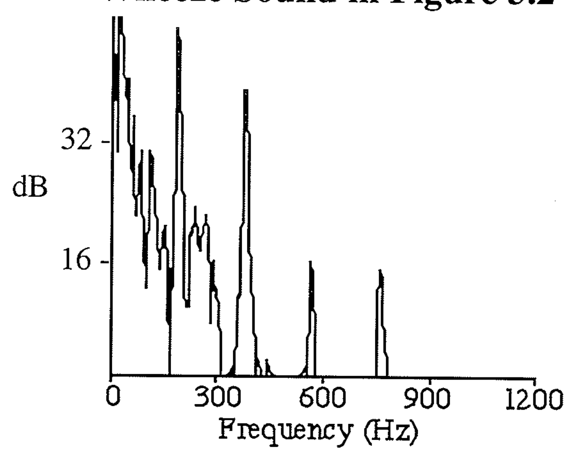


Figure 3.4: FT Spectrum of the Wheeze Sound in Figure 3.2



3.2.3 Fast Fourier Transform Algorithm

The Fast Fourier Transform (FFT) algorithm is a relatively quick method for calculating the frequency representation of a set of sampled data points. The FT representation, H_n of N points can be calculated as follows:

$$H_n = \sum_{k=0}^{N-1} h_k \exp(2\Pi ikn/N)$$

Note: $i = \sqrt{-1}$

(Press et al., p 503)

Similarly, the inverse FT can be calculated from the H_n's computed in the above equation as follows:

$$h_k = 1/N \sum_{k=0}^{N-1} h_k \exp(2\Pi ikn/N)$$

Note: $i = \sqrt{-1}$

(Press et al., p 503)

Computing the inverse FT changes a signal from the frequency representation to the time representation.

Implementing either of above two formulas on a computer can be very time consuming because they execute in time $\Theta(N^2)$* . However, the FFT algorithm utilizes a special way of setting up the points so that the FT can be calculated in time $\Theta(N \log N)$, which is substantially quicker for large values of N (Manber, p 310). For a detailed explanation of how this is accomplished, the reader is referred to Section 9.6 of (Manber).

The FFT algorithm implemented in the research is included in Appendix A. The FFT algorithm returns the real and imaginary number pairs, representing the FT spectrum, in the

* The notation "$\Theta(N^2)$" is read "Big-O of N squared", and denotes that as the number of inputs, N, increases, the time required to process the FT increases by a factor of N^2. For a more detailed discussion of Big-O notation, the reader is referred to (Manber).

array DATA[ ]. The real and imaginary values occupy two consecutive positions in the array.

Once the FFT is calculated, the FT spectrum as depicted in Figure 3.3 and Figure 3.4, is formed. This representation is also commonly referred to as the FT representation. The FT spectrum is calculated by taking the square root of the sum of the squares of the real and imaginary parts, or

$$f_i = \sqrt{((data_i)^2 + (data_{i+1})^2)}$$

Note: $data_i$ = real part of the complex number

$data_{i+1}$ = imaginary part of the complex number

The relative frequency (f_i) for which the power is being calculated depends on the sampling rate and the number of points in the sample. The real and imaginary values at data[1] and data[2] represent the values at 0 Hz. Actually it represents a frequency range from (0 * (half the sampling rate) / (number of points being sampled)) Hz to ((1 * (half the sampling rate)/(the number of points being sampled)) -1) Hz. Similarly, the values at data[3] and data[4] represent the frequency content from (1 * ((half the sampling rate)/(number of points being sampled)) Hz to ((2*(half the sampling rate)/(number of points being sampled)) - 1) Hz.

After the FT spectrum is calculated, the spectrum values are usually linearly scaled so that they range in value between 0.0 and 2.0. This is done so that large FT spectrum values do not saturate the neural network.

3.3 Sonograms

Sonograms are another method of displaying sounds. A sonogram displays time on the horizontal axis and frequency on the vertical axis (R.A.L.E.). A color scheme is then used to indicate the intensity of the frequency at a specified point in time. White indicates the absence of that frequency, and increasingly darker shades of gray indicate an increasing presence of that frequency. Sonograms are very good representations of sounds because they allow the user to compare frequency content of different time segments very easily. The sonogram of a non-wheeze and wheeze are shown in Figures 3.5 and 3.6 respectively.

Figure 3.5: Sonogram of a Non-Wheeze Lung Sound

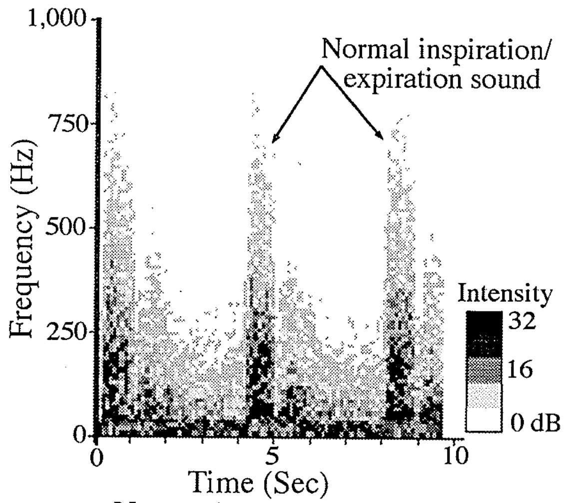
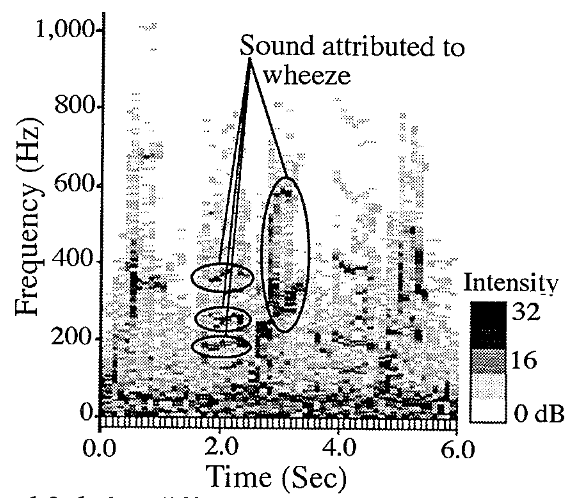


Figure 3.6: Sonogram of a Wheeze Lung Sound



Note: The sonograms in Figures 3.5 and 3.6 also differ in breathing rates.

Sonograms have only been used as an aid to manually classify lung sounds in this thesis.

3.4 Summary

In Chapter 3 the characteristics of sounds including the frequency, intensity, duration, and quality, were presented. These characteristics were then applied to lung sounds. The general principles of signal processing, including signal sampling, were also introduced. The time and frequency representations of lung sounds were then presented and the Fourier Transform algorithm which interchanges between the two representations was also introduced. The sonogram representation was also presented. In subsequent chapters, both the raw signal and the FT representation of the signal will be used in a variety of classification systems.

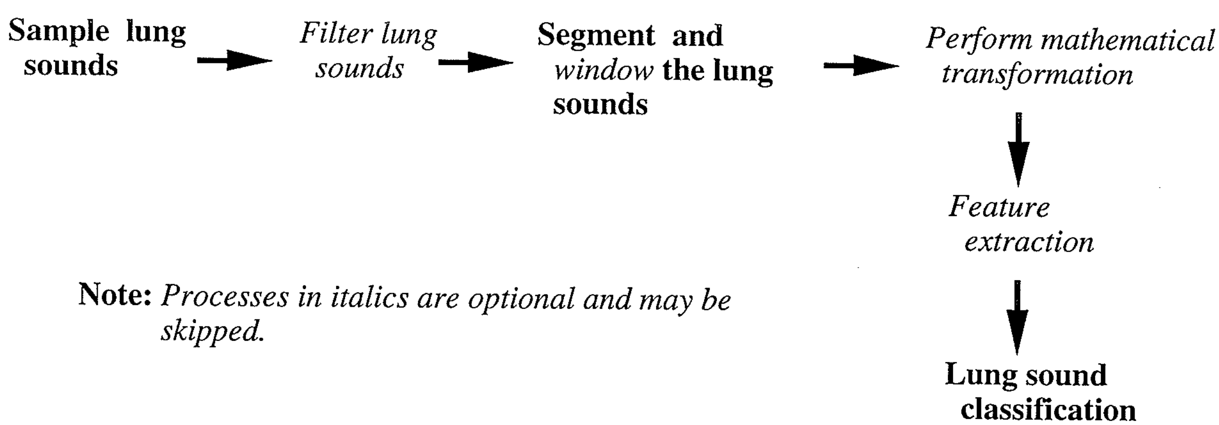
Chapter 4

Lung Sound Analysis and Recording

4.1 Lung Sound Analysis Algorithms

Several authors have presented various methods for detecting adventitious lung sounds. These methods have ranged from algorithmic techniques that utilize both time expanded waveform and FT representations, to methods that use neural networks, (neural networks will be described in Chapter 5). The typical classification scheme used by many authors is outlined in Figure 4.1.

Figure 4.1: Typical Classification Scheme for Lung Sounds



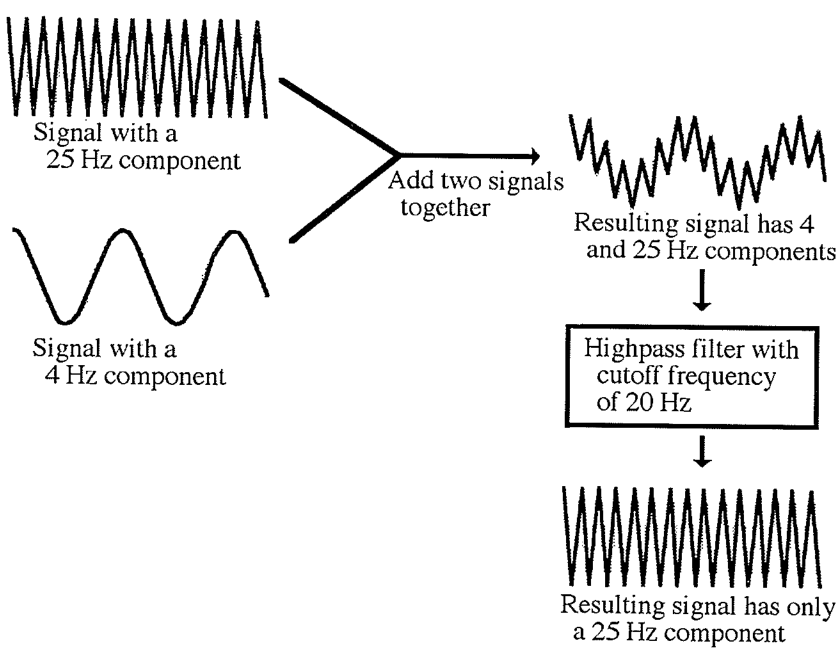
Each process outlined in Figure 4.1 has a variety of ways of being performed. For example, sampling the lung sounds may be done from a variety of anatomical sites, including the trachea or anywhere over the five lobes of the lungs. Sampling may also involve the use of a variety of microphones, microphone attachment methods, and sampling rates. For a more comprehensive review of the equipment, protocols, and analysis methods, the reader is referred to (Mussell, pp 129-139). A comparison of various sensors has also been performed by (Pasterkamp et al., 1993, pp 1518-1525).

4.1.1 Crackle Detection

Computer analysis of crackle sounds has included both detection and classification algorithms using stationary-nonstationary filters and neural networks. As mentioned in Section 2.4.2.1, the presence and type of crackle can be very characteristic of certain diseases. Therefore, computer-aided crackle analysis is an important field.

As mentioned in Section 2.4.2.1, crackles have a frequency range of 200 to 2,000 Hz. The effects of highpass filtering has been studied by (Karp et al., pp 383-384). Highpass filters selectively remove frequency components that are below the cutoff frequency (Oppenheim et al., p 401). For example, suppose a signal is constructed with frequency components 4 Hz and 25 Hz. If a highpass filter with a cutoff frequency of 20 Hz is applied to this signal, the resulting signal will contain only the 25 Hz component. This process is depicted in Figure 4.2. (Note: Some highpass filters may leave a small portion of the lower frequency components in the signal.)

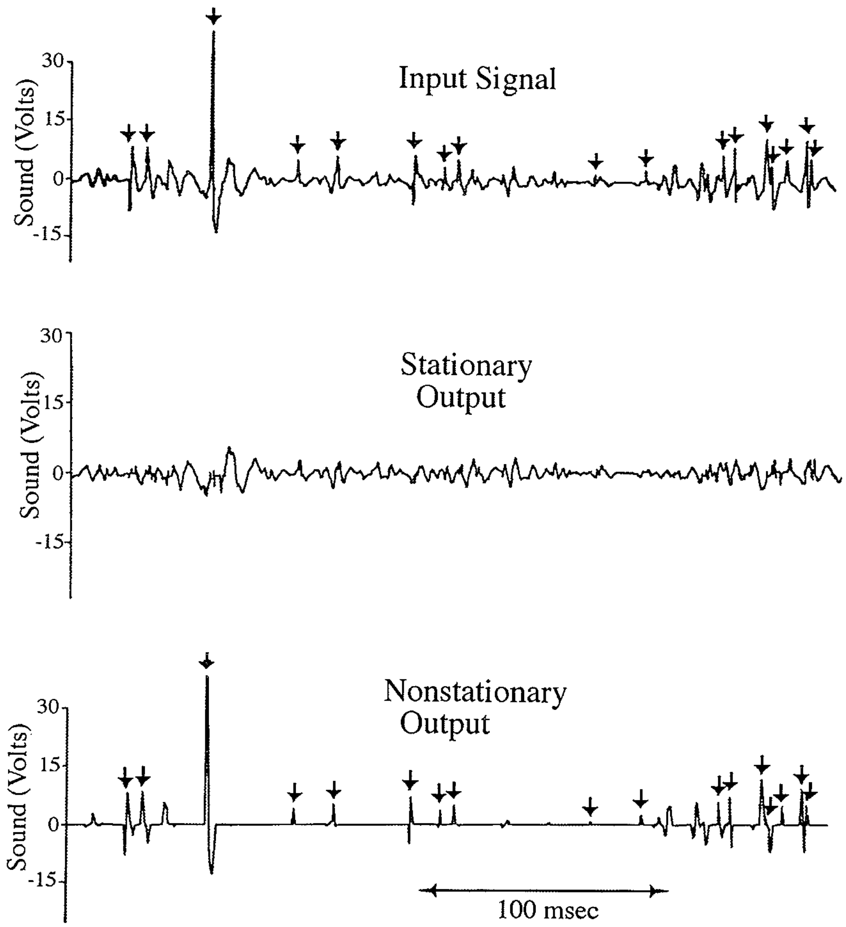
Figure 4.2: Highpass Filtering



A study presented by (Karp et al., pp 383-384) highpass filtered both analog and digital signals at 50, 100, 200, and 400 Hz. It was found that filtering increased the number of visible deflections but decreased their amplitudes. In addition, filtering the digital signal at 200 and 400 Hz inverted the polarity of the initial deflection while filtering the analog signal conserved the polarity. Since many authors first filter the lung sounds to remove unwanted noise, the effects of filtering must be taken into account (Xu et al., 1989, p 1676).

One method to detect crackles within vesicular breath sounds utilizes a nonlinear digital filter which was designed to separate nonstationary from stationary signals (Ono et al., pp 286-291). Vesicular breath sounds were found to make up most of the stationary signals, while crackles were found to form most of the nonstationary signals. This process is demonstrated in Figure 4.3.

Figure 4.3: Stationary and Nonstationary Crackle Signals



Note: The arrowheads (↓) indicate those waves which were visually identified as crackles.

(Figure 4.3 adapted from Ono et al., p 289.)

The process of stationary-nonstationary filtering was reported to have worked relatively well for crackle detection, although numeric results were not presented. This detection method was also found to identify the beginning of inspiration, heart sounds, and short wheezes in the nonstationary output (Ono et al., pp 286-291).

A feedforward network has also been implemented to detect crackles in (Kalayci et al., pp 2580-2581). The neural network consisted of 128 input units, 16 hidden units, and 6 output units. The network was reported to be successful at classifying and detecting crackle sounds, although results were not quantified.

4.1.2 Wheeze Detection

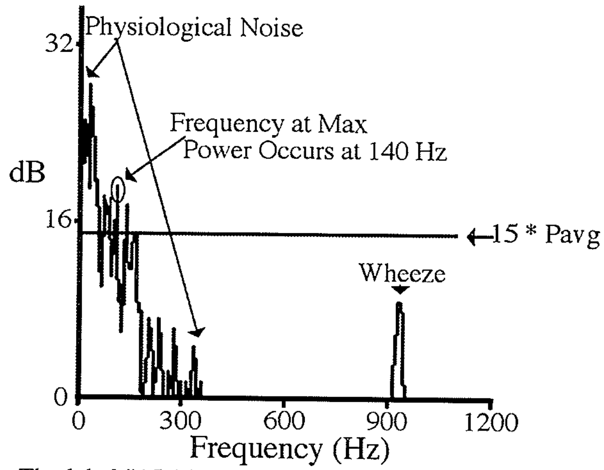
Computerized wheeze detection has been performed by algorithmic and Self Organizing Map (SOM) techniques. Two different algorithmic techniques have been implemented and tested with the FT representation of the data sets of Section 7.1. The SOM technique could not be tested because the SOM processing was not fully explained. A description of SOM's is provided in Section 5.5.2.1 and the results of testing a different type of SOM are presented in Chapter 7.

4.1.2.1 Simple Wheeze Detection Algorithm

One method of detecting wheezes over the lungs is to determine the frequency of maximum power (f_p) between 100 Hz and 1300 Hz for the lung sounds. "A value of f_p greater than 200 Hz may be considered as abnormal, and if present, is generally adequate to identify wheezing in lung sounds" (Fenton et al., Jan. 1985, p 51). A value less than or equal to 200 Hz indicates that no wheeze is present. Lung sounds that contain a wheeze usually have a higher mean peak frequency than lung sounds that do not contain a wheeze (Pasterkamp et al., 1983, vol. 127, p 217).

This simple algorithm for wheeze detection was implemented and tested on the FT representation of the data set of Section 7.1. (Note: The FT was calculated as outlined in Section 3.2.3, except no normalization between 0.0 and 2.0 was performed). The algorithm had an overall performance of 73% on the data sets Sounds_1.A and Sounds_1.B. However, the performance with the data set Sounds_2 was less than 60%. After reviewing the data sets it was observed that many of wheezes that had been misclassified had frequencies that were of smaller amplitude than the physiological noise between 100 Hz and 200 Hz. A typical wheeze that was misclassified is displayed in Figure 4.4.

Figure 4.4: An Example of an Incorrectly Classified Wheeze

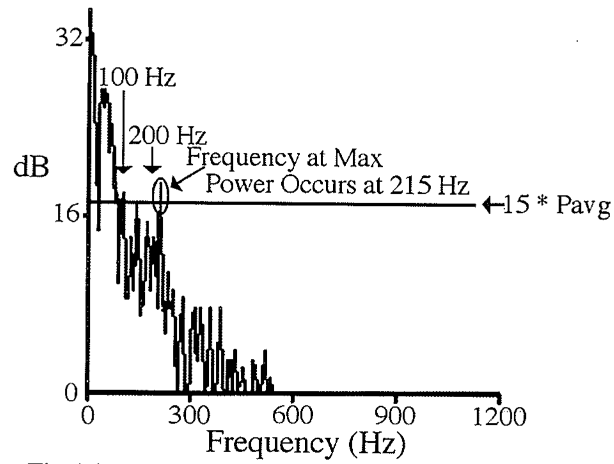


Note: The label "15 * Pavg" will be explained in Section 4.1.2.2.

Figure 4.4 depicts a wheeze that is of lower intensity than the frequency of maximum power. The frequency of maximum power above 100 Hz occurs at 140 Hz and is due to physiological noise. By the simple algorithmic technique, this segment is incorrectly classified as non-wheeze. Some wheezes were also between 100 Hz and 200 Hz and will be misclassified.

Some misclassified normal lung sounds contained the peak power just above 200 Hz which was probably attributable to physiological noise. An example of one of the misclassified non-wheezes is displayed in Figure 4.5.

Figure 4.5: An Example of an Incorrectly Classified Non-Wheeze



Note: The label "15 * Pavg" will be explained in Section 4.1.2.2.

In Figure 4.5, the frequency of maximum power occurs at 215 Hz and is attributable to physiological noise. This segment is incorrectly classified as containing a wheeze.

From the results, this approach is insufficient for detecting wheezes within the three data sets.

4.1.2.2 Amplitude Qualification Wheeze Detection Algorithm

Another more sophisticated method for wheeze analysis of tracheal sounds has been proposed in (Fenton et al., pp 50-55). Unlike the chest, sounds recorded at the trachea contain many high frequency components because the chest functions as a lowpass filter and attenuates high frequency sounds (Forgacs, Lung Sounds, p 41). Therefore, a form of "amplitude qualification" is required to take into account the higher frequency of tracheal sounds.

The amplitude qualification method calculates the average power (P_{avg}) between 110 and 1,200 Hz from the FT data. Any frequency that is more than A_f times P_{avg} is classified as being a wheeze. (Note: An A_f value of 15 was used by Fenton et al. because tracheal sounds were analyzed.) The lower range of 110 Hz was used to exclude noise of the muscles and heart. No lung sounds in Fenton et al.'s study were above 1,000 Hz so restricting processing to 1,200 Hz does not cause a problem. This method was reported to have false positive[1] and false negative values of less than 2% when analyzing wheezes of medium to loud intensity at the trachea.

The amplitude qualification algorithm was implemented and tested on the data of Section 7.1. Fenton et al. originally used the algorithm only on sounds recorded at the trachea; however, the data sets of Section 7.1, were recorded from the anterior and posterior sides of the thorax region. Initial results using $A_f = 15$ were poor so different values of A_f were tried. The FT spectrum was calculated as outlined in Section 3.2.3, (except normalization of the values between 0.0 and 2.0 was not performed). The results for Sounds_1.A and Sounds_1.B are displayed in Table 4.2 and the results for Sounds_2 are displayed in Table 4.2. A description of contents of the data files Sounds_1 and Sounds_2 is provided in Section 7.1

Table 4.1: Amplitude Qualification Results on Data Sets Sounds_1.A and Sounds_1.B

A_f Values	Number of Correct Non-Wheezes	Number of Correct Wheezes	Specificity* (Non-Wheeze Performance)	Sensitivity* (Wheeze Performance)	Overall Performance*
2	0	235	0%	100%	50%
3	30	235	13%	100%	56%
4	80	222	33%	97%	65%
6	199	206	83%	88%	85%
7	228	194	95%	83%	89%
8	233	181	97%	77%	87%
10	240	137	100%	58%	79%
12	240	101	100%	43%	72%
14	240	77	100%	33%	68%
16	240	55	100%	23%	62%
18	240	41	100%	17%	59%

* Refer to Appendix B for an explanation of how specificity, sensitivity, and overall performance were calculated.

[1] Refer to Appendix B for a definition of false positive and false negative results.

Table 4.2: Amplitude Qualification Results on Data Set Sounds_2

A_f Values	Number of Correct Non-Wheezes	Number of Correct Wheezes	Specificity	Sensitivity	Overall Performance
2	0	116	0%	100%	55%
3	12	111	13%	96%	58%
4	35	69	37%	60%	49%
6	73	53	77%	46%	60%
7	83	47	87%	41%	62%
8	92	34	96%	29%	60%
10	95	12	100%	10%	51%
12	95	1	100%	1%	45%
14	95	0	100%	0%	45%
16	95	0	100%	0%	45%
18	95	0	100%	0%	45%

The results displayed in Table 4.1 indicate that the A_f value that maximizes overall performance is $A_f = 7$ for the data set Sounds_1.A and Sounds_1.B. This provides a specificity of 95% and a sensitivity of 83% and an overall performance of 89%. For data set Sounds_2, the optimal A_f value was also 7; however, specificity was 87% and sensitivity was only 47%. This result can be explained by the fact that Sounds_2 contained wheezes that were of lower intensity than the wheezes in Sounds_1.A and Sounds_1.B. The lower intensity wheezes will have smaller spikes in the corresponding FT spectrum, and so this method will have a worse performance when detecting wheezes. This finding was consistent with our results.

In Figures 4.4 and 4.5, an A_f value of 15 is represented as a horizontal line on the FT spectrum. Anything above the line, and with a frequency greater than 110 Hz, indicates that the algorithm would classify the segment as containing a wheeze. The lung sounds of Figure 4.4 will be correctly classified as containing a wheeze; however, not because of the wheeze that occurs at 920 Hz, but rather because of the physiological noise. If a lowpass filter is used to remove the wheeze part of the signal, the signal would still be classified as a wheeze. Figure 4.5 would be incorrectly classified as a wheeze with this algorithm because the spike at 215 Hz, which is attributable to physiological noise, is more than 15 times the average frequency.

The results displayed in Table 4.1 are not as good as the results obtained from (Fenton et al., pp 50-55) who analyzed lung sounds recorded at the trachea and obtained specificity and sensitivity levels better than 98% for both non-wheeze and wheeze data. One reason that the results may not have been as good is that the lung sounds at the trachea are generally of higher pitch than the lung sounds in the thorax region (Fenton et al., p 51), and so they may appear louder because they are not covered up by physiological noise or attenuated by the chest wall. This may allow the wheezes at the trachea to have larger spikes in the FT spectrum.

It may be possible to improve performance of the algorithm by using different filtering or post-processing techniques such as thresholding or consensus diagnosis. These methods will be further discussed in Chapter 7.

4.1.2.3 SOM Wheeze Detection Algorithm

Self-Organizing Maps (SOM) have also been applied to lung sound analysis. However, the research was presented as an extended abstract so only preliminary findings were stated. The SOM method presented by (Kallio et al.) used 16 frequency components derived from a 256 point FT spectrum of the lung sounds. No information was provided regarding how the 16 frequency components were acquired. The only mention of the SOM's performance was "... [the SOM] follows natural lung sound spectra saving their features and it can be used for automatic classification of various lung sound types" (Kallio et al.).

An SOM was also presented to diagnose patients with various diseases by the same research group (Rajala et al.). The training set for the SOM consisted of twenty-eight patients having crackles, four patients having wheezes, and six healthy patients. The patients with crackles were diagnosed with either fibrosing alveolitis, bronchiectasis, Chronic Obstructive Pulmonary Disease (COPD), or heart failure. The system was then tested with six patients who had crackles (and the diagnosis of either fibrosing alveolitis, bronchiectasis, or COPD), and one healthy patient. The patients in the test set were not the same as in the training set. The results are displayed in Table 4.3.

Table 4.3: SOM to Diagnose Patients with Various Diseases By Lung Sound Analysis

Type of Disease	Overall Performance
Healthy patient (no abnormalities): expiration	100%
Healthy patient (no abnormalities): inspiration	60%
Fibrosing Alveolitis	60%
Bronchiectasis	60%
Chronic Obstructive Pulmonary Disease	30%

(Table 4.3 adapted from Rajala et al.)

With the exception of the healthy patient during expiration, the results proved to be poor. In the case of COPD, the system had a worse performance than if it had just guessed at the diagnosis; in which case we would expect a level of performance of about 50%.

4.2 Lung Sound Recording

All of the lung sounds utilized in this thesis were acquired at the Respiratory Acoustics Laboratory at the Health Sciences Center in Winnipeg, Manitoba. All of the lung sounds were reused from previously performed experiments at the Respiratory Acoustics Laboratory. All subjects were between the ages of 5 and 30 years and most subjects were male.

The lung sounds were taken from any one of the locations numbered one through seven shown in Figures 4.6 and 4.7. Tracheal sounds were not analyzed because they usually have a relatively large amount of power in the higher frequency spectrum and so are different from sounds recorded at the chest or back (Fenton et al., 1985, p 51).

Figure 4.6: Anterior View of the Lungs

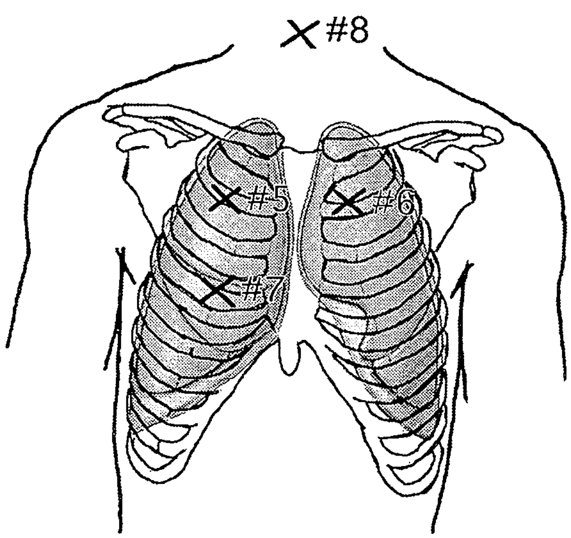
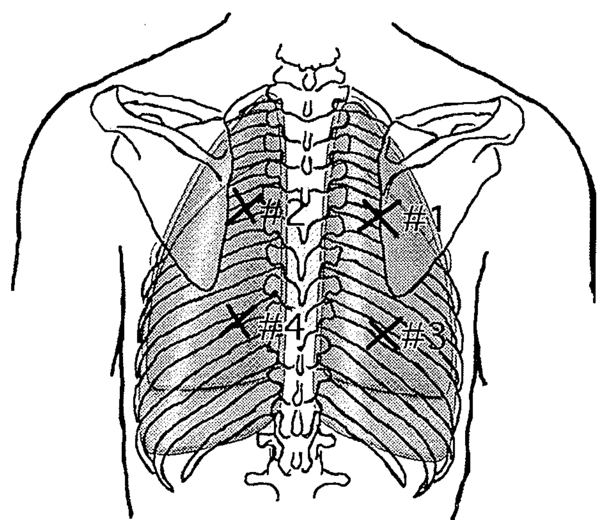


Figure 4.7: Posterior View of the Lungs



(Figures 4.6 and 4.7 adapted from Moore, p 62.)

The lung sounds were recorded with sensitive piezoelectric accelerometers called Siemens EMT25C, which had custom-built signal conditioners. As used, they have an output of 12 V/m/sec^2 with a flat response (to acceleration, not displacement) from 20 to 900 Hz. After 1,200 Hz the EMT25C exhibits a roll off effect. The microphones were attached to the skin with double-sided adhesive rings (#2181 from 3M- 1.25" with 11/32" aperture). Each sensor was calibrated by feeding white-noise from a speaker onto a rigid metal plate with the EMT25C attached. A reference sensor was also attached, and the outputs were compared so that they were within a couple of decibels of each other. Each EMT25C weighs 15 grams. No shielding was utilized for the accelerometers.

The sound signals were then sampled with an analog-to-digital converter (model DT2801A, DataTranslation) and transferred to an IBM-compatible personal computer (80386 microprocessor and 80387 mathematics coprocessor, 20 MHz clock speed), using a sampling rate of 10,024 Hz per channel. In some experiments data were sampled at 5,120 Hz, and so all data were down sampled to 5,120 Hz (by taking an average of every two consecutive data points), before the data were processed by the classifier. The data were stored using a customized computer program (R.A.L.E.: Respiration Acoustics Laboratory Environment).

In addition to lung sounds, a signal from a calibrated pneumotachograph or nasal flow thermistor was recorded as a representation of airflow in and out of the lungs. A multi-lead (usually three) EKG was also recorded in some experiments.

Lung sound data were classified by playing the lung sounds back and using the time expanded waveform, sonogram, and FT representations to help classify the signal. Lung sounds were played back several times and several breathing cycles were compared. All data were classified by the author of the thesis after receiving a course on lung sound classification. All lung sounds which could not be classified by a high degree of certainty were excluded from the study.

4.3 Uses of an Automated Wheeze Classifier

There are several uses of the automated wheeze classifiers implemented in this thesis. The author has listed some of the uses in the following section. Although not an exhaustive list, it should give the reader an indication of the importance for obtaining an algorithm that classifies wheezes with a high degree of accuracy.

An automated wheeze classifier can be used in making the diagnosis of asthma. Asthma has been defined as:

> ... a disease characterized by an increased responsiveness of the trachea and bronchi to various stimuli and manifested by a widespread narrowing of the airways that changes in severity either spontaneously or as a result of therapy.
>
> (American Thoracic Society, 1962, p 763)

Asthma affects approximately 5% of adults and 7 to 10% of children (Wilson et al., p 1047). The point prevalence (prevalence of active disease) of asthma "... reaches a peak in childhood, declines during adolescence, and increases again during middle age" (Dodge and Burrows, p 573). Patients with asthma may suffer from asthmatic attacks which may involve the following manifestations:

1) Wheezing, coughing, or dyspnea;
2) Over inflation of the chest;
3) Decreased lung capacity and increased airway resistance which is alleviated by the administration of a bronchodialator.

(American Thoracic Society, 1962, pp 763-764)

Of the three manifestations, wheezing is the symptom of interest in this research. There have been diagnostic descriptors that population studies have used to diagnose asthma. In some studies, the diagnosis is based on the subject's report of wheezing (McNickol and William, pp 7-11). Other studies may require that wheezing be provoked by a certain stimuli (Smith and Knowler, pp 16-38). However, one drawback with using a wheeze as the only diagnostic descriptor is that approximately 30% of the population wheezes, but only one in seven of these people are found to have asthma (Dodge and Burrows, p 569). Some studies may require that the patient has seen a physician for problems related to asthma

(Dodge and Burrows, pp 568). There are many ways that a physician may diagnose asthma.

One method used by physicians to diagnose asthma is by use of a Methacholine Challenge (MCH). During a MCH, the patient breathes a small dose of methacholine which may cause the maximum volume of air that can be expelled in one second (FEV_1) to decrease by 20% or more if the patient has asthma (Hopp et al., p 157). This decrease in FEV_1 is often associated with the production of a wheeze. During experiments with the MCH, microphones are placed on the patient's body to detect wheeze sounds and lung function is assessed by spirometry (see Figures 4.6 and 4.7 for microphone placement). A study of children by (Sanchez et al., 1993; vol. 15, pp 28-35) found that wheezing during a MCH had a sensitivity of 68% and a specificity of 82%[1]. It was concluded that "...wheezing during a MCH strongly suggests airway hyperresponsiveness; however, wheeze detection can not fully replace spirometry" (Sanchez et al., 1993; vol. 15, p 28).

Another experiment was also performed by (Sanchez et al., 1993; vol. 147, pp 705-709) that investigated wheezing during a MCH of normal children and children suffering from cystic fibrosis. Both acoustic and spirometric assessment were performed with the later assessment taken as the true assessment. This experiment found that the sensitivity of a wheeze as an indicator of bronchial hyperactivity was only 50%, indicating that only half the subjects with a positive spirometric response (FEV_1 reduced by at least 20%) wheezed. However, the specificity was 100%, indicating that no normal subjects wheezed.

Although the sensitivity of wheezing is low as compared to spirometric assessment, in instances where the patient is unable to provide a spirometric assessment, as in the case of infants or extremely ill patients, the appearance of a wheeze may be the most diagnostic sign

[1] Sensitivity and specificity have been defined in Appendix B.

of disease. For example, the response to treatment for infants suffering from airflow obstruction due to bronchiolitis, infantile asthma, bronchopulmonary dysplasia, or cystic fibrosis is difficult to make in an objective manner. The assessment of response to treatment can not be done spirometrically, and so is based on many manifestations, including wheeze severity. An objective method for measuring wheeze severity would allow the evaluation of treatment to be more objective.

A study by (Asher et al., pp 1405-1410), investigating the use of computerized lung sound analysis as a method to evaluate treatment effectiveness of children with bronchiolitis found the method was a "good non-invasive indicator of treatment response." However they also pointed out that infants have different wheeze patterns than adults and the lack of objective acoustical standards of wheezes forces the physician to make "generalized classifications." A study by (Baughman and Loudon, Nov. 1984, pp 718-722) found that there was a very good correlation between the decrease in the amount of time spent wheezing and the increase in FEV_1.

Another benefit of the technique of computerized analysis of lung sounds is that it is noninvasive and can be performed without disturbing the patient's physiological state. This property can be very important to the diagnosis of nocturnal asthma where it is very important to assess the severity of asthma, and the stage of REM sleep without disturbing the patient. The computerized analysis method was used by (Baughman and Loudon, Sept. 1985, pp 364-368) to assess the severity of the asthma and correlate it with the patient's stage of REM sleep. They found that the technique was useful for continuous, noninvasive monitoring of wheezy patients.

Finally, a computerized lung sound analysis technique can be an effective tool for monitoring patients in critical care facilities. The monitoring device can free up medical personnel from the arduous task of continually auscultating a patient to evaluate pulmonary function.

4.4 Summary

This chapter has outlined some of the previous methods for crackle and wheeze detection. Two of the wheeze detection algorithms were also implemented and tested with the same data that is used in Chapter 7. The best results of the algorithmic techniques are displayed in Table 4.4.

Table 4.4: Best Results of the Algorithmic Techniques

Algorithm	Overall Performance with Sounds_1.A and Sounds_1.B	Overall Performance with Sounds_2
Simple Algorithmic Technique	73%	< 60%
Amplitude Qualification ($A_f = 7$)	89%	62%

From Table 4.4, it can be seen that the overall performances of the two algorithmic techniques degraded when tested with the data set Sounds_2. Sounds_2 contained wheezes of low to medium intensity. These lung sounds have a slightly smaller signal to noise ratio and therefore will be generally harder to classify. From the results summarized in Table 4.4, it appears that the two algorithmic techniques have difficulty classifying wheezes that are of low to medium intensity.

Chapter 4 also discussed how lung sounds were recorded and categorized for use in the thesis. The chapter ended with a brief description of some of the uses of a non-invasive computerized wheeze classification system which include: 1) a wheeze detection system for a methacholine challenge test; 2) a system capable of evaluating severity of airway narrowing or response to treatment in subjects unable to perform a conventional spirometric assessment; 3) a system capable of evaluating severity of airway narrowing without disturbing a subject's physiological state; and 4) a continuous monitoring system for patients in critical care facilities.

Chapter 5

Neural Networks

5.1 A Brief Introduction

Neural networks are powerful tools of Artificial Intelligence (AI) that are capable of performing pattern recognition tasks quite well. Stated in simple terms, a neural network is:

> ...simply a collection of processing units ... each of which is connected to many other [processing units]. All of the knowledge of the network is in the strength of the connections between the processors. You get a network to perform a specific task by showing it lots of training examples.
>
> (Horenblas)

This chapter will briefly outline the structure and biological basis of neural networks. It will also examine five types of neural networks--Backpropagation, Self Organizing Maps, Learning Vector Quantization, Probabilistic Networks, and Radial Basis Functions--that were used in the research. Finally, it will end with a brief summary of other successful applications of neural networks. A more comprehensive review of the field of neural networks can be found in several texts such as those by (Hertz et al.), (Wasserman, 1993), (Wasserman, 1989), and (Rumelhart et al.).

5.2 A Comparison With Other Techniques.

Neural networks differ from conventional programming techniques and traditional AI techniques in several ways. Conventional programming languages such as Pascal and Fortran implement an algorithm that is already known. Traditional AI tools, such as expert or rule-based systems, use explicitly represented knowledge and carefully designed search algorithms to find a solution (Luger, p 21). However, neural networks use a set of data called the training set from which the network learns relationships. After being trained with the training set on several example relationships, the neural network can interpolate or extrapolate a new relationship that is consistent with the training set (Hertz et al., p 9).

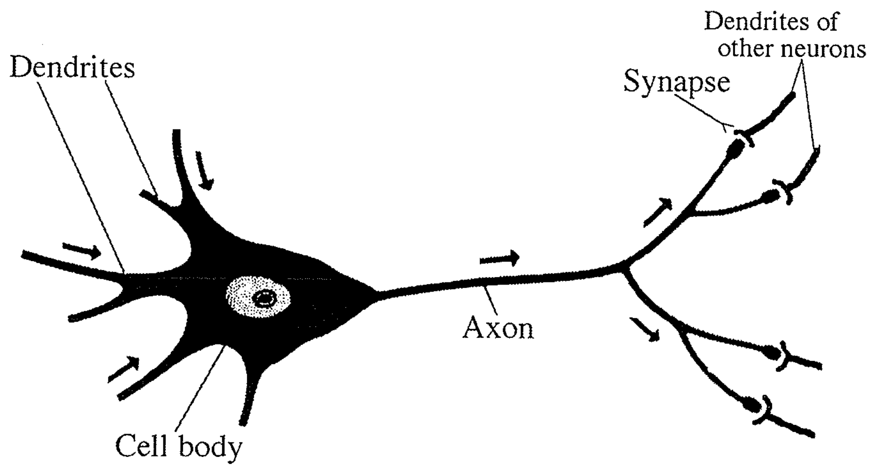
There are several advantages to neural networks in comparison to algorithmic and rule-based systems. Some of them are:

> 1) Neural networks do not require any "hard-coded" rules or algorithms and therefore can be applied to problems that are not rigidly defined, such as lung sound classification;
>
> 2) Neural networks are fault tolerant and can also work well in environments in which the data contain noise because the knowledge is distributed somewhat uniformly around the network; and
>
> 3) Once trained, some types of neural networks are faster than algorithmic or rule-based systems. Research is ongoing into hardware implementations that can provide faster processing for time-critical applications.
>
> (Luger, p 22)

5.3 Biological Neural Networks

The field of artificial neural networks, also known as Parallel Distributed Processors (PDP), or more simply neural networks, has its origin in the work of W. S. McCulloch and W. A. Pitts (1943) with the publication of *A Logical Calculus of the Ideas Immanent in Nervous Activity*. In it, they proposed that the activity of biological neural networks, such as found in the brain, could be modeled by using logical networks of simplified artificial neurons or processing units. This was the basis for the field of neural networks (McCulloch and Pitts). Figure 5.1 is a simple schematic of a biological neuron.

Figure 5.1: Biological Neuron



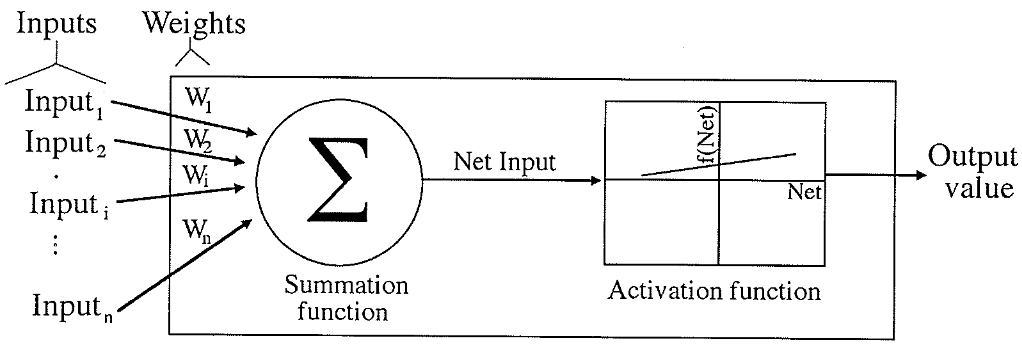
(Figure 5.1 adapted from Shapiro, p 1017.)

The biological neuron has a number of branched protrusions, called dendrites, that bring information to the cell body of the neuron. A single branch, called the axon, exits the cell body, and carries information away from the neuron. The axon may later divide to form other axon branches. Dendrites allow the biological neuron to receive signals from other neurons. If the activity of all signals reaching the cell body exceeds a threshold, the neuron will be allowed to "fire" and an impulse or "spike" will be propagated down the neuron's axon. Synapses are points of contact between the axons of one neuron and the dendrites of another neuron. Synapses are either excitatory, which adds to the total signal reaching the neuron, or inhibitory, which subtracts from the total signal reaching the neuron (Luger, p 22). Many biological neurons may be interconnected in this manner. This interconnection allows an organism to control other parts of its body or perform complex functions. The above description of a biological neuron is rather simple but is similar to the process that is modeled in an artificial neuron.

5.4 Artificial Neural Networks

Artificial neural networks are created by interconnecting a large number of processing units by weighted connections. The weights of the connections determine the strength of the relationship between adjacent processing units. A typical processing unit is shown in Figure 5.2.

Figure 5.2: Typical Processing Unit



Each processing unit consists of:

Input values, $Input_i$. These data may be the actual input values to the neural network, or the outputs of other processing units. Different neural networks may have different ranges of input values. However, the values are usually discrete values from the set {0,1} or {-1, 1} or real valued numbers between this range.

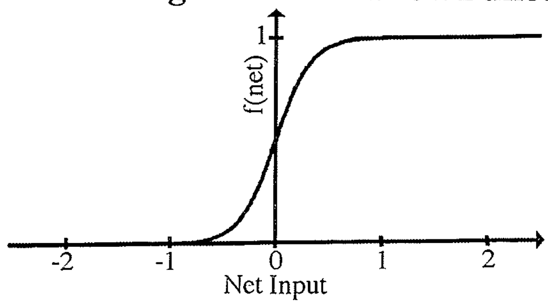
A set of real valued weights, w_i. The weights determine the strength of the relationship between connected processing units.

A Summation Function, Σ. This function computes a weighted sum called the **Net input**, which is based on the connection weights and the inputs.

Activation Function. This function computes a function of the **Net input**. As the net input increases, the **Output value** of the unit increases towards 1. As the Net input decreases, the output value of the unit decreases towards 0. Different neural networks may have different activation functions. For example, the activation function depicted in Figure 5.2 is called the linear activation function because as Net input increases, the output value increases in a linear relation. The sigmoid activation function is displayed in Figure 5.3.

(Luger, pp 518-519)

Figure 5.3: Sigmoid Activation Function



The sigmoid activation function is different from the linear activation function because for the sigmoid activation function, the output value may have a relatively large change in value from 0 to 1 when the Net input is close to 0. Outside of this range, large changes in the Net input have little effect on the output value (Hertz et al., p 27). For the linear activation function, the output value changes gradually even when the net input was not close to 0.

In addition to the properties possessed by individual neurons, a neural network is also characterized by global properties such as:

> **Network Architecture**. The way individual processing units are arranged and interconnected.
>
> **Learning Algorithm**. The algorithm that is used to learn the relationships of a training set.
>
> **Environment of the Neural Network.** This includes the relationships of the inputs and outputs and any interpretation of the input or output data.
> (Luger, pp 518-519)

Neural network processing generally consists of two phases, a training phase and a test phase. There are two types of training, supervised and unsupervised training. During **supervised training**, the network is presented with a set of input patterns and the associated set of desired output patterns. The inputs typically represent a state of a system such as the lung sounds being produced. In addition, the input values are usually normalized to have values within a small range such as {0.0, 2.0} to avoid saturating the network with large input values. Output values may represent a categorization of input values, as for example, the outputs used in the thesis represent the categorization of lung sounds. There must exist a relationship between all input and output patterns. It is the relationship that the neural network "learns" and then uses to classify new data in the test set.

Unsupervised training is a bit different from supervised training because during training, the network is not given a set of associated output values as in the case of supervised training. Instead, the network "learns" to group similar patterns together.

Both types of training involve presenting the network with a set of training patterns. The network uses the layers of processing units and weighted connections to determine the relationships of the training data. The weighted connections are successively altered by the learning algorithm so that the relationships can be learnt. Initially, the connections have random weights. However, as the network is trained and gains experience, the weighted connections are altered and store the knowledge acquired through training. Each cycle that the neural network is trained with the complete set of training examples is called an **epoch**.

The majority of the neural networks developed in this study were created using the NeuralWorks Professional II package (NeuralWare); however some of the initial work was carried out using the Desire/Neunet system (Korn). The networks were implemented on an IBM-compatible personal computer with a 80486 microprocessor with a clock speed of 33 MHz and a 80387 mathematics coprocessor.

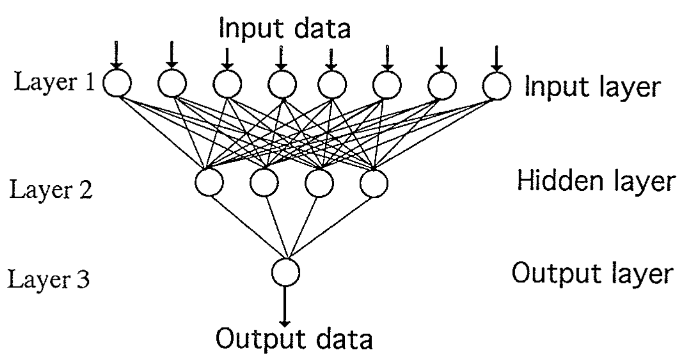
5.5 Types of Neural Networks

There are several types of neural networks, each with its own unique way of altering the connection weights and thresholds. Some of the models used in the thesis will be discussed in the following sections.

5.5.1 Backpropagation

Backpropagation (BP) neural networks are multi-layer feed-forward networks with an input layer, one or more hidden layers, and an output layer. The architecture of a BP neural network is presented in Figure 5.4.

Figure 5.4: Typical Architecture of a BP Neural Network



One of the problems with BP networks is that they are not guaranteed to learn the input patterns perfectly. However, BP networks exhibit interpolative behavior[1] which allows them to generalize and produce the correct output data for input patterns for which they were not trained.

The processing of an individual processing unit in a BP network is relatively straight forward. The **Net input** of a processing unit is calculated by:

[1] **Interpolative behavior** involves creating a new output (different from any output pattern in the training set), for a new pattern by generalizing from the training set. It often involves learning a set of "rules" that helps classify data within a given environment. Conversely, probabilistic neural networks involve **acretive behavior** that can not generalize new relationships. The network will respond with the same output pattern that is associated with the new pattern that most closely matches a pattern in the training set.

Net input = $\sum_i$ input$_i$ * w_i

where input$_i$ = Inputs to the processing unit.
w_i = Connection weights of the inputs.

The activation function used in BP is the sigmoid function and it generates the **Output value, O**, of the processing unit. The values of this function were displayed in Figure 5.3. The sigmoid function is defined as:

$$\mathbf{O} = \frac{1}{1 + \exp(-\textbf{Net input})}$$

The error in an individual processing unit is calculated as the difference between the calculated output O_a, and the correct or desired output value O_c. When the network is learning, the connection weights (w_i) are altered so that the error is minimized. This is accomplished by the formula:

$$w_i(\text{new}) = w_i(\text{old}) + (O_c - O_a) * \text{Input}_i * \varepsilon$$

Note: ε = The learning rate which is gradually reduced as learning progresses.

This formula requires that the desired output be known for each processing unit. However, only the output layer's desired output values are known during training. To overcome this problem, the BP network approximates the amount of error in the hidden and input layers by using the errors of the units to which each hidden unit is connected.

For a unit in layer n (where n is not the output layer), the unit's error is the weighted sum of the errors of all units that use the unit's output. The error in the weights connecting a unit in layer n to all units in layer n+1 is calculated by:

> Error = f'(**Net Input**) * $\sum$w (For all connections between a processing unit in layer n, and any other processing unit in layer n+1) * Error(previously calculated for each of the connected units in layer n+1).
>
> (Hertz et al., Chapter 6)

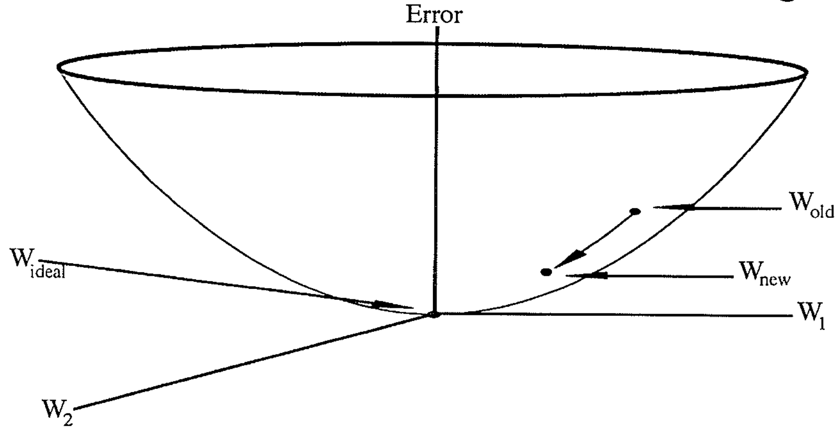
The neural network's error for an individual pattern is commonly measured as the sum of squares over all values in the output layer by using the following equation:

$$\text{Total Error} = \sum_{\text{all outputs}} (O_c - O_a)^2$$

One measure of how well a neural network is learning the patterns in the training set during the learning phase is the Root Mean Square error (RMS). To calculate the RMS, sum the squares of the error for each unit in the output layer, divide by the number of units in the output layer, and then take the square root of the resulting value (NeuralWare, p UN-42). The lower the RMS value, the better the network has learned the patterns. A RMS value of less than 0.1 is usually a good indication that the training patterns have been learnt by the network.

If a plot is made of the error vs. the possible connection weights, a bowl-like surface is obtained, with the bottom corresponding to the ideal weight assignments (Luger, p 522). This bowl-like surface is displayed in Figure 5.5.

Figure 5.5: Delta Rule Error vs. Connection Weights



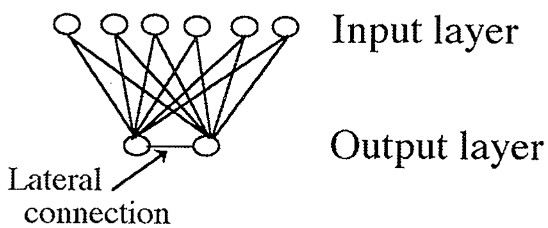
(Figure 5.5 adapted from Luger, p 523.)

For a more in-depth study of BP neural networks, the reader is referred to (Hertz et al., Chapter 6).

5.5.2 Competitive Neural Networks

The Self-Organizing Map (SOM) and Learning Vector Quantization (LVQ) neural networks are both examples of competitive neural networks. Competitive neural networks are usually composed of two layers, an input layer and an output or processing layer. The network is fully connected between layers and may also have lateral inhibitory connections between units in the output layer as shown in Figure 5.6.

Figure 5.6: Example of a Competitive Neural Network



However, unlike backpropagation, only one processing unit in the output layer is permitted to fire when processing an input pattern. This type of processing is often referred to as "winner-take-all processing" (Hertz et al., p 217).

Competitive networks often use unsupervised learning. Before any learning can occur, the number of output units, which equals the number of classification categories, must be defined and all connection weights must be randomized. The input patterns are normalized so that they have unit length by using the following formula:

$\text{Input}_j = \text{Input}_j / (\sum_i \text{Input}_i^2)^{1/2}$

Note: Input_i varies to include all possible input values.

The connection weights are also normalized in a similar manner. Each pattern is then processed by the network and the net input is calculated the same way as in backpropagation.

The processing unit in the output layer with the connection weights that most closely match the current input pattern wins the competition and is allowed to "fire". All other output units have an output value of zero. The metric used to measure distance between the input pattern, $Input_j$ and the weight vector associated with a processing unit, w_i is usually calculated by the using the Euclidean distance metric equation:

$$\| w_i - Input_j \| = ((w_{i1} - Input_{j1})^2 + (w_{i2} - Input_{j2})^2 + \ldots)^{1/2}$$

The Kohonen Learning Law is then used to update the weights of the output unit with the smallest distance by using the following equation:

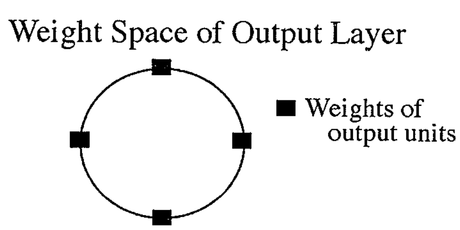
$$w_i(new) = (1 - \varepsilon)*w_i(old) + \varepsilon(Input_j)$$

Note: ε is a learning rate which is gradually decreased as learning progresses.

This learning law causes the weight vector to move, but its length remains unchanged. The connection weights of the units that did not win remain unchanged.

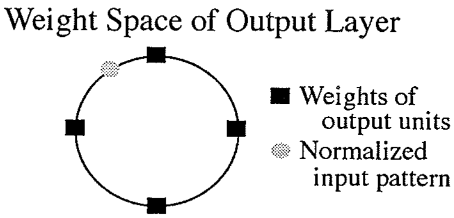
The learning process can also be interpreted geometrically. Suppose we have a competitive network with two input units (ie., two characteristics), and four output units (ie., four possible classifications). The weights on the output units are normalized and positioned along the boarder of a unit circle as shown in Figure 5.7.1.

Figure 5.7.1: Normalized Connections of Output Layer



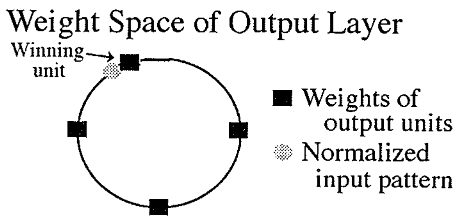
The normalized input pattern is then placed at its appropriate location on the unit circle as in Figure 5.7.2.

Figure 5.7.2: Weight Space With Input Vector



The distance metric is then calculated and the closest pattern wins. The weights of the winning unit are then altered so that they become closer to the input pattern as shown in Figure 5.7.3.

Figure 5.7.3: Weight Space With Updated Weight Vector



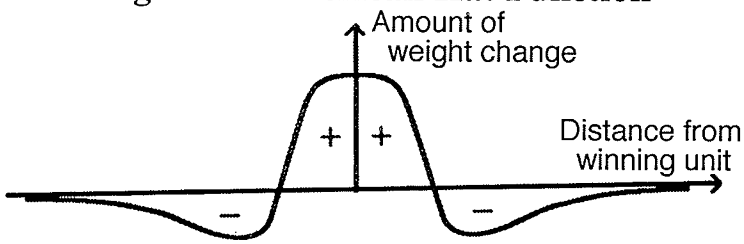
Note: The above process can be adapted to n dimensions (where n = the number of input units), by using a hypersphere instead of a two dimensional circle.

Most competitive neural networks include a mechanism, called a **conscience**, that ensures that a few processing units do not win most of the competitions. This mechanism improves learning.

5.5.2.1 Self-Organizing Map

The Self-Organizing Map (SOM) neural network is a variation on the basic competitive network model described in Section 5.5.2. The main difference between the basic competitive network and the SOM network involves the weight changes of the connections during learning. In the basic competitive model, the unit that wins the competition is the only unit that has its weights altered. However, in the SOM network, the units in the "neighborhood" of the unit that won the competition will have their weights updated. The amount that the weights are updated decreases as the distance from the winning unit increases. Units that are far away from the winning unit will have their weights slightly decreased (Scuse 74.720, p 204). The amount that the weights are modified is defined by the Mexican Hat function as picture in Figure 5.8.

Figure 5.8: Mexican Hat Function

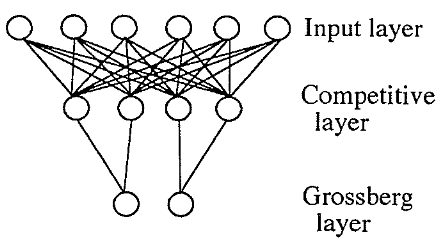


The SOM is very good at performing a mapping from the input space to the categorization or output space. The SOM differs from backpropagation because the SOM does not concentrate on individual features of the input pattern, but rather the general features and relationships of the signal are examined as a whole (Scuse 74.720, p 205).

5.5.2.2 Learning Vector Quantization

Learning Vector Quantization (LVQ) is an extension of competitive learning that combines supervised learning with unsupervised learning. The network usually has three layers--an input, competitive, and Grossberg or output layer--as shown in Figure 5.9.

Figure 5.9: LVQ Network



The competitive layer functions like the basic competitive network previously described in Section 5.5.2 and performs the categorization of the input patterns. The competitive layer uses unsupervised learning.

The Grossberg layer maps the outputs of the competitive layer to one of the output categories. The net input to each unit in the Grossberg layer is calculated the same way as in backpropagation. The Grossberg layer utilizes supervised learning and the desired outputs are supplied with the input patterns. During training the connection weights, w_{ij} in the Grossberg layer are modified according to the following formula:

$$w_{ij}(new) = w_{ij}(old) + \varepsilon(T_{ik} - w_{ik}(old)) * Input_k$$

Where ε = learning rate
T_{ik} = desired output
w_{ij} = connection weight between units i and j in adjacent layers

The overall effect of including the Grossberg layer is that each unit in the competitive layer is mapped to a unit representing a categorization in the Grossberg layer. This allows multiple competitive units to be mapped to a single class and permits a more generalized categorization process at the competitive layer. LVQ networks are usually not as powerful as backpropagation, but they can be trained more rapidly than backpropagation. LVQ networks can tolerate noise better than SOM networks because of their acretive behavior. However, the LVQ network has a somewhat limited ability to generalize (Scuse 74.719, Nov. 18 '93 p 3b).

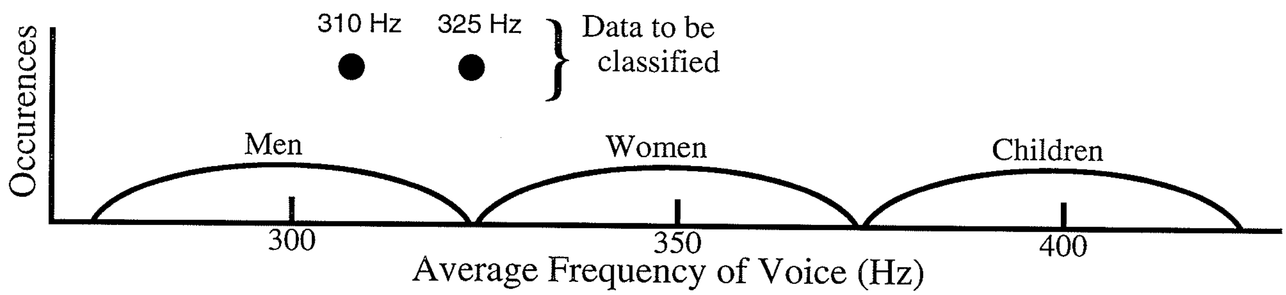
5.5.3 Probabilistic Neural Networks

Probabilistic Neural Networks (PNN) provide a general method for pattern classification problems. They utilize *a priori* probability and distribution functions to determine the classification of a new pattern. The PNN stores the training patterns in the weights of the network. It then estimates the distribution of classification and chooses the most probable classification category.

PNN's implement a Baysian classifier which takes into account the relative likelihood of an event occurring. For example, if an input pattern is equally likely to be in the category of wheeze or non-wheeze and wheeze is a rare event, then selecting the non-wheeze categorization is more likely to be correct. The use of this *a priori* information can help to improve the classification (NeuralWare, p NC-253).

The best way to illustrate how PNN's function is with an example. Suppose that you can differentiate between a man, woman, and child's voice on the basis of average frequency. Men have the lowest average frequency voices, woman have "in-between" pitched frequency voices, and children have the highest frequency voices. The population is sampled and the hypothetical results are shown in the histogram of Figure 5.10.

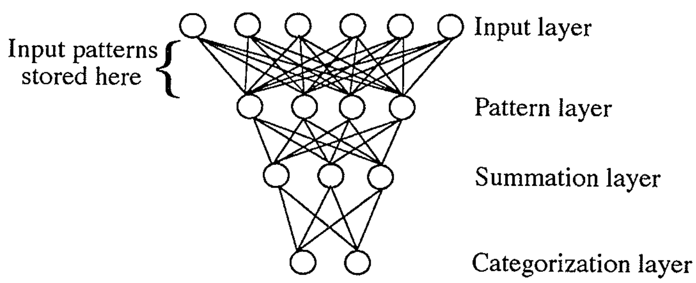
Figure 5.10: Hypothetical Histogram of Average Frequency of Voice Content



Suppose that someone has a voice of average frequency 310 Hz. According to the probability distribution, it is most probable that the voice belongs to a man. If someone has a voice of average frequency 325 Hz, the person is probably either a man or a woman. However, if we know that there are more men than women in the population under study, then the categorization of man is more probable (Wasserman 1993, pp 37-38).

The PNN usually has the architecture shown in Figure 5.11 with the following layers: 1) an input layer that performs some type of normalization; 2) a pattern layer where each pattern in the training set is stored as the weights between the input and pattern layer; 3) a summation layer which sums the probability that an input pattern belongs to a class; 4) a categorization or output layer which selects the class with the largest output or highest probability (NeuralWare, p NC-256).

Figure 5.11: Typical Architecture of a PNN



The processing of each node in the summation layer follows the following formula:

$$\text{Summation} = \sum_{\text{all inputs}} \exp((\text{Net input of connected nodes} - 1)/\sigma^2)$$

(Wasserman 1993, p 53)

σ is a smoothing variable which determines the width of categories, or may also be thought of as the standard deviation of a categorization. In the PNN each unit has the same σ value (Wasserman 1993, p 43). The value of σ is calculated with the following equation:

$$\sigma = \text{Sigma Scale} / \text{Number of Pattern Units}^{(\text{Sigma Exponent} / \text{Number of Input Units})}$$

(Wasserman 1993, p 150)

The categorization layer finds the unit with the largest summation and that unit is allowed to fire. All other units have an output value of 0. The summation layer can be thought of as being a separate competitive network.

PNN's offer the following advantages:

1) Rapid training: PNN's may be five times faster than backpropagation for training because training mainly involves normalization and storage of patterns.

2) Optimal Solution: Because PNN's utilize a Baysian classification scheme, a PNN is guaranteed to converge to an optimal solution.

3) Versatility of Data Sets: Data may be added to, or deleted from the training set without the lengthy retraining of the network as must occur in backpropagation.

4) Certainty of Categorization: The PNN can be easily adapted to provide an output value indicating the amount of evidence upon which the network is basing its decision.

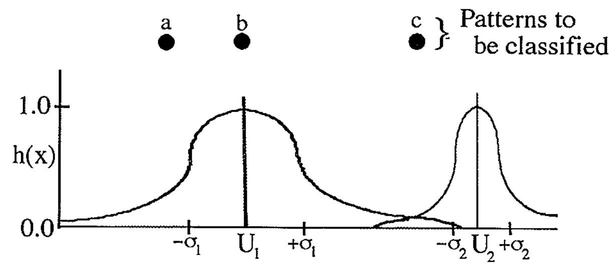
(Wasserman 1993, p 36)

PNN's work well in noisy environments because they demonstrate acretive behavior. However, they are unable to interpolate to new a classification when a novel pattern is shown to the network.

5.5.4 Radial Basis Function

The Radial Basis Function (RBF) network is very similar to the PNN discussed in Section 5.5.3. One main difference is that a basis function is defined for the receptive field of each training pattern. In the case of PNN's, a test pattern receives the same classification as the training pattern it most closely matches. However, in RBF networks, the output classification is scaled so that it represents the scaled classification with respect to each training pattern in its receptive field. The width of the receptive field is analogous to the σ value for PNN's and can be different for individual processing units. This process is demonstrated in the example depicted in Figure 5.12.

Figure 5.12: RBF Pattern Classification



In the above example, U_1 and U_2 represent learning patterns with their receptive fields defined such that they follow the Gaussian basis function (ie., 95% of the data are within $\pm 2\sigma$). The Gaussian basis function is:

$$h(x) = \exp(-\| Input - U \| / (2\sigma^2))$$

Input represents an input vector, and U is a stored vector
|| Input - U || represents the Euclidean distance between the two vectors
The value of σ is calculated in the same way as for PNN's, except each processing unit may have its own value for σ.

From the above equation, notice that the value of the function h(x) decreases as the input vector diverges from stored vector U. The input patterns a, b, and c, in Figure 5.12 need to be classified. Pattern b is identical to training pattern U_1 and so its classification will be the same as pattern U_1. Input pattern a, also most closely matches stored pattern U_1 but it is not an exact match. Therefore, a weighted output which uses the Gaussian basis function, h(x), is used to calculate the output classification. The output layer or mapping layer of the RBF network accomplishes this task with the following equation:

$$\sum h(x) * w_{ij}$$

This value is calculated for each of the adjoining connections. Similarly, the input pattern c is categorized by a weighted classification of stored patterns U_1 and U_2.

The RBF neural network generally has four layers, an input layer, a pattern layer where the learning patterns are stored, a hidden layer that performs most of the processing, and an output layer. The RBF network generally shares the same advantages as the PNN network.

However, the RBF network can also be utilized as a mapping network because it has good interpolative behavior.

5.6 Successful Applications of Neural Networks

The processing ability of neural networks has led to their utilization in a variety of fields. For example, neural networks have been utilized to drive a vehicle along a road (Pomerleau), to classify sonar signals (Gorman and Sejnowski), to compress images (Cottrell et al.), to classify EKGs (Bortolan et al.), and to diagnose malignant melanoma (Chawla and Ercal). There are other techniques available for solving the above mentioned problems. However, neural network solutions can often be built quicker than the other more traditional approaches (Rich and Knight, p 517).

5.7 Summary

This chapter presented several different types of neural networks and the biological basis for the neural network. It also made a comparison of neural networks to algorithmic and rule-based systems and found that there are several advantages to neural networks. The area of lung sound classification is very subjective and lacks objective mathematical definitions for adventitious lung sounds such as wheezes. Therefore, the field of neural networks should perform quite well at wheeze detection.

Although each type of neural network may have a different architecture, each neural network is composed of the same fundamental building block, the processing unit; however, different types of neural networks may contain different types of processing units. Because each type of neural network has its own characteristics, each network may process data differently and therefore have different performances. The BP neural network performs a global optimization of the set of training patterns and so it attempts to learn the general features of the patterns. The other network models--RBF, LVQ, SOM, and PNN--perform a local optimization of the training patterns. With these types of neural networks, each "hidden" unit represents an individual training pattern or a prototype of the pattern. During the testing phase, the new pattern that is presented to the network is compared with the pattern stored at each hidden unit. A competition occurs and the hidden unit that generates the largest response wins the competition and is allowed to generate the output value for the network.

Chapter 6

Comparison of Lung Sound Representations

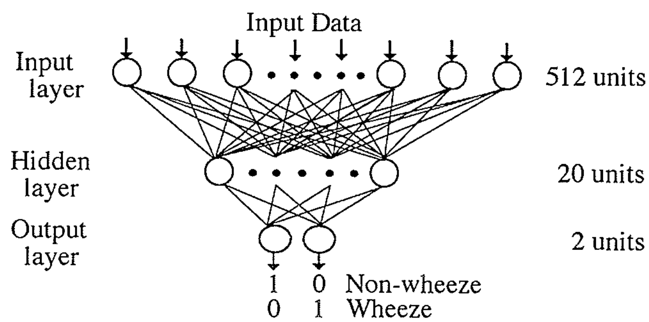
6.1 Data Sets Used in the Experiments

The data sets used in this thesis were composed of lung sounds recorded at the Respiratory Acoustics Laboratory as outlined in Section 4.2. All lung sounds were recorded from the microphones numbered one to seven in Figures 4.6 and 4.7. Data from microphone eight, representing sounds from the trachea, were not used for reasons outlined in Section 4.2.

All of the data used in Chapter 6 were taken from eight patients who wheezed and eight patients who did not. The data were sampled at 10 kHz and then down sampled to 5 kHz by taking an average of every two consecutive data values.

A three layer backpropagation network with the architecture shown in Figure 6.1, was created with NeuralWorks (NeuralWare) and used to determine the best representation for the lung sounds. Any type of neural networks described in Chapter 5 could have been used for this purpose, but the backpropagation network was chosen because it could be readily implemented and easily manipulated.

Figure 6.1: Architecture of the Neural Network Used in Chapter 6



All neural networks were trained for 30,000 epochs with a training schedule which progressively reduced the learning rate from 0.9 to 0.00009. If after 30,000 epochs the RMS error was not sufficiently small, more training was performed. The weights of each network were initially randomized so that they would have a value between -1.9 and +1.9.

All networks implemented in Chapter 6 had 512 input units, 20 hidden units and 2 output units. A network output of (1 0) was used to indicate that the neural network classified the sound as non-wheeze and an output of (0 1) was used to indicate that the neural network detected a wheeze in the lung sound. Output values of the network consisted of real values between 0.0 and 1.0. Therefore, after the network performed its processing, values below 0.5 were rounded to 0 and all other values were rounded to 1. In the unlikely event that both values are rounded to the same value, the classification was interpreted as being incorrect. This event never occurred during any of the experiments.

6.2 Representation of the Signal

A number of different representations of the signal were examined. Each representation explored in Section 6.2 uses data taken from the same patient population. The number of data patterns in the test and training sets was kept constant throughout the experiments and the composition of the data set is displayed in Table 6.1.

Table 6.1: Data Set Used in Chapter 6

Data Set	Number of Non-Wheezes	Number of Wheezes
Train	106	105
Test	123	121

All raw signal representations use 0.1 seconds of lung sounds per data pattern. The FT spectrum representation uses 0.2 seconds of lung sounds per data pattern because 0.2 seconds of data were required to compute the FT frequency components between 1 and 2048 Hz. After segmenting the data, lung sounds were classified manually as either non-wheeze or wheeze as outlined in Section 4.2. All data were then linearly scaled so that the data ranged in value between 0.0 and 2.0. This scaling prevents large input values from saturating the network.

6.2.1 Raw Signal Representation

Lung sound data on which no transformation was performed were used to test and train the backpropagation network described in Section 6.1. The RMS of the network was reduced to 0.2 during learning. Additional training was performed; however, the RMS was not reduced nor was network performance increased.

When the network was tested with the training set, it was able to correctly classify all of the patterns. However, when the neural network was tested with lung sounds that were not in the training set (ie., the test set), it was able to classify 102 of the 123 non-wheeze segments correctly, and 87 of the 121 wheeze segments correctly. In this instance, the network had an overall performance of 77%. The lower performance for the test set suggests that the network was not able to generalize the concepts of wheeze detection from the raw signal. It probably had memorized the training patterns and so was able to classify correctly all of the data in the training set.

6.2.2 Raw Data With a Highpass Filter

A highpass filter which used the FFT was used to preprocess the data for the neural network. The FFT was calculated on the lung sounds, and then the frequencies between 1 and 100 Hz were removed by zeroing their corresponding FFT values. The inverse FFT was then calculated to obtain the filtered lung sound representation. The data were then divided into 0.1 second segments and used to train and test the network outlined in Section 6.1.

The network was able to classify all patterns in the training set correctly. The network was able to classify 104 of the 123 non-wheeze segments and 95 of the 121 wheeze segments. The overall network performance on the test set was 82% which was better than the 77% performance obtained with raw signal data. This indicates that the filtered representation is a better representation for the data in the test set than the unfiltered representation.

In early experiments, a crude highpass filter was developed using the first derivative. The first derivative is not exactly a highpass filter, but rather enhances high frequency components more than low frequency components. The first derivative is also a fair method for performing trend analysis (Stein, p 35). The above experiments were repeated using the first derivative as the filter. The network was able to classify all patterns in the training set correctly. The network was able to classify correctly 106 of the 123 non-wheeze segments and 93 of the 121 wheeze segments. The overall network performance on the test set was 82% which was the same as for the highpass filter.

6.2.3 FT Representation of the Lung Sounds

The FT was used to transform the lung sounds from the time domain to the frequency domain. The lung sounds were taken from the same set of patients as outlined in Section

6.1, but the data were divided into 0.2 second segments instead of 0.1 second segments. The Hanning window function was applied and a 1024 point FT was calculated. A periodogram was then constructed as outlined in Section 3.2.3. The resulting 512 unique spectrum values were then linearly scaled so that they ranged in value between 0.0 and 2.0. The training set contained 106 non-wheeze segments and 105 wheeze segments. The test set consisted of 123 non-wheeze segments and 121 wheeze segments. Except for the data set, all other experiment parameters were identical to the parameters of the network used in Section 6.2.1.

During training, the RMS of the network was reduced to 0.07, which indicates that the network did a "good" job at learning the training patterns. When the neural network was tested with the training set, it was able to correctly classify all of the patterns. However, when the neural network was tested with lung sounds that were not in the training set (ie., the test set), it was able to classify 112 of the 123 non-wheeze segments correctly and 103 of the 121 wheeze segments correctly. The network's reduced overall performance of 88% with the test set indicates that the neural network was able to generalize; however, it did not learn all the concepts of wheeze detection.

6.3 Number of Data Points Used to Represent the Lung Sounds

Experiments were also performed that involved altering the number of data points that represent the lung sounds. Once again, all data came from the same patients as outlined in Section 6.1. Each data set contained the same number of non-wheeze and wheeze segments as the data set of Section 6.2. The neural network was the same type of network as used in Section 6.2.1 The data were divided into 0.05 second segments (or 256 input values), 0.1 second segments (or 512 input values), and 0.2 second segments (or 1024 input values) and used to train and test three different networks. Each segment was reviewed to make sure that it either contained no wheezes or contained some wheeze sounds. The overall performance rates for each test set are displayed in Table 6.2.

Table 6.2: Affect of Altering the Number of Input Values for the Raw Signal

Number of Input Units	Overall Performance
256	80%
512	77%
1024	70%

The results in Table 6.2 indicate that as the number of data points representing the signal decreases, the performance increases. This probably occurs because as the signal is recorded for a longer time period, there is a greater chance that some part of a wheeze signal contains a non-wheeze segment. If a signal contains both wheeze and non-wheeze segments, it should be classified as containing a wheeze. However, if the signal is made up of very little wheezing and much noise, the neural network will have a difficult time detecting the wheeze with the raw signal representation. In addition, as the signal is analyzed for a longer time period, there is an increased chance that the appearance of the signal may be altered by the inclusion of noise, time offset, or alterations in frequency and amplitude.

One problem with using 256 data values to represent the signal is that it represents only 0.05 seconds of data (because the effective sampling rate is 5,120 Hz). However, by the definition outlined in Section 2.4.2.2, a wheeze must be at least 0.1 seconds long. Therefore, using less than 0.1 seconds of data (or 512 data values) is impractical because it is less than the duration of a wheeze.

The number of data points processed by the network from the FT spectrum was also altered. The frequency range of 1 Hz to 2048 Hz was used and compared with a much smaller frequency range of 70 Hz to 1,300 Hz[1]. The former representation consisted of 512 data input values and the latter consisted of 247 data values. When the backpropagation network of Section 6.1 was trained and tested on the data sets, overall performance improved from 88% with the 512 data point representation to 91% with the more compact representation. It is believed that this increase in performance is due to a more narrowed focus of the network in the compact representation. The compact representation allows the network to focus on the frequencies where a wheeze may be found and so the network does not have to contend with as much noise as in the case of the larger representation. The noise may be incorrectly interpreted by the network as significant and therefore cause a reduced network performance.

6.4 Discussion of the Results

The results of this section are summarized in Table 6.3.

Table 6.3: Performance of Representations When Tested With Test Set

Representation	Specificity* (Non-Wheeze Performance)	Sensitivity* (Wheeze Performance)	Overall Performance*
Raw Signal	83%	72%	77%
Filtered Data	85%	79%	82%
Derivative of Signal	86%	77%	82%
FT Spectrum	91%	85%	88%

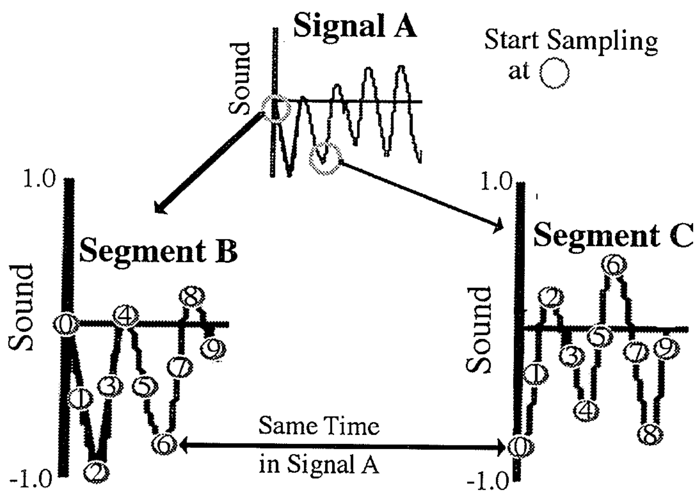
* Specificity, sensitivity, and overall performance have been defined in Appendix B.

It was observed that the FT spectrum representation provided the best performance for the data. The poor performance of the raw signal representation was probably due to the wide range of possible input values representing a wheeze in the raw signal format. For example, suppose that two segments (Segment B and Segment C) of the same signal (Signal A) are

[1] Note: Section 2.4.2.2 mentions that a wheeze may range from 60 to 2,000 Hz. However, wheezes lower than 70 Hz or above 1,300 Hz were not encountered in any of the data sets used in this research.

taken. However, Segment C is sampled at an offset in time that is slightly after Segment B. This scenario is depicted in Figure 6.2.

Figure 6.2: Timeform Result of Sampling Signal at a Different Time Offset



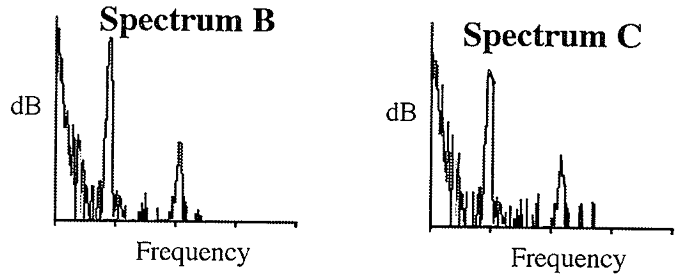
The first ten values for Segment B and Segment C are displayed in Table 6.4.

Table 6.4: Sampled Values for Time Offset Signals

Sample Number	Segment B	Segment C
0	0.0	-0.8
1	-0.5	-0.3
2	-1.0	+0.2
3	-0.45	-0.1
4	0.0	-0.5
5	-0.45	-0.05
6	-0.8	+0.45
7	-0.3	-0.1
8	+0.2	-0.8
9	-0.1	-0.07

The neural network input units will each receive different values for Segment B and Segment C. It will be very difficult for the neural network to recognize that Segment B and Segment C are similar signals. However, the FT spectrum will appear very similar, and is depicted in Figure 6.3. (Note: The actual FT spectra were computed on a larger portion of the original signal than depicted in Figure 6.2.)

Figure 6.3: FT-form Result of Sampling Signal at a Different Time Offset



The Fourier Transform representation is more time offset invariant than the raw signal format.

In addition to time offset variation, the signal may contain low frequency noise which alters its appearance in the raw signal representation. When the signal was filtered, low frequency noise was removed and the patterns became less variable and so classification performance improved slightly. However, the results were still not as good as the FT representation. Wheezes of different frequencies are also less variable in the Fourier Transform representation than the raw signal representation. The variability of the raw signal format is also applicable for segments not containing wheezes.

All representations and associated neural networks were also tested on lung sounds recorded from patients not included in the original patient population. The neural networks were again trained on the training set. However, they were tested on the data set Test_2 which was composed of lung sounds extracted from patients not in the previous study. This was done to see how the neural network would perform in a situation more similar to a real world setting where it may not be possible to train with data from the patient before the data is processed by the network. The data set Test_2 had 121 non-wheeze segments and 70 wheeze segments. The results are summarized in Table 6.5.

Table 6.5: Performance of Representations When Tested With Test_2

Representation	Specificity	Sensitivity	Overall Performance
Raw Signal	73%	67%	71%
Filter Raw Signal	75%	71%	74%
Derivative	77%	70%	74%
FT of Raw Signal	86%	80%	84%

All four neural networks exhibited a decrease in performance when tested on new patient data. This probably occurred because every person possesses different physiological characteristics which may cause unique lung sounds to be produced. The FT of the raw signal still had the best performance and was the best representation of the data. Although its overall performance decreased by 4%, it is believed that the performance could be improved if the size of the training set were increased and taken from a much larger patient population. This would give the network more "experience" in categorizing lung sounds from a large range of individuals.

It was also found that more compact and discrete representations of the lung sounds return better results. However, the representation must not be so compact that it does not represent a wheeze in the true sense of its definition as was found when only 0.05 seconds of data were processed.

Chapter 7

Comparison of Neural Network Classifiers and Post-Processing Techniques

7.1 Data Sets Used in Chapter 7

The data used in Chapter 7 were recorded at the Respirology Acoustics Laboratory as outlined in Section 4.2. All data were taken from microphones located at positions one through seven as displayed in Figures 4.6 and 4.7. The experiments were performed with the same software and equipment as used in Chapter 6. However, the data sets used in Chapters 6 and 7 are different because the data sets in Chapter 7 were taken from a much larger range of individuals than the data sets of Chapter 6.

The lung sounds used in the experiments of this chapter were recorded from 28 patients who did not wheeze and 15 patients who did wheeze. (Note: Only 8 non-wheeze and 8 wheeze patients were used in Chapter 6). All patients were randomly placed into one of two groups, Sounds_1 and Sounds_2. (Note: Patients in Sounds_1 were different from patients in Sounds_2.) The lung sounds from Sounds_1 were then randomly placed into data sets Sounds_1.A or Sounds_1.B. The data set Sounds_2 contained data taken from patients that the neural network had not seen before and is used to represent how the classifier would function with patients being tested for the first time, as in a real world case scenario. The wheeze lung sounds in Sounds_2 were generally of lower intensity and contained more high frequency noise than the wheezes found in Sounds_1. (Note: Low intensity wheezes were between 10 and 30 dB.) The lung sounds of Sounds_2 also had a decreased signal to noise ratio, which should make it harder to classify the lung sounds. All of the lung sounds were broken into 0.2 sec segments and classified as outlined in Section 4.2. The type and number of lung sounds in the data sets are shown in Table 7.1.

Table 7.1: Number of Patterns in the Data Sets

Data Set	Non-Wheeze Patterns	Wheeze Patterns	Total Patterns
Sounds_1.A	114	128	242
Sounds_1.B	126	107	233
Sounds_2	95	116	211
Total	335	479	686

The lung sounds were originally sampled at 10 kHz and then down sampled to 5 kHz by taking an average of every two consecutive data values. The Hanning window function was then performed and a 1024 point FFT was computed as outlined in Section 3.2.3. The data within the range of 70 Hz to 1,300 Hz were then used to calculate the FT spectrum. The resulting 247 FT spectrum values were linearly scaled so that they ranged between 0.0 and 2.0. This scaling was necessary to prevent large values from saturating the network.

7.2 Comparison of Neural Network Models

The five neural networks presented in Chapter 5 were tested with the three data sets of Section 7.1. Most of the neural networks had two output units, one representing a non-wheeze classification, and the other representing a wheeze classification. In all subsections of Section 7.3, a simple method of interpreting the network outputs was implemented; the output unit with the largest output value was used as the network's classification category. If both output units had the same value, the classification was taken as being incorrect; however, this event never occurred in any of the experiments. In most instances, the data set

Sounds_1.A was used to train the network and all three data sets were then used to evaluate the network's performance.

7.2.1 Backpropagation Neural Network

A backpropagation (BP) network with 247 input units, 20 hidden units, and 2 output units was constructed. The neural network was trained for 50,000 epochs with a training schedule which progressively reduced the learning rate from 0.9 to 0.00009. The network weights were originally randomized so that they would have values between -1.9 and +1.9. The results of training the network with Sounds_1.A are displayed in Table 7.2.

Table 7.2: Backpropagation Network Results

Data Set	Specificity* (Non-Wheeze Performance)	Sensitivity* (Wheeze Performance)	Overall Performance*
Sounds_1.A	100%	100%	100%
Sounds_1.B	88%	91%	89%
Sounds_2	91%	96%	95%

* Refer to Appendix B for an explanation of how specificity, sensitivity, and overall performance were calculated

Several backpropagation networks were used in the study. All the BP networks had similar architectures: an input layer, one hidden layer, and an output layer. A BP network with two hidden layers was also implemented, but the results were significantly worse than with only one hidden layer. In addition, it was also found that as many as forty and as few as five hidden units could be used without causing a large drop in network performance. Networks that had more than forty hidden units had a poor performance when tested on the data sets Sounds_1.B and Sounds_2 because the network had just memorized the patterns in the training set. Networks that contained fewer than five hidden units also had a poor performance with all three data sets, including the training set. This occurred because the network had a reduced processing capacity and so was not able to learn the general features of wheeze classification or memorize the patterns in the training set.

The phenomena with the number of hidden units also occurred in some early experiments when the raw signal representation was used to train the network. As the number of hidden units increased past forty, memorization occurred. As the number of units decreased below ten, the network was unable to learn the patterns.

In other experiments, the number of output units was decreased to one and the classification of non-wheeze was made if the output value was less than 0.5, otherwise the pattern was classified as a wheeze. The overall performance of this network was very poor, (only slightly better than 60%). Including two output units improved performance substantially. Using two output units is analogous to having two separate neural networks. Howard Card, an expert within the field of neural networks explained that: "the two output units each have their own set of weights which can be tuned independently by training. With a single unit the common weights have to satisfy both classes' needs" (Card).

The results of the BP network were somewhat surprising because the performance with the data set that contained new patients, data set Sounds_2, was better than the performance on patient's with which it had trained, data set Sounds_1.B. One possible explanation for this occurrence is that the training set, Sounds_1.A and Sounds_2 contained more wheeze patterns than non-wheeze patterns. As a result, the network may have become more proficient at characterizing wheeze patterns than non-wheeze patterns. A new BP network with the same characteristics as the network above was trained with the data set Sounds_1.B instead of the data set Sounds_1.A. This resulted in an overall improvement in performance of both test sets. The results are shown in Table 7.3.

Table 7.3: Backpropagation Results When Trained With Sounds_1.B

Data Set	Specificity	Sensitivity	Overall Performance
Sounds_1.A	95%	91%	93%
Sounds_1.B	100%	100%	100%
Sounds_2	93%	95%	94%

From Table 7.3 it may be noted that the network's wheeze performance either decreased slightly or stayed the same but the non-wheeze performance increased for each test set. This was likely due to the reduced number of wheezes and increased number of non-wheezes within the training set.

7.2.2 Radial Basis Function Neural Network

A Radial Basis Function (RBF) network with 247 input units, 50 prototype units, 20 hidden units, and 2 output units was created. The network used a simple Euclidean function to calculate distance between patterns and used the same basis function as stated in Section 5.5.4 for processing. The delta learning rule and sigmoid transfer function were used during training and the connection weights were randomized between -1.9 and +1.9. The network was trained for 30,000 epochs and the results are displayed in Table 7.4.

Table 7.4: RBF Network Results

Data Set	Specificity	Sensitivity	Overall Performance
Sounds_1.A	97%	95%	96%
Sounds_1.B	91%	94%	93%
Sounds_2	95%	97%	96%

The network architecture was also altered and the experiments were performed again. The number of input and output units remained fixed at 247 and 2 respectively. The number of prototype units was also fixed at fifty. Experiments were performed with the number of hidden units ranging between five and sixty. It was found that when the number of hidden units was as low as five or as high as fifty, performance degraded by only 3%. The same experiments were performed by varying the number of prototype units and keeping all other parameters constant. The number of hidden units was twenty for the duration of these experiments and the number of prototype units was varied between ten and one hundred. It was found that when the number of prototype units was as low as ten that performance degraded only by 4%. No significant change in performance was noted when one hundred prototype units were used. Therefore, the number of hidden and prototype units did not alter performance by any significant amount when they were changed within the specified ranges.

7.2.3 Self-Organizing Map Neural Network

A Self-Organizing Map (SOM) network with 247 input units, 100 hidden units, and 2 output units was trained and tested with the data of Section 7.1. The neural network was trained for 20,000 epochs with a training schedule which progressively reduced the learning constant from 0.9 to 0.00009. The network weights were originally randomized so that they would have a value between -1.9 and +1.9. The network used the delta learning rule and the sigmoid transfer function. A neighborhood value, which determines which weights become updated during learning, was set to have a square shape. The results are displayed in Table 7.5.

Table 7.5: SOM Network Results

Data Set	Specificity	Sensitivity	Overall Performance
Sounds_1.A	93%	94%	93%
Sounds_1.B	89%	93%	91%
Sounds_2	89%	92%	91%

A number of network parameters were altered in an attempt to maximize performance. Changing the number of hidden units to any value between twenty-five and four hundred had very little effect on network performance, other than as the number of competitive units increased, processing time became substantially larger. Several neighborhood values were also tried; however, performance was unchanged.

7.2.4 Learning Vector Quantization Neural Network

A Learning Vector Quantization (LVQ) neural network with 247 input units, 48 competitive (hidden) units, and two Grossberg (output) units was trained and tested with the data of Section 7.1. The network was implemented with a conscience factor that causes processing elements that win too often to develop a "guilty conscience" and so they become penalized for winning too often (NeuralWare, p RF-38). The network was trained with two learning algorithms. First the LVQ1 algorithm was used for 10,000 epochs with a learning rate of 0.06. The LVQ1 algorithm allows the network to learn a "coarse" classification. The LVQ2 algorithm was then used for 5,000 epochs with a learning rate of 0.03 to refine the solution and provide a "fine" classification (NeuralWare, p RF-38). The results are shown in Table 7.6.

Table 7.6: LVQ Network Results

Data Set	Specificity	Sensitivity	Overall Performance
Sounds_1.A	99%	98%	98%
Sounds_1.B	95%	91%	93%
Sounds_2	97%	96%	96%

Another experiment was performed with the same network architecture but a reduced number of training cycles. The NeuralWare manual suggests that the number of training cycles should be:

For LVQ1: 30 * (# training patterns)
For LVQ2: 10 * (# training patterns)
(NeuralWare, p RF-41)

Since there are 242 patterns in the training set, the network was trained for 7,260 epochs with LVQ1 and 2,420 epochs with LVQ2. The learning rates were the same as above. This resulted in a wheeze performance reduction of about 8%, and a decrease in non-wheeze performance of 6% from the results stated in Table 7.6. When the network was trained for an additional 8,000 epochs with each training set, performance was the same as in Table 7.6. When the number of competitive units was reduced to only ten and the network had the same parameters as the network in Table 7.6. The performance was similar, except non-wheeze performance had decreased by 2% with the file Sounds_1.B. Overall, it was found that altering the number of competitive units between the range of ten to sixty had little affect on network performance. However, reducing the number of training cycles does have an effect on training. It was found that for the lung sound data used in the experiment, the recommended number of training cycles was too low to learn the general features of non-wheeze and wheeze classification.

7.2.5 Probabilistic Neural Network

Several experiments were performed with a Probabilistic Neural Network (PNN) with 247 input nodes that processed the data sets of Section 7.1. An optimal performance of 65% on Sounds_1.B was achieved when the system was trained on Sounds_1.A with 247 input values per pattern. This result was achieved when the network had 200 pattern units, 2 output units, a sigma scale value of 0.392, and a sigma exponent value of 0.5. (Note: PNN's require only one pass through the training set.) Several different networks were constructed which altered the number of pattern units and it was found that when fewer than fifty or more than three hundred pattern units were included, performance dropped to 50%, with most patterns being classified as non-wheeze. The values of sigma exponent and sigma scale were also altered. It was found that altering the sigma scale value between 0.4 and 0.3 had little effect on the performance; however, outside of these values, performance decreased to 50%. Similarly, the values of sigma exponent could be altered between 0.2 and 0.9 without affecting performance; however, outside of these values performance degraded to 50%.

In an attempt to decrease the input vector size and increase performance, every five consecutive FT spectrum values were averaged into one value. Several network configurations were tried. The best results were obtained when the network had 300 pattern units, 2 output units, a sigma scale value of 0.392, and a sigma exponent value of 0.5. The results are displayed in Table 7.7.

Table 7.7: PNN Network Results

Data Set	Specificity	Sensitivity	Overall Performance
Sounds_1.A	100%	81%	90%
Sounds_1.B	100%	65%	84%
Sounds_2	100%	80%	89%

It appears that the size of the input pattern was too large to obtain adequate results with the PNN classifier. One reason that the input vector had to be reduced in size is that the input pattern space may have been too large for the network to process the data adequately. Further discussion of the PNN results is beyond the scope of this thesis, and so no further comment will be made regarding the PNN.

7.2.6 Discussion of the Neural Network Results

A summary of the overall neural network performance on the two test sets (Sounds_1.B and Sounds_2) is displayed in Table 7.8.

Table 7.8: Comparison of Neural Network Types

Data Set	Specificity	Sensitivity	Overall Performance
LVQ	96%	94%	95%
RBF	93%	96%	95%
BP	90%	94%	92%
SOM	89%	93%	91%
PNN	100%	73%	87%

It was found that the LVQ and RBF networks had the best results, with an overall performance of 95%. The BP and SOM were slightly worse with a performance of 92%

and 91% respectively. The PNN had the worst performance at 87%; however, it used a different representation of the data.

Each neural network performs a different type of processing. The BP network performs a global optimization of the set of training patterns. This will cause the BP network to learn the general features of pattern classification (assuming that there is an adequate number of hidden nodes) (Hertz et al., Chapter 6). The other models--RBF, LVQ, SOM, and PNN--perform a local optimization of the training patterns. With these models, each unit in the "hidden" layer represents an individual training pattern or prototype pattern. During the testing phase, the test patterns are compared to patterns stored at each hidden unit and a competition is carried out. The hidden unit that contains the pattern that most closely matches the test pattern will generate the largest response, win the competition, and generate the appropriate output. However, as the number of hidden units was altered, network performance generally remained constant. This indicates that the training set may have many similar lung sound patterns. When the lung sounds were reviewed, it was found that lung sounds taken from the same patient and at the same time in the respiratory cycle were very similar. Because the number of hidden units could be reduced slightly without degrading performance, it is possible that similar patterns were assigned to the same hidden units. Therefore, some of the hidden units may have contained redundant data, and so could be deleted without degrading network performance. However, training the network with lung sounds taken at a large range of times in the respiratory cycle or taken from different patients should cause performance to degrade when the number of hidden units are reduced.

7.3 Increasing the Size of the Training Set

Increasing the size of the data set and sampling the data at different times in the respiratory cycle should improve network performance because the network obtains more "experience" with classifying different types of data. The patterns in the data sets Sounds_1.A and Sounds_1.B were combined into one large data set and used to test and train the BP network of section 7.2.1. This was performed for the BP network and the results are shown in Table 7.9.

Table 7.9: Backpropagation Network Results With Large Data Set

Data Set	Specificity	Sensitivity	Overall Performance
Sounds_1.A + Sounds_1.B	100%	100%	100%
Sounds_2	95%	98%	96%

The results displayed in Table 7.9 support the hypothesis that performance should improve when the size of the training set is increased. In addition, performance started to degrade as the number of hidden units was reduced below ten. This also supports the hypothesis that a larger and more diverse training set requires more hidden units to store the general information of pattern classification for larger training sets. (Remember the BP network that was trained with Sounds_1.A had its performance degrade when only five hidden units were used as opposed to ten hidden units for the current situation.)

A BP network was trained with all the data from both groups of patients and then the network was tested with all of the data. This process ensures that all patterns are consistent with each other and no patterns have been misclassified. The BP network had a performance of 100%. Both the RBF and LVQ networks were also tested in the same way and had performances of 95% and 99% respectively. It is likely that by adjusting the number of patterns in the hidden layer for the RBF and LVQ that the performance would have improved.

7.4 Thresholding the Results of the Neural Network

In the previous sections, a pattern was assigned to a class (wheeze or non-wheeze) by a simple process; the output unit with the larger value determines the class. This method of classification results in a fairly good classification rate (overall performance of at least 91%, as in the case of SOM). However, the strategy causes some data to be misclassified (approximately 9% of the patterns, as in the case of SOM), which is very undesirable in a medical setting. However, if a certainty value is associated with the classification, then classifications with small certainty values could be re-examined by another lung sound classification device or by a physician. This would decrease the number of incorrect classifications and improve performance.

A post-processing mechanism was used to interpret the data and convey certainty in a diagnosis. In addition to the non-wheeze and wheeze categories, an "unknown" category was created by incorporating threshold values for the output units of the neural network. The threshold was implemented in the following way:

if (Output_Layer_Unit_with_Larger_Value) >= (Larger_Threshold)
and
(Output_Layer_Unit_with_Smaller_Value) <= (Smaller_Threshold)
then classification is known (as either wheeze or non-wheeze)
otherwise classification is unknown

Once it has been determined that the classification is within the "known" category, it is classified as either being a non-wheeze or wheeze according to which output unit has the larger value; otherwise, it is classified as "unknown". One of the main advantages of this classification mechanism is that if the network is uncertain about a how to classify a lung sound, then the likelihood of an incorrect classification is reduced as it will be placed into the "unknown" category.

Experiments were performed with an RBF network with an architecture similar to the one outlined in Section 7.2.2. The network was trained on the data set Sounds_1.A and tested with the three data sets of Section 7.1. An upper threshold value of 0.8 and a lower threshold value of 0.2 were used. The results are shown in Table 7.10.

Table 7.10: RBF Network Results With Thresholding (0.8/0.2)

Data Set	Correct Response*	Incorrect Response*	Unknown Response*
Sounds_1.A (training)	89%	2%	9%
Sounds_1.B	82%	4%	14%
Sounds_2	86%	2%	12%

* Please refer to Appendix B for more information regarding the calculation of these values.

The experiment was repeated using an upper threshold value of 0.9 and a lower threshold value 0.1. The results are displayed in Table 7.11. These threshold values require that the classification be known with a higher degree of certainty than the previous threshold values.

Table 7.11: RBF Network Results With Thresholding (0.9/0.1)

Data Set	Correct Response	Incorrect Response	Unknown Response
Sounds_1.A (training)	74%	2%	24%
Sounds_1.B	71%	2%	27%
Sounds_2	72%	2%	26%

As can be seen from the results displayed above, the combined error rate with Sounds_1.B and Sounds_2 dropped from 5.5% (when a simple comparison technique was implemented as in Table 7.4) to 2% (when a threshold of 0.9/0.1 was used). However, there is a tradeoff; as the number of incorrect responses decreases, the number of correct responses also decreases. In this case, the overall number of incorrect responses dropped from 5.5% to

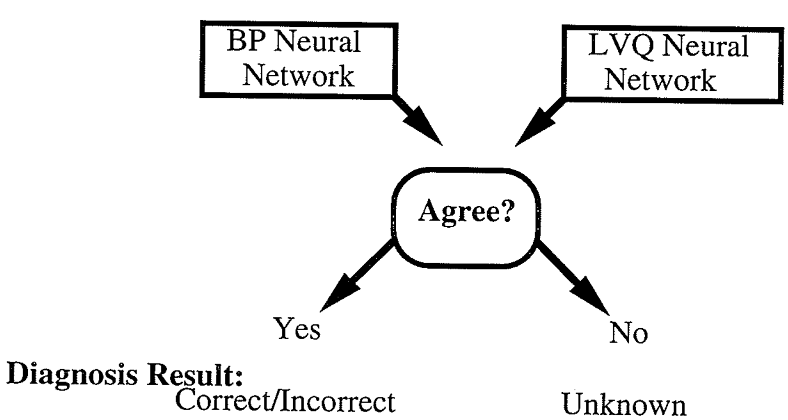
2% when a threshold of 0.9/0.1 was introduced but the number of correct responses also dropped from 95% to 72%. However, it must be remembered that it is better to have a computer classify a data pattern as unknown than to classify the pattern incorrectly. The 26% of the data that were classified as unknown can be further classified by another approach such as having the physician classify the data.

7.5 Consensus Diagnosis

Consensus diagnosis is a post-processing technique that combines the results of two or more classifiers in an attempt to increase the performance rate (Ng and Abramson, p 916). Combining more than one classifier was implemented by (Xu et al., 1992, pp 418-435) to improve the classification performance of handwriting recognition tools. The study found that combining the results of more than one classifier always provided as good or better results than the results than could be obtained from an individual classifier.

In the research presented here, consensus diagnosis was performed by comparing the results of two different neural networks. If both networks agree on the classification, the classification is accepted and it is either correct or incorrect; otherwise the classification was identified as "unknown". This process is outlined in Figure 7.1.

Figure 7.1: Consensus Diagnosis with Two Neural Networks



The BP, RBF, and LVQ neural networks from Sections 7.2.1, 7.2.2, and 7.2.4 were each trained with the data set Sounds_1.A and then each network was paired with another network to formulate a consensus diagnosis. In each network, the output unit with the largest value was taken as being the network's classification, and so no threshold was implemented. The results are displayed in Table 7.12.

Table 7.12: Two Network Consensus Diagnosis Results (No Threshold)

Classifier	Correct Response	Incorrect Response	Unknown Response
LVQ and RBF	93%	2%	5%
LVQ and BP	92%	2%	6%
BP and RBF	91%	2%	7%

Although all consensus systems had similar incorrect response rates, the LVQ and RBF consensus system had the highest correct response rate and lowest unknown response rate. The consensus systems all had lower incorrect response rates than the individual networks with threshold values.

Another method of consensus diagnosis that was implemented involved combining the results of three networks--BP, LVQ, and RBF--and then arriving at a consensus diagnosis. A simple comparison of the results of the three networks was made and if they all agree, the non-wheeze or wheeze category was chosen depending on which consensus was reached. If they do not all agree, then the "unknown" category was chosen. In each network, the output unit with the largest value was taken as being the network's classification. The networks had the same architectures as outlined previously and were also trained and tested in the same manner. The results are displayed in Table 7.13.

Table 7.13: Three Network Consensus Diagnosis Results (No Threshold)

Classifier	Correct Response	Incorrect Response	Unknown Response
BP + LVQ + RBF	90%	1%	9%

The results shown in Table 7.13 indicate that combining three network classifiers together results in a slightly lower correct response rate. However, the incorrect response rate is also decreased slightly. It is more important to reduce the incorrect response rate, even if it is at the expense of increasing the unknown response rate slightly, because the data classified as unknown can always be reviewed by another classifier or a physician.

Finally the output values of the combined BP and LVQ networks were also thresholded and compared to formulate a consensus diagnosis. If after thresholding the results, the networks were in agreement then the classification was accepted. Four categories were formed: 1) an agreed correct response category, which contained all the data that was correctly classified by both networks; 2) an agreed incorrect category which contained data that was incorrectly classified by both networks; 3) an agreed unknown category which contained data whose classification did not meet the threshold requirements on both networks; and 4) a disagreed unknown response category which contained data whose classification did not meet the threshold requirements for one of the two neural networks. Threshold values of 0.9 and 0.1 were used and the results are shown in Table 7.14.

Table 7.14: Combined BP and LVQ Results (0.9/0.1)

Data Set	Agreed Correct Response	Agreed Incorrect Response	Agreed **Unknown** Response	Disagreed **Unknown** Response
All Patterns	71%	1%	2%	26%

The incorrect classification rate was reduced to 1% but again as in Section 7.4, at the expense of a significantly lower correct classification rate of only 71% and a significantly higher **unknown** classification rate of 28%.

7.6 Discussion of the Results

The results from the single neural network trials without thresholding indicates that the LVQ and RBF networks were the best networks at classifying the data sets with each having a performance of 95%. The BP and SOM networks had reduced performances of 92% and 91% respectively; however all performances were within 4% of each other, (except for the PNN which used a different representation of the data). It was also noted that it is important to choose the correct network architecture as the architecture may have a large effect on the network's performance. The size of the input vector may be too large and should be experimented with, as was seen in the case of the PNN.

Thresholding the network's results by requiring that the values of the output units be within a certain range reduced the amount of incorrectly classified data; however, the amount of correctly classified data was also reduced. However, it was noted that the data that were classified as unknown could be reviewed by another classifier or a physician. It was also noted that as the thresholds became tighter (ie., allowing less uncertainty), the amount of data being misclassified was reduced; however, the amount of data being correctly classified was also reduced.

Consensus diagnosis involved combining the results of two or more classifiers. The best results for the consensus diagnosis system came from combining three networks--LVQ, RBF, and BP--together. The correct response rate was 90% and the incorrect response rate was decreased to only 1%, with the other 9% classified as unknown. Results of combing a

BP and LVQ network and implementing a threshold method also reduced the incorrect response rate to 1%; however, the correct response rate was only 71%, with 28% of the data classified as unknown.

Our experiments found that the network that minimizes the amount of misclassified data and maximizes the amount of correct data is the consensus diagnosis system that is composed of three networks--BP, LVQ, and RBF--and does not use any thresholding. In the case of thresholding, it was found that as more certainty is required in a diagnosis, the number of incorrect diagnoses decreases; however, so does the number of correct diagnoses. Given this approach, it is important that the system err on the side of being conservative. If a sound is not clearly identifiable as being normal, it should be identified as a possible abnormal sound and placed into the unknown category.

Chapter 8

Conclusion

8.1 Summary of the Results

In Sections 4.1.2.1 and 4.1.2.2 it was found that the existing algorithmic approaches for wheeze detection have a poor performance when analyzing wheezes of low to moderate intensity. **The goal of this research was to evaluate neural networks as a tool for wheeze detection and to construct a suitable representation and neural network system that would be able to classify lung sounds with a high degree of accuracy and minimize the amount of data that is misclassified.**

An examination of a variety of representations for the lung sounds, including the raw signal format, the filtered signal format, and the Fourier Transform (FT) representation was made. After reviewing the results of the various representations, the FT representation was found to return the best results with the data sets used in the experiments.

Determining the best neural network classifier was accomplished by testing five different neural networks--Backpropagation, Self Organizing Maps, Learning Vector Quantization, Probabilistic Networks, and Radial Basis Functions--and comparing the results. The overall performances of each neural network (using the simple classification scheme of determining the network's classification according to which unit had the largest output value) are displayed in Table 8.1. It was found that the LVQ and RBF networks had the best performances. The PNN had a lower performance than the other neural networks; however, it was found that the size of the input vector was too large to adequately process the data.

Table 8.1: Results of Different Neural Network Models

Network	Overall Performance (on Sounds_1.B and Sounds_2)
LVQ	95%
RBF	95%
BP	92%
SOM	91%
PNN	87%

Various parameters such as network architecture and amount of training data were also altered. It was found that the network architecture may have a large affect on performance and so the architecture must be carefully developed. In addition, it was found that training the network with a larger training set improved performance. Also the number of hidden units and the size of the training set increased the training time.

Although all of the neural networks had very good overall performances (with the exception of the PNN, which used a smaller representation of the data), they still misclassified between 5% to 13% of the data depending on which neural network was used. A thresholding technique which required that the output unit's values are within a certain range before a classification can be made, was used to decrease the amount of data that was misclassified. This technique was evaluated with data that the network was not trained and the results are summarized in Table 8.2. It was found that the number of incorrect responses could be decreased by requiring more certainty in a diagnosis. However, this also reduced the number of correct responses.

Table 8.2: Results of Thresholding the Outputs of an RBF Network

Threshold	Correct Response	Incorrect Response	Unknown Response
No Threshold	92%	8%	0%
0.8/0.2 Threshold	84%	3%	13%
0.9/0.1 Threshold	72%	2%	26%

A consensus diagnosis system was also constructed by combining the results of several neural networks and not using any thresholds. If the neural networks agreed on a classification, the classification was accepted. Otherwise, the pattern was placed into an unknown category. Three neural networks--BP, LVQ, and RBF--were combined two at a time and a consensus diagnosis was formulated. A consensus system was also constructed from a BP and RBF network with thresholds applied to the output units. The BP, LVQ, and RBF neural networks were also combined into one large consensus system. The results of the consensus diagnosis systems are outlined in Table 8.3. The best performance was obtained from the three network consensus system and resulted in only 1% of the data being misclassified and a correct response rate of 90%. Thresholding the results of the BP and RBF networks resulted in a very low incorrect response rate; however, the correct response rate decreased substantially.

Table 8.3: Consensus Diagnosis Results (No Threshold)

Classifier	Correct Response	Incorrect Response	Unknown Response
LVQ and RBF	93%	2%	5%
LVQ and BP	92%	2%	6%
BP and RBF (No Threshold)	91%	2%	7%
BP and RBF (Threshold)	71%	1%	26%
BP + LVQ + RBF	90%	1%	9%

In conclusion, the best results of the study were obtained by using the FT spectrum representation with the consensus system constructed from three neural networks--BP, LVQ, and RBF--without using any thresholding.

8.2 Future Work

In general, the accuracy attained by the neural network classifiers was quite high. However, there are several aspects to the classification of lung sounds that require further attention:

- objective definitions for lung sound nomenclature
- having more than one individual classify lung sounds in the training set
- correlating lung volumes and airflow with lung sounds
- finding the ideal representation of lung sounds
- incorporating the algorithmic technique into the consensus system
- increasing the size of the patient population
- single segment monitoring versus continuous monitoring
- classification of more complex lung sounds
- eliminating extraneous environment noise from recordings
- incorporation of the system into a clinical setting

First, as was demonstrated in Chapter 2, the lung sound nomenclature is vague and there is still much disagreement on how to classify adventitious lung sounds. A question regarding what adventitious lung sounds shorter than 100 msec, but longer than 30 msec sound be called was posed to Dr. R. L. H. Murphy, founder of the International Lung Sound Association. His reply summed up some of the shortcomings of lung sound classification. He said "the science [of lung sound classification] is an emerging one. There are only guidelines, not reliable mathematical formulas to classify these sounds" (Murphy, 1995). The lung sound nomenclature should be re-examined, by experienced respirologists, especially continuous lung sound nomenclature, so that scientists may have objective definitions with which to work.

The research presented in this thesis relied on data that was classified by a single individual. It may be best to have a separate individual review the data in the training sets so that incorrect data is not used to train the neural network. It may even be necessary to have several individuals review the data and utilize a type of consensus classification system as was used in Chapter 7. Having incorrect data within the training set of the neural network may cause incorrect classifications.

In the future, an attempt should be made to correlate lung sounds with lung volume and airflow information into the neural network's data. It has been noted that if patient's do not have a critical flow rate, no wheeze sound will be produced.

Other than re-examining the lung sound nomenclature, there is also much work that still needs to be done specifically with neural network processing. First, the representation of the lung sounds should be further researched. Our findings support the view that the FT spectrum provides a good representation. However, the number of data values used to represent the lung sounds is large (247 data values). In the case of the PNN, performance improved when the size of the input vector was reduced. The dimension of the input vector could be reduced further by: 1) using principle components analysis; 2) using an LVQ network to identify the dominant characteristics of the spectrum and only processing these characteristics; or 3) summing or averaging several adjacent input values. In addition, the wavelet representation should be compared with the FT representation. Also, all input values for the neural network were scaled linearly to avoid having large values that might saturate the network. A logarithmic scaling scheme should also be tested and the results compared. As mentioned in Section 4.1.1, a stationary-nonstationary filter for crackle detection was also able to detect short wheezes in the nonstationary output. This process should be further investigated as a method for detecting wheezes of longer duration.

Research should also be performed which examines the results of combining a neural network classifier and an algorithmic classifier into a hybrid consensus system. The algorithmic classification scheme was previously found to have good results when tested with loud wheezes of high intensity. Using a hybrid system constructed of a neural network and an algorithmic scheme may provide a superior classifier. Further research should also be performed with the stationary/non-stationary filter of Section 4.1.1.

Our results support the hypothesis that training the network with different types of lung sounds provides a better classification rate. One of the next steps in this research is to maintain the same high level of performance when testing the system with data from a very large patient population, as might be encountered in a clinical setting. This endeavor will require a larger training set and a larger test set composed of lung sounds taken from patients with a wide range of lung sounds.

In the research presented in this thesis, wheezes were identified within isolated data segments. However, in practice, a physician may have several data segments from the same patient. These segments should be processed and the results accumulated so that the number of wheeze classifications per unit time can be accurately measured.

The networks examined in this research were trained to identify only wheezes, and not other adventitious lung sounds such as crackles. The network should be tested in situations where wheeze lung sounds are mixed with crackle sounds. It is believed that crackle detection will require a separate network that has been trained specifically with crackles. The networks could then be joined together to formulate a consensus diagnosis. In addition, it was mentioned that wheezes produced at the trachea are different from wheezes produced at the chest. If wheezes at both locations are to be monitored, it may be necessary to have customized neural networks for each location. In addition, it will also be important to treat neonatal wheezes differently than the wheezes of older patients. It may be necessary to re-evaluate the different types of neural networks for each lung sound, as each network may have its own advantages at analyzing the various types of lung sounds.

The system should also be capable of analyzing breath sounds within a noisy environment. This may require better shielded microphones and adaptive filter processing techniques that eliminate constant background noise. Spontaneous background noise may be reduced slightly by processing only information within the bandwidth where lung sounds are likely to be found. It may also be necessary to perform a breath to breath comparison of lung sounds so that noise can be more easily eliminated.

Finally, there are still many questions that must be answered before a lung sound classifier, such as the one presented in the research, can be incorporated into a clinical setting. One of the main questions is how this system can be validated. It is suggested that such a system could be first evaluated on several cases and scenarios and then modified as the need arises. The final tests should be carried out by researchers distant from the research project to ensure that no bias is introduced. The most likely use of the system would be to perform an initial screening to identify abnormal lung sounds. These sounds would then be analyzed by an experienced physician and a diagnosis would then be made by the physician. It will be important that the system err on the side of being conservative and that lung sounds that can not be categorized with a large amount of certainty be categorized as unknown. The amount of certainty and system testing required before the system can be used in a clinical setting is still a question that remains to be answered.

8.3 Other Applications of the Research

There are many applications of this research in addition to the wheeze analysis applications mentioned in Section 4.3. Within the field of respirology, the neural networks could be adapted to classify crackles by altering the training and test sets. It will also be necessary to find the best representation for a crackle, which may be different from the FT spectrum representation chosen for wheezes. The system can also be modified to detect snoring sounds and used in applications that help patients stop snoring.

The thresholding neural networks and consensus diagnosis systems constructed in this research can be modified and used in almost any pattern recognition problem. The systems constructed can be used on almost any application where the input patterns and associated classifications can be quantified in a numerical or ordinal representation. For example, with some minor modifications, this research can be applied to EKG classifications, underwater acoustics problems, or some of the various image processing applications.

It is hoped that the results of this research not only further the field of respirology, but also the multi-faceted field of signal classification.

8.4 Author's Closing Comment

A review of the lung sound nomenclature and neural networks has been presented in the early chapters of this research so that other researchers may continue work in this field. This project is only one of the many examples of the results that may be attained when computer scientists and medical doctors work together to integrate computerized technology into the medical system. Computer science is a relatively young discipline; medicine is a very old discipline. However, the combination and integration of the two disciplines is just in its infancy. It is only through continued cooperation by both groups of people that this integration will continue to be successful. It was the intent, and remains the hope, that the research presented in this thesis will provide a contribution to this endeavor.

Appendix A

Fast Fourier Transform

A.1 Fast Fourier Transform Algorithm

The Fast Fourier Transform Algorithm outlined in Section 3.2.3 is described below.

The algorithm is passed the number of complex data points, nn (which must be a power of 2), and isign, which should be set to +1. The algorithm is also passed the sampled points in a complex number format. (Complex numbers are composed of a real and an imaginary part. The real and imaginary parts occupy two consecutive positions in the array data[ ], with the real part always occurring before the imaginary part.) The real part of the complex number is used to store the sampled data values. The imaginary part is not used when computing the FT, and so it is zeroed. The algorithm computes the real and imaginary parts of the FT in the array data[ ]. The array is set up such that data[1] and data[2] contain the real and imaginary parts of f_0. Only the values up to data[(nn/2) - 1] and data[(nn/2)] are unique. The values in the other half of the array represent the mirror image of the first half of the array. For a more detailed explanation, the reader is referred to Chapter 12 of (Press et al.).

```
{                       Fast Fourier Transform Algorithm: In Pascal                    }
{data[ ] = the data for which the FT is to be computed.  Data[ ] will contain the FT   }
{            representation when the algorithm is finished processing.                  }
{nn       = 2 * the number of data pts sampled and used to compute the FT.              }
{            NOTE: One complex number has both a real and imaginary componet;          }
{                       each stored at consecutive array positions.                     }
{isign    = +1 when performing a FT, and -1 when performing an inverse FT.              }
{            NOTE: You must also divide each FT coefficient by nn/2 to compute          }
{                       the inverse FT.                                                 }
{DataType: Array [1..nn] of real                                                        }


procedure Four1(Var data:DataType;
                      nn,isign: integer);
var
  ii,jj,n,mmax,m,j,istep,i:integer;
  wtemp,wr,wpr,wpi,wi,theta:double;
  tempr,tempi,wrs,wis:real;

Begin
  n:= 2*nn;
  j:= 1;
  {Bit reversal routine}
  for ii:= 1 to nn Do Begin
    i:= 2*ii-1;

    if j > i then Begin
      tempr:= data[j];
      tempi:= data[j+1];          {Exchange the two complex numbers}
      data[j] := data[i];
      data[j+1]:= data[i+1];
      data[i]:=tempr;
      data[i+1]:= tempi;
    end; {if j > i}

    m := n DIV 2;

    while (m >= 2) and (j > m) Do Begin
      j:= j-m;
      m:= m DIV 2;
    end; {While}

    j:= j+m;
  end; {For}

  mmax:=2;
  {Start the Danielson-Lanczos routine}
  while n>mmax Do Begin          {Outer loop executed log2nn times}
    istep:= 2 * mmax;
    theta:= 6.28318530717959/(isign*mmax); {init for trig recurrence}
    wpr:= -2.0*sqr(sin(0.5*theta));
    wpi := sin(theta);
    wr:= 1.0;
    wi := 0.0;
```

```
For ii := 1 TO mmax DIV 2 Do Begin      {Here are the two nested loops}
 m:= 2*ii-1;
 wrs:= wr;
 wis := wi;

 FOR jj := 0 TO (n-m) DIV istep DO BEGIN {This is the Danielson-Lanczos}
  i:= m+JJ*istep;                  {formula}
  j:= i+mmax;
  tempr := wrs*data[j]-wis*data[j+1];
  tempi := wrs*data[j+1]+wis*data[j];
  data[j] := data[i]-tempr;
  data[j+1]:= data[i+1]-tempi;
  data[i]:= data[i]+tempr;
  data[i+1] := data[i+1]+tempi;
 end; {for}

 wtemp:= wr;                    {Trigonometric recurrence}
 wr:= wr*wpr-wi*wpi+wr;
 wi:= wi*wpr+wtemp*wpi+wi;
end; {for}

mmax := istep;
end; {while}

end; {Procedure}
```

(Adapted from Press et al., pp 435-436.)

Appendix B

Measures of Performance

B.1 Calculation of the Performance Values

The results for the various lung sound classifiers presented in this thesis use different measures of performance. In the experiments where only two classifications were possible (ie., wheeze or non-wheeze), specificity, sensitivity, and overall performance were used and their definitions are provided in Section B.2. In instances where three classifications were possible (ie., correct, incorrect, or unknown), performance was calculated with the method outlined in Section B.3.

B.2 Calculating Performance in a Two-Class System

In the experiments that had a two-class classification (ie., no unknown category was included), the characteristic of interest was the presence of a wheeze. A lung sound segment may either contain a wheeze segment, in which case a **positive** event has occurred; or a lung sound segment may not contain a wheeze, in which case a **negative** event has occurred. The classification may be **true** (ie., correct), or it may be **false** (ie., incorrect). If a lung sound that contains a wheeze is correctly classified as a wheeze, the classification is said to be a **True Positive**. Similarly, if a lung sound that does not contain a wheeze is correctly classified as not containing a wheeze, the classification is said to be a **True Negative**. Conversely, if a lung sound contains a wheeze and is incorrectly classified as not containing a wheeze, the classification is termed a **False Negative**. If a lung sound that does not contain a wheeze is incorrectly classified as containing a wheeze, the classification is termed a **False Positive**. Table B.1 outlines the possible results of a two-class classification experiment.

Table B.1: Possible Results of a Two-Class Classification Scheme

		Actual Classification	
		Class Positive (contains a wheeze)	Class Negative (no wheeze present)
Predicted Classification	Positive Predicition (contains a wheeze)	True Positive	False Positive
	Negative Prediction (no wheeze present)	False Negative	True Negative

(Table B.1 adapted from Weiss, p 19.)

The **specificity** of a classifier is defined as the percentage of segments without the characteristic that have a negative result. In this thesis, a high specificity indicates that the classifier is very good at classifying **non-wheeze** segments (Kotz and Johnson, pp 370-371). Specificity may be calculated with the following formula:

$$\textbf{Specificity} = \frac{\text{(Number of True Negatives)}}{\text{(Number of True Negatives)} + \text{(Number of False Positives)}}$$

The **sensitivity** of a classifier is defined as the percentage of segments with the characteristic that have a positive result. In this thesis, a high sensitivity indicates that the classifier is very good at detecting **wheezes** (Kotz and Johnson, pp 370-371). Sensitivity may be calculated with the following formula:

$$\textbf{Sensitivity} = \frac{\text{(Number of True Positives)}}{\text{(Number of True Positives)} + \text{(Number of False Negatives)}}$$

The **overall performance** of a classifier is defined as the percentage of correct classifications. Overall performance may be calculated with the following formula:

$$\textbf{Overall Performance} = \frac{\text{(Number of Correct Responses)}}{\text{(Number of Correct Responses)} + \text{(Number of Incorrect Responses)}}$$

B.3 Calculating Performance in a Three-Class System

In the three-class systems used in this research, three results were possible: correct, incorrect, or unknown. Each classification had its associated response rate which was calculated as the number of values within the response category divided by the summation of the number of values in the three possible categories. For example, unknown response rate was calculated as follows:

$$\textbf{Unknown Response Rate} = \frac{\text{Number of Unknown Responses}}{\text{Number of Correct Responses} + \text{Number of Incorrect Responses} + \text{Number of Unknown Responses}}$$

References

American Thoracic Society. "Definitions and Classifications of Chronic Bronchitis, Asthma, and Pulmonary Emphysema," *American Review of Respiratory Disorders*, vol. 85, 1962, pp 762-768.

American Thoracic Society Ad Hoc Committee on Pulmonary Nomenclature. "Updated Nomenclature For Membership Reaction," *ATS News*, Fall 1977; 3 pp 5-6.

Asher T., Sanchez I., and Pasterkamp, H. "Respirosonography in Infants With Acute Bronchiolitis," *American Journal of Diseases of Children*, vol. 145, Dec. 1991, pp 1405-1410.

Bates B., Bickley L., and Hoekelman R. *A Guide to Physical Examination and History Taking 6th ed.*, Philadelphia: J. B. Lippincott Company, 1992.

Baughman R. P. and Loudon R. G. "Wheezing in Asthma and Its Response to Bronchodialators: Waveform Analysis Using Fast Fourier Transform Technique," *American Review of Respiratory Disease*, 1982; vol. 125, p 96.

Baughman R. P. and Loudon R. G. "Quantitation of Wheezing in Acute Asthma," *Chest*, Nov. 1984; 86:5, pp 718-722.

Berg, R. E. and Stork D. G. *The Physics of Sound*, New Jersey: Prentice-Hall, 1982.

Bortolan G., Degani R., and Willems J. L. "ECG Classication with Neural Networks and Clustering Analysis," *IEEE Computers in Cardiology Proceedings*, 1991, p 177.

Card H. C. Personal communication with Dr. Howard Card of the University of Manitoba, Winnipeg, Manitoba, Department of Engineering, via the internet. August 28, 1995.

Chawla A. and Ercal F. *Diagnosis of Malignant Melanoma Using a Neural Network, MSc Thesis,* Department of Computer Science, University of Missouri - Rolla, CSc-93-06, 1993.

Cohen A. *Biomedical Signal Processing,* vol. 1, Florida: CRC Press, 1986.

Cohen A. and Landsberg D. "Analysis and Automatic Classification of Breath Sounds," *IEEE Transactions on Biomedical Engineering,* Sept. 1984, pp 585-590.

Cottrell G. W., Munro P., and Zipser D. "Learning Internal Representations From Grayscale Images: An Example of Extensional Programming," *Proceedings of the Ninth Annual Conference of Cognitive Science Society*, 1987.

Dodge R. R. and Burrows B. "The Prevalence and Incidence of Asthma and Asthma-like Symptoms in a General Population Sample," *American Review of Respiratory Disease*, 1980; vol. 122, pp 567-575.

Fenton T., Pasterkamp H., Tal A., and Chernick V. "Automated Spectral Characterization of Wheezing in Asthmatic Children," *IEEE Transactions on Biomedical Engineering*, 1985; 32, pp 50-55.

Fitzgerald, F. T. Physical Diagnosis Versus Modern Technology. A Review," *J. Med.* 1990; 152, pp 377-382.

Forgacs, P. "The Functional Basis of Pulmonary Sounds," *Chest*, March 1978; 73, pp 399-405.

Forgacs P. *Lung Sounds*, London: Baillière Tindall, 1978.

Godfrey S., Edwards R. H. T., Campbell E. J. M., Armitage P., and Oppenheimer EA. "Repeatability of Physical Signs in Airways Obstruction," *Thorax*, 1969; 24, pp 4-9.

Gorman R. and T. J. Sejnowski. "Analysis of Hidden Units in a Layered Network to Classify Sonar Targets," *Neural Networks*, 1988; 1(1), pp 75-89.

Grotberg J. and Gavriely N. "Flutter in Collapsible Tubes: A Theoretical Model of Wheezes," *Journal of Applied Physiology*, 66(5), 1989, pp2262-2273.

Hale-White W. *Translation of Selected Passages from De l'Auscultation Médiate by R. Théophile H. Laennec*, London: John Bale, Sons and Danielsson, 1923.

Hertz J., Krogh A., and Palmer R. *Introduction to the Theory of Neural Computation*, Massachusetts: Addison-Wesley Publishing Co., 1991.

Holford S. K. *Discontinuous Adventitious Lung Sounds: Measurement, classification and modeling. Doctoral Thesis,* Cambridge: Massachusetts Institute of Technology, 1982.

Hopp R. J, Againdra B. K., Nair N. M., and Townley R. G. "Specifity and Sensitivty of Methacholine Inhalation Challenge in Normal and Asthmatic Children," *J. Allergy Clin. Immunol.*, 1984;74, pp 154-158.

Horenblas R. "Stopping the Brain Drain," *Graduate Computer World*, Fall 1992.

Kaisla T., Sovijärvi A., Piirilä P., Rajala H. M., Haltsonen S., and Rosqvist T. "Validated Method for Automatic Detection of Lung Sound Crackles," *Medical and Biological Engineering and Computing*, Sept 1991; vol. 29, pp 517-521.

Kalayci T. Çelebi G., Öztürk Y., and Özhan M. "Automatic Detection and Classification of Crackles by Using a Neural Network," *1992 IEEE Engineering in Medicine and Biology 14th Annual Conference*, Paris France, Oct. 29-Nov. 1, 1992; vol. 6, pp 2580-2581.

Kallio K., Rosqvist T., Haltsonen T., Karp S., Katila T., Paajanen E., Piirilä P., and Sovijärvi A. R. A. "Classifcation of Lung Sounds by Using Self-Organizing Feature Maps," *ILSA Proceedings*, New Orleans: 1990. (Abstract)

Karp P., Katila T., Paajanen E., Piirilä P., and Sovijärvi A. "Effect of Highpass Filtering on the Original Waveform of Lung Sound Crackles," *Annual International Conference of the IEEE Engineering In Medicine and Biology Society*, 1991; vol. 13, no. 1, pp 383-384.

Kotz S. and Johnson N. L. *Encyclopedia of Statistical Sciences*, vol. 8, New York: John Wiley & Sons, 1988.

Korn G. *Desire/Neunet Reference Manual*, G. A. and T. M. Korn Inc, 1994.

Kryger M. *Introduction to Respiratory Medicine, 2nd ed.,* New York: Church Livingstone, 1990.

Loudon R. G. "Auscultation of the Lung," *Clin Notes Resp Dis*, 1982; vol. 21, pp 3-7.

Loudon R. G. "Clinical Application of Wheeze Analysis," *Archives of Chest Disease*, 1993; 48,(5), pp 583-585.

Loudon R. G and Murphy, R. L. H. Jr. "State of the Art. Lung Sounds," *American Review of Respiratory Disease*, 1984;130[2] pp 663-673.

Luger G. and Stubblefield W. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 2nd ed.,* California: Benjamin/Cummings Publishing Co, 1992.

Manber U. *Introduction to Algorithms: A Creative Approach*, Massachusetts: Addison-Wesley, 1989.

McCulloch W. S. and Pitts W. A. "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bull. Math Biophys,* 1943; 5, pp 115-133.

McNichol K. N. and William H. B. "Spectrum of Asthma in Children. I. Clinical and Physiological Components," *British Medical Journal*, 1973; 4, pp 7-11.

Mikami R., Murao M., Cugell D., Chrétien, J., Cole P., Meier-Sydow J., Murphy R., and Loudon, R. "International Symposium on Lung Sounds", *Chest*, August, 1987, pp 342-345.

Moore K. L. *Clinically Oriented Anatomy 3rd ed.,* Baltimore: Williams & Wilkins, 1992.

Murphy R. L. H. Jr. "Discontinuous Adventitious Lung Sounds," *Seminars in Respiratory Medicine*, Jan. 1985; vol. 6, no. 3, pp 210-219.

Murphy R. L. H. Jr. Personal communication with Dr. Raymond L. H. Murphy, Faulkner Hospital, Pulmonary Service, Boston MA, via the internet. July 21, 1995.

Mussell M. J. "The need for standards in recording and analyzing respiratory sounds," *Medical and Biological Engineering and Computing*, March 1992; 30, pp 129-139.

Murray J. and Nadel J. *Textbook of Respiratory Medicine*, Pennsylvania: W. B. Saunders Co., 1994.

NeuralWare. *Neural Computing: A Technology Handbook for Professional II/Plus and NeuralWorks Explorer*, Pittsburgh: NeuralWare, 1993.

Ng K. C. and Abramson B. "Consensus Diagnosis: A Simulation Study," *IEEE Transactions on Systems, Man, and Cybernetics*, Sept./Oct. 1992; vol. 22, no. 5, pp 916-928.

Ono M., Arakawa K., Mori M., Sugimoto T., and Harashima H. "Separation of Fine Crackles from Vesicular Sounds by a Nonlinear Digital Filter," *IEEE Transactions on Biomedical Engineering*, Feb. 1989; vol. 36, no. 2, pp 286-291.

Oppenheim A. and Willsky A. *Signals and Systems*, London: Prentice-Hall, 1983.

Pasterkamp H., Wiebicke W., and Fenton R. "Subjective Assessment vs. Computer Analysis of Wheezing in Asthma," *Chest*, March, 1987; pp 376-381.

Pasterkamp H., Montgomery M., and Wiebicke W. "Nomenclature Used by Health Care Professionals to Describe Breath Sounds in Asthma," *Chest*, August, 1987; pp 346-352.

Pasterkamp H., Carson C., Daien D., and Oh Y. "Digital Respirosonography: New Images of Lung Sounds," *Chest* Dec. 1989; vol. 96, pp 1405-1412.

Pasterkamp H., Kraman S. S., DeFrain P. D., and Wodicka G. R. "Measurement of Respiratory Acoustical Signals: Comparison of Sensors", *Chest*, Nov. 1993; vol. 104, pp 1518-1525.

Pomerleau D. "ALVINN: An Autonomous Land Vehicle in a Neural Network," *Advances in Neural Information Processing Systems I*, ed. D. Touretzky, San Mateo, CA: Morgan Kaufmann, 1989, pp 305-313..

Press W. H., Flannery B. P., Teukolsky S. A., and Vetterling W. T. *Numerical Recipes in Pascal*, Cambridge: Cambridge University Press, 1989.

Rajala H. M., Kallio K., Karp P., Katila T., Paajanen E., Piirilä P., Rosqvist T., and Sovijärvi A. R. A. "Comparison of Methods for Automatic Classification of Respiratory Sounds," *ILSA Proceedings*, Helsinki, 1992;53, (Abstract).

R.A.L.E. *RALE Educational Program: Lung Sounds*, Winnipeg, Manitoba, Canada: PixSoft Corporation, 1990.

Rubin E. H. and Rubin M. *Thoracic Diseases*, Philadelphia and London: W.B. Saunders, 1961.

Rumelhart D., McClelland J., and The PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1* Massachusetts: MIT Press, 1986.

Sanchez I., Powell R. E., and Pasterkamp H. "Wheezing and Airflow Obstruction During a Methacholine Challenge in Children with Cystic Fibrosis and in Normal Children," *American Review of Respiratory Disease*, 1993; vol. 147, pp 705-709.

Sanchez I., Avital A., Wong, I., Tal A., and Pasterkamp H. "Acoustic versus Spirometric Assessment of Bronchial Responsiveness to Methacholine in Children," *Pediatric Pulmonology*, 1993; vol. 15, pp 28-35.

Scuse D. *74.719 Lecture Notes: Neural Networks*, University of Manitoba, Fall 1993.

Scuse D. *74.720 Lecture Notes: Advanced Topics in Neural Networks*, University of Manitoba, Spring 1995.

Seeley R. and Stephens T. *Anatomy and Physiology*, St. Louis: Times Mirror/Mosby College Publishing, 1989.

Shapiro S. *Encyclopedia of Artificial Intelligence, 2nd ed.*, New York: John Wiley & Sons, 1992.

Smith J. M. and Knowler L. A. "Epidemiology of Asthma and Allergic Rhinitis. I. In a Rural Area," *American Review of Respiratory Disease* 1965; 91 pp 16-38.

Smyllie H. C., Blendis L. M., and Armitage P. "Observer disagreement in Physical Signs of the Respiratory System," *Lancet*, 1965; 2, pp 412-13.

Stein R., "Preprocessing Data for Neural Networks," *AI Expert*, March 1993, pp 32-37.

Van Heuvelen A. *Physics. A General Introduction, 2nd ed.*, Boston: Little, Brown and Company, 1985.

Wasserman P. *Advanced Methods in Neural Computing*, New York: Van Nostrand Reinhold, 1993.

Wasserman P. *Neural Computing. Theory and Practice*, New York: Van Nostrand Reinhold, 1989.

Wilson J., Braunwald E., Isselbacher K., Petersdorf R., Martin J., Fauci A., and Root R. *Harrison's Principles of Internal Medicine*, vol. 2, New York: McGraw-Hill Inc, 1991.

Weiss S. M. *Computer Systems That Learn: Classification and Prediction Methods from Statistics and Neural Nets*. San Mateo: M. Kaufmann Publishers, 1991.

Xu J., Chen Q., Zhang Y., and Liu S. "Spectrum Analysis of Lung Sounds," *IEEE Engineering in Medicine and Biology Society 11th Annual International Conference*, 1989, pp 1676-1677.

Xu L., Krzyzak A., and Suen C. Y. "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, May/June 1992; vol. 22, no. 3, pp 418-433.