Advanced signal processing methods for analysis of respiratory sounds

Jin, Feng

2010

Jin, F. (2010). Advanced signal processing methods for analysis of respiratory sounds.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/36292>

<https://doi.org/10.32657/10356/36292>

Advanced Signal Processing Methods for Analysis of Respiratory Sounds

JIN FENG

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2010

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me and has not been submitted for a higher degree to any other University or Institute.

.....

Date

.....

JIN FENG

*To Dad and Mom,
for their encouragements and love.*

Abstract

Respiratory sounds (RSs) auscultation using stethoscope is often the first noninvasive resource available to physicians for the detection and analysis of respiratory ailments. Motivated by the subjectivity of human auditory system, computerized auscultation combined with signal processing techniques have boosted the diagnostic capabilities of RS. This PhD thesis concentrates on the advanced signals processing methods for RS analysis from several aspects including: RS denoising by means of heart sounds (HSs) cancellation (Part I), respiratory phase detection (Part II), as well as RS analysis and feature extraction methods for RS classification (Part III).

Part I solves the problem of HS cancellation through a two-steps scheme consisting of HS localization and HS removal from different types of single-channel RS recordings. HS contaminated segments have been localized by introducing a wavelet based localization scheme. A detection function based on multiscale decomposition is calculated and HS segments in the noisy RS signal are localized based on the cumulative sums of likelihood ratios capturing the dynamic behaviour of the signal. HS removal from RS is then addressed by proposing a novel semi-blind single-channel source extraction algorithm which incorporates the filter banks and template based matching using FIR filters. It has shown that the proposed methods are successful attempts to solve the clinical application challenge faced by the existing HS cancellation methods in terms of respiratory ailments.

In Part II, the problem of noninvasive automatic respiratory rate monitoring by acoustical means is addressed first. Respiratory phase segmentation is then formulated as an optimization problem and the boundaries of the respiratory phase segments are

detected using a multi-population genetic algorithm (MPGA). A new annotating index for the identification of inspiratory and expiratory phases from pathological respiratory sounds (PRSs) is lastly proposed to complete the reliable respiratory phase detection method, targeting irregular breathing for pathological objects.

Part III focuses on the time-frequency (TF) analysis and classification of RS. Identification of nonlinear parts in RS using third-order cumulant has been firstly performed followed by the analysis in TF domain using a novel nonlinear analysis method based on optimally weighted Wigner-Ville distributions (WVDs) of the weighted subband signals from a filter bank. The existence of nonlinearities is quantified and the relationship between the degrees of nonlinearity and RS signal types has been discussed. A new set of features based on the temporal characteristics of the filtered narrow band signals are finally proposed to facilitate accurate RS classification into normal and continuous adventitious types. New features are extracted from the proposed discriminating function based on the autoregressive (AR) averaging, the recursively measured instantaneous kurtosis along frequency and time, as well as the mean distortion between sample entropy (*SampEn*) histograms over the selected frequency bins. The presented features are then evaluated and compared with the existing features using a modified clustering index with different distance metrics.

Acknowledgements

In the first place I would like to express my at most gratitude to my supervisor, Dr. Farook Sattar, for his supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences through out the work. Above all and the most needed, he provided me unflinching encouragement and support in various ways. His truly scientist intuition has made him as a constant oasis of ideas and passions in science, which exceptionally inspire and enrich my growth as a student and a researcher want to be.

I gratefully acknowledge Dr. Goh Yam Thiam Daniel from National University Hospital, for his advice and for providing the enlightening discussions and comments with his abundant knowledge to improve my works from medical point of view. His thoughtful comments and encouragement on my research are highly appreciated. Many thanks go in particular to Dr. Irene Melinda Louis. I am much indebted to Irene for her valuable advice in discussion, using her precious times to complete the data collection. I would also acknowledge Siti Dahlia and Maheswari Arumugam for their advice and their willingness to share their knowledge and experience in handling medical equipments with me. Without their help, the thesis could never be finished.

I gratefully thank Dr. Sirajudeen Gulam Razul for contributing useful discussion regarding the use of Markov Chain for respiratory phase identification, and his thoughtful comments and encouragement on my research are highly appreciated. I also want to thank Dr. Gan Woon Seng for his comments and generous help during the thesis revision.

Special thanks to Ding Huijun, Shen Minmin, and Luo Xuwen for their friendships

and countless helps during my whole graduate studies. My memorable days of life in NTU can never be complete without them. Collective and individual acknowledgments are also owed to my colleagues at Nanyang Technological University whose present somehow perpetually refreshed, helpful, and memorable. I would also like to express my appreciation to Ajay Kumaraswamy Kattepur and Smita Shekar for their kind helps and enjoyable discussions in a wide range. And I would like to thank all the technical staff in the Media Technology lab and Digital Signal Processing lab.

I am so grateful to Chen Xu, Lee Kunqi, Goh Sihui, Zhao Hui, and Jiang Dafei for their encouragements and supports all the time. Lastly, my family deserve special mention for their inseparable supports and blessings. My dad, mom, and grandma in the first place, are the important ones who put the fundament my learning character, showing me the joy of intellectual pursuit ever since I was a child. Their dedication, love and persistent confidence in me is the reason to keep me moving forward all the time. Qu Bo, thanks for being supportive, caring, and encouraging always.

Contents

Abstract	i
Acknowledgements	iii
List of Figures	x
List of Tables	xiv
List of Abbreviations	xvi
List of Symbol	xviii
1 Introduction	1
1.1 Motivation	2
1.2 Objectives	4
1.3 Thesis Contribution	5
1.4 Thesis Organization	6
2 Research Background	8
2.1 Historical Overview of Respiratory Sound Analysis	8
2.2 Anatomy and Physiology of Respiratory System	10
2.3 Respiratory Sound Characteristics and Categorization	12
2.4 Existing Techniques for Respiratory Sound Analysis	17
2.4.1 Heart Sound Cancellation Methods for Comparison	17
2.4.2 Respiratory Phase Detection Methods for Comparison	22

2.4.3	RS Feature Extraction Methods for Comparison	24
2.5	Respiratory Sounds Recording	26
2.5.1	Subject Conditions and Procedures	27
2.6	Recording Setup	28
Part I: Heart Sound Cancellation From Respiratory Sounds		30
3	Wavelet Based Approach for Heart Sound Localization	32
3.1	Introduction	32
3.2	Methodology	34
3.2.1	Data	34
3.2.2	Signal Model	36
3.2.3	Calculation of Cumulant	36
3.2.4	The Wavelet Based Localization Scheme	40
3.3	Results and Discussion	47
3.3.1	Results on Synthesized Data	47
3.3.2	Results on Real Data	49
3.4	Chapter Conclusion	51
4	Filter Bank Based Source Extraction for Heart Sound Removal	53
4.1	Introduction	53
4.2	Methodology	54
4.2.1	Signal Characteristics	54
4.2.2	Basic Concept	55
4.3	Algorithm Implementation	57
4.3.1	Outline of the Proposed Algorithm	57
4.3.2	The Proposed Template Design Scheme	58
4.3.3	The Implementation of Shelving Filter	61
4.3.4	The Proposed Signal Extraction Scheme	61
4.4	Experimental Results and Discussion	65
4.4.1	Experimental Dataset	65
4.4.2	Analysis of γ_{opt}	66

4.4.3	Auditory Test	68
4.4.4	Quantitative Analysis	68
4.5	Chapter Conclusion	73
Part II: Acoustical Respiratory Phase Detection		75
5	Respiratory Phase Segmentation Using Genetic Approach	78
5.1	Introduction	78
5.2	Methodology	79
5.2.1	Data Acquisition	80
5.2.2	Signal Preprocessing	81
5.2.3	Respiratory Rate Estimation	85
5.2.4	Sample Entropy	87
5.2.5	Multi-population Genetic Algorithm (MPGA)	89
5.2.6	Parameter Setting	92
5.3	Results and Discussion	97
5.3.1	Respiratory Rate Estimation	97
5.3.2	Respiratory Phase Segmentation	100
5.4	Chapter Conclusion	103
6	Annotating Index for Respiratory Phase Identification	104
6.1	Introduction	104
6.2	The Proposed Annotating Index	106
6.2.1	Mixing Index (MI)	106
6.2.2	The Choice of Frequency Bins	108
6.2.3	The Choice of Window Length	109
6.3	Analysis and Discussion	110
6.3.1	Analysis on the Choice of $K2$	110
6.3.2	Analysis on the Choice of N_w	112
6.4	Results and Comparison	114
6.4.1	Experimental Dataset	114
6.4.2	Comparison With Time-Frequency Domain Features	114

6.4.3	Comparison with Wavelet Domain Feature	116
6.5	Chapter Conclusion	118
Part III: Time-Frequency Analysis and Classification of Respiratory Sounds		119
7	Tracking and Analysis of Nonlinearity in Respiratory Sounds	122
7.1	Introduction	122
7.2	Methodology	123
7.2.1	Experimental Dataset	123
7.2.2	Nonlinearity Tracking	123
7.2.3	The Nonlinear Analysis Method	125
7.3	Results and Discussion	128
7.4	Chapter Conclusion	131
8	Feature Extraction Methods for Respiratory Sound Classification	132
8.1	Introduction	132
8.2	Methodology	134
8.2.1	Data Acquisition	134
8.2.2	Time-Frequency Analysis	135
8.2.3	New Feature Set Based on Instantaneous Kurtosis	136
8.2.4	New Feature Set Based on Discriminating Function	140
8.2.5	New Feature Set Based on Sample Entropy (<i>SampEn</i>)	145
8.2.6	Histogram Distortion Calculation	146
8.2.7	Evaluation of The Proposed Feature Sets	147
8.3	Results and Discussion	149
8.3.1	Experimental Dataset	149
8.3.2	Separability Index (<i>SI</i>)	150
8.3.3	Classification Results	151
8.4	Chapter Conclusion	155
9	Conclusion and Recommendations	156
9.1	Summary and Concluding Remarks	156

9.2	Future Works	158
9.2.1	Issues for HS Cancellation	158
9.2.2	Issues for Flow Volume Estimation	159
9.2.3	Issues for RS Classification	159
9.2.4	Applications	160
Appendices		162
Appendix A: Implementation of GLRT		163
Appendix B: Time Scaling Approach for Source Extraction Scheme		166
Appendix C: Sample Entropy Calculation		168
Appendix D: Computation of Gabor Transform		169
Appendix E: First-Order AR Averaging		170
Author's Publications		171
Bibliography		173

List of Figures

2.1	Anatomy of the respiratory system and the zoomed in picture of a bronchiolus branch and alveolar ducts (reprint from [30]).	11
2.2	A typical LS signal spectrogram (left) and a typical TBS spectrogram (right) (reprint from [39]).	13
2.3	Waveform and spectrogram of a typical wheeze sound (reprint from [39]).	14
2.4	Waveform and spectrogram of a typical stridor sound (reprint from [39]).	15
2.5	Waveform and spectrogram of the typical rhonchi sound (reprint from [39]).	15
2.6	Waveform and spectrogram of typical squawk (reprint from [39]). . . .	15
2.7	Waveform and spectrogram of typical pleural sound (reprint from [39]).	16
2.8	Waveform and spectrogram of typical fine crackle (top) and coarse crackle (bottom) (reprint from [39]).	16
2.9	Histograms of the RS segments including HS (solid curve) and void of HS (dashed curve) (reprint from [40]).	19
2.10	Block representations of a 3 level SWT decomposition of $x(n)$ (above) and the upsampling process of the SWT filters (below).	20
2.11	(a) Waveform of the original RS recording. (b)-(d) Approximation coefficients, first to third scales. (e) Multiscale products of first and second scales. (f) Multiscale products of first through third scales (reprint from [7]).	21
2.12	The measurement scales $\delta t = 32, 64, 128$ for D_σ calculation (reprint from [43]).	23
2.13	Apparatus setup for RS recording.	28
3.1	(a) An overall block scheme of the proposed HS localization method; (b) The detailed steps involved in the wavelet based localization scheme for linear/nonlinear parts of the noisy RS signal.	35

3.2	(a) The proposed nonlinear identification function for a real RS signal; (b) Original waveform of the RS signal with its respective linear (dashed line)/ nonlinear (solid line) parts being labeled manually.	39
3.3	(a) Decision function $\mathbf{d}$ for the noisy signal $\mathbf{y}$ displayed in (c); (b) Envelope $\mathbf{e}$ of the same signal; (c) Noisy RS signal with HS segment locations $\mathbf{A}$	45
3.4	(a) Noisy RS signal with nonlinear parts located by third-order cumulant (solid line); (b) Linear parts of the noisy RS signal with HS segment locations $\mathbf{A}_g^L$; (c) Nonlinear parts of the noisy RS signal with HS segment locations $\mathbf{A}_g^{NL}$; (d) Noisy RS signal with the output HS segment locations $\mathbf{A}$	46
3.5	(a) HS localization results for the pure HS signal using the proposed method (solid line) and the entropy based method in [40] (dotted line); (b) HS localization results for the synthesized noisy mild wheeze signal using the proposed method (solid line) and the entropy based method (dotted line); (c) The synchronized ECG signal.	48
3.6	Normalized histograms of a HS including segment (solid line) and a HS void segment (dashed line) of the signal in Fig. 3.4.	51
4.1	(a) Waveforms of pure HS and RS recordings; (b) Spectra of the HS and RS signals in (a).	55
4.2	An overall block diagram of the proposed semi-blind single-channel source extraction algorithm.	58
4.3	A block diagram of the proposed template design scheme.	59
4.4	Magnitude responses of (a) a prototype lowpass filter; and (b) the linear reconstruction bandpass filter bank used in Fig. 4.2 with $L = 32$ and filter length of 256.	62
4.5	The temporal continuity ζ with, order of the optimum filters $\mathbf{b}_{iopt}$, $N_b = 5$	65
4.6	The average power differences with changing γ in different frequency bands.	67
4.7	The waveforms and spectrograms of a real recorded wheeze signal (top two) and its reconstructed HS-free RS signal (bottom two).	69
4.8	The waveform (top) and spectrogram (bottom) of the reconstructed HS-free RS signal for the same noisy wheeze recording in Fig. 4.7 without template design scheme.	70
4.9	Average PSDs of the original HS-free normal LS signal, the simulated HS corrupted LS signal, and the reconstructed LS signal using the proposed method.	71

4.10	Average PSDs of the original HS-free RS signal, the HS corrupted RS signal, and the reconstructed RS signal using the proposed method and the linear prediction method in [7].	71
5.1	A general block scheme of the proposed respiratory phase segmentation method.	80
5.2	PSDs of a spike-contaminated RS segment (solid line), a spike-free RS segment of same length (dash line), and a spike-contaminated RS segment after spike removal using NEO (dotted line).	82
5.3	(a) Original waveform $y(n)$ of a real recorded stridor signal; (b) The bandpass filtered signal $y_f(n)$; (c) The resulting spike-free signal $y_{sf}(n)$ after spike removal.	83
5.4	(a) The resulting $y_{sf}(n)$ after bandpass filtering and spike removal of a real recorded wheeze signal; (b) The corresponding $\Gamma(y_{sf}^w)$ displayed together with threshold $\bar{\Gamma}$	84
5.5	$SampEn$ of $y_{sf}(n)$ in Fig. 5.4 (dotted line) and $SampEn$ of the corresponding input signal $x(n)$ after sparseness reduction (solid line). . . .	85
5.6	(a) Original waveform $y(n)$ of a real recorded signal as mixture of real wheeze and pleural sound (WNP) recording; (b) The preprocessed clean RS signal $x(n)$; (c) $ \hat{e}_N(n) $, $ \hat{e}_N^m(n) $, and $\bar{e}_N^m$ of $x(n)$ with $R_t = 6$; (d) Time-frequency plot of the signal in (a).	88
5.7	Averaged kurtosis of the spike-contaminated parts of $y_{sf}(n)$ by varying c values (solid line with indents) and the averaged kurtosis of their neighboring spike-free parts of $y_{sf}(n)$ (solid line).	93
5.8	Performance of the noise gain function $G_N(k)$ with varying β at $a = 2$ and $\kappa = 2$	95
5.9	(a) Waveform of a complete respiratory cycle (inspiration-pause-expiration-pause) from a preprocessed stridor signal; (b) Effects of m on $SampEn$ with $r = 0.2\sigma$ for the signal in (a); (c) Effects of r on $SampEn$ with $m = 1$	96
5.10	Mean segmentation error of the MPGA with different $maxgen$ for TBS recordings of different types.	98
6.1	Scatter plots of the estimated signal roots for an inspiratory normal RS segment $y_1(n)$ of small spectral support with (a) $N_w = 2048$ and (b) $N_w = 1024$; scatter plots for its consecutive expiratory wheeze segment $y_2(n)$ of large spectral support with (c) $N_w = 2048$ and (d) $N_w = 1024$	112

6.2	(a) Original waveform of a normal LS signal extracted from [37]; (b) The wavelet domain feature (dash-dot) displayed with the phase identification results by [44] (solid line) and the proposed method (dotted line) with 1, 2, 3 correspond to transition, expiration and inspiration.	117
6.3	(a) Original waveform of a real recorded wheeze signal; (b) The wavelet domain feature displayed with the phase identification results by [44] (solid line) and the proposed method (dotted line) with 1, 2, 3 correspond to pause, expiration and inspiration.	117
7.1	(a) Third-order cumulant of a typical real RS with its threshold; (b) Original waveform of the RS with its corresponding nonlinear regions segmented.	128
7.2	Illustrative TF plots of (a) $\hat{\mathbf{R}}_{ls}$ (b) $\hat{\mathbf{W}}$ for a nonlinear segment and (c) $\hat{\mathbf{R}}_{ls}$ (d) $\hat{\mathbf{W}}$ for a linear segment from heart sound contaminated real recording.	129
8.1	Gabor spectrogram of a CAS containing signal with CAS occurred in frequency bin $k = 8$, from 0.4 s to 0.5 s.	138
8.2	(a)-(c) Waveforms of $x(n, k)$ for $k = 5, 8, 20$ of the CAS signal as used in Fig. 8.1; (d) The resulting mean instantaneous kurtosis $K_{\text{mean}}(k)$ for varying frequency bin k of the same signal.	139
8.3	The averaged instantaneous kurtosis $\overline{K}_k^{bl}(n_i)$ of the broadband signal used in Fig. 8.1 without TF analysis, and that of the filtered narrowband signals at $k = 5, k_{\text{max}} = 8, 10, 20, 30$	140
8.4	(a) The block scheme for the proposed discriminating feature extraction method; (b) Generation of the discriminating function for the k th frequency bin.	141
8.5	(a) The discriminating function of the broadband CAS signal used in Fig. 8.1 without TF analysis and that for the filtered narrowband signals at $k = 5, k_{\text{max}} = 8, 10, 20, 30$; (b) The discriminating function of a broadband normal breath sound signal without TF analysis, and that for the filtered narrowband signals at $k = 5, k_{\text{max}} = 7, 10, 20, 30$	143
8.6	The block scheme for the proposed <i>SampEn</i> based feature extraction method.	145
8.7	(a) A partial representation with $k = [1 \ 10]$ of a concatenating histogram H_{SampEn} for the normal breath sound used in Fig. 8.5(b); and (b) The same partial H_{SampEn} for the CAS signal used in Fig. 8.5(a).	147
8.8	Performance of the 1NN classifier using different feature sets obtained by the proposed methods.	153

List of Tables

2.1	Summary of recommendations for short-time RS recording [18]	29
3.1	Comparison of the localization accuracy for different HS localization methods on synthesized data in terms of the mean and standard deviation error ϵ (%)	49
3.2	The performance of the proposed method on different types of real recorded TBS signals in terms of the mean and standard deviation error ϵ (%)	50
4.1	SNR^k and γ_{opt}^k for different frequency bands on experimental dataset #1	67
4.2	Comparison of the average power difference $(\mu \pm \sigma)$ (dB) between the PSDs of the original and the reconstructed normal RS by different HS cancellation methods	72
4.3	The average power difference $(\mu \pm \sigma)$ (dB) between PSDs of the original and the reconstructed RS of different types by the proposed method . .	73
5.1	RR estimation accuracies (%) of different methods on different types of RS	99
5.2	The estimation errors $(\mu \pm \sigma)$ (ms) of the proposed method on clean LS and real recorded TBS signals of different types	101
5.3	The estimation errors $(\mu \pm \sigma)$ (ms) for different respiratory phase segmentation methods on real recorded TBS signals of different types . . .	102
6.1	The resulting choices of $K2$ and N_w for different types of RS signals . .	111
6.2	Performance comparison between the proposed MI and the method in [9] on different types of RS signals	115
7.1	Nonlinearity detection and λ_S for different types of RS	130
7.2	Eigenvalue λ_S $(\mu \pm \sigma)$ for the noisy and preprocessed RS recordings of various types	131

8.1	Composition and size of the proposed feature sets	149
8.2	Feature evaluation results for different feature sets using SI with different distance metrics	150

List of Abbreviations

ABBREVIATIONS	FULL EXPRESSIONS
AR	Autoregressive
CAS	Continuous Adventitious Sound
CORSA	Computerized Respiratory Sound Analysis
DAS	Discontinuous Adventitious Sound
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
ECG	Electrocardiogram
EMG	Electromyogram
FEV	Forced Expiratory Volume
FFT	Fast Fourier Transform
GA	Genetic Algorithm
GLRT	Generalized Likelihood Ratio Test
HS	Heart Sound
IFFT	Inverse Fast Fourier Transform
KFD	Katz Fractal Dimension
LPC	Linear Predictive Coding
LS	Lung Sound
MA	Moving Average
MI	Mixing Index
MFCC	Mel-Frequency Cepstral Coefficient
MPGA	Multi-Population Genetic Algorithm
MSE	Mean Square Error
NGO	Nonlinear Energy Operator
NN	Nearest Neighbor
PCA	Principal Component Analysis

PCG	Phonocardiogram
PRS	Pathological Respiratory Sound
PSD	Power Spectral Density
SC	Signal Coherence
SNR	Signal-to-Noise Ratio
SVD	Singular-Value-Decomposition
SWT	Static Wavelet Transform
RS	Respiratory Sound
RR	Respiratory Rate
TBS	Tracheal Breath Sound
TF	Time-Frequency
TFR	Time-Frequency Representation
VFD	Variance Fractal Dimension
WT	Wavelet Transform
WVD	Wigner-Ville Distribution

List of Symbols

SYMBOL	MEANING
$\star$	Complex conjugate
$*$	Convolution operator
$ \cdot $	Magnitude
$L(\cdot)$	Log likelihood ratio
n	Index for discrete time
F_s	Sampling frequency
$(\cdot)^T$	Matrix transposition
$(\cdot)^{-1}$	Matrix inversion
$\mathbf{y}$	Input noisy RS sequence
$\ \cdot\ $	Euclidean norm
μ	Mean
σ	Standard deviation
$E\{\cdot\}$	Expectation operator
$(\cdot)^H$	Conjugate transposition of matrix
$\hat{\mathbf{a}}$	Estimate of $\mathbf{a}$
$\tilde{\mathbf{a}}$	Median of $\mathbf{a}$
$\bar{\mathbf{a}}$	Arithmetic mean of $\mathbf{a}$
$\mathbf{R}_{\mathbf{aa}}$	Covariance matrix of $\mathbf{a}$
$\mathbf{y}_j$	The corresponding $\mathbf{y}$ segment corrupted by the j th HS event
$\mathbf{s}_j$	The underlying RS component in $\mathbf{y}_j$
$\mathbf{v}_j$	The underlying HS component in $\mathbf{y}_j$
N	Length of $\mathbf{y}$
j	Index for the HS event
J	Total number of HS events in $\mathbf{y}$

$w(b)$	Window function in time domain
b	Time-lag
$\mathbf{y}^p$	Representation for $\mathbf{y}_g^L$ or $\mathbf{y}_g^{NL}$
$\mathbf{y}_n$	Wavelet transform coefficients of $\mathbf{y}^p$
$\mathbf{s}_n$	Wavelet transform coefficients of $\mathbf{s}$
$\mathbf{v}_n$	Wavelet transform coefficients of $\mathbf{v}$
i	Index for basis function
M	Total number of basis functions
L	Total number of subfilters in the filter bank
$\mathbf{h}$	Impulse response of the filter
k	Index for frequency bin or frequency band
$\mathbf{x}$	Preprocessed RS signal
$H(k)$	Transfer function in frequency-domain
K	Total number of frequency bins

Chapter 1

Introduction

Respiratory sounds (RSs) have long been the most important indicator of respiratory health and disease, and pulmonary auscultation has been the key method to detect and evaluate these respiratory dysfunctions for many years. Prior to the advances in computer technology and digital signal processing techniques, physicians relied on their hearing to detect abnormalities in RS of the patients. Although human auditory system can accurately identify the deterministic contents of sound in both time and frequency domains, it is not sensitive when dealing with noise. Furthermore, in many cases, the abnormal breath sounds known as adventitious sounds are intermittent or occurring only during sleep (e.g. intermittent wheeze) [1] which makes the diagnosis of such sound signals depends heavily on the history from the patient or care-giver. As stated in [2], the auscultation depends heavily on the individual's own hearing, experience and ability to distinguish between different sounds. Agreements on the presence or absence of physical signs in auscultation between the care-giver (e.g. parents) and the experienced physicians, or even among physicians are therefore far from perfect [3].

Automatic recording and analysis of RS using signal processing techniques, therefore play an increasingly important role in the evaluation of pulmonary diseases by

providing possible solution to remove the inherent subjectivity in auscultation via stethoscope. The research in digital signal processing is not only providing means to digitize, denoise, and store the information in the RS signals recordings, but also enables the quantification and the graphical representations which aid to visualize the defects in RS. It helps us to explore deeper into the acoustic mechanisms, and provides an alternative measurement of RS. By virtue of the advanced signal processing methods, more useful information could be extracted from RS for both diagnostic and assessment of treatment purposes.

Computerized auscultation has been traditionally analyzed and characterized by: 1) morphological changes in the time domain using statistical measures; 2) spectral properties in the frequency domain using simple spectral analysis; or 3) nonstationary properties in the joint time-frequency (TF) domain based on short-time Fourier transform [4]. However, the perspectives in RS analysis are broadened by the advanced signal processing methods which have started to emerge since the last decade. Through the years, RS analysis has branched out in various directions, and those of the most interests include heart sound (HS) cancellation [5][6][7], respiratory phase detection and flow estimation [8][9][10], nonlinear analysis of RS [11][12][13], feature extraction and classification [14][15][16][17].

1.1 Motivation

Although RS interpretation is still based on the auscultative findings of the physician in practice, modern computerized-based analysis have been proposed as new means to extend RS capabilities. This can be achieved by retrieving information based on many other underlying characteristics that foster the diagnostic information besides those based on sound perception.

In order to process RS for diagnosis of abnormalities (such as crackles, squawks, wheezes, stridors, rhonchi), the RS signal has to be first free (or contains minimum amount) of interferences originating from both inside and outside the body. The environmental noises such as speech or the noises generated by fans in electronic equipments can be easily minimized by proper shielding of the data acquisition devices during RS recording. Also, the nonrespiratory sounds and body sounds (as defined by [18]) related to breathing can be suppressed by proper choice of sensor locations. However, those body sounds which are not related to breathing such as HS and murmurs cannot be removed by adjusting the recording condition. Therefore, it should be necessary for RS signal preprocessing to remove HS which is the most significant and unavoidable interference.

Besides, respiratory phase information such as accurate timing and duration of each respiratory phase, is crucial for quantification of RS as well as heart flow and MRI studies. However, without the help of respiratory flow readings, respiratory phase localization and identification could be challenging especially for RS captured over the trachea or at the presence of adventitious sounds which alters the amplitude and frequency relationship between different respiratory phases. The above mentioned challenge therefore motivates the requirement of accurate respiratory phase detection.

Furthermore, as the respiratory acoustic measurements have shown to be very useful in the detail investigation of different upper airways pathologies [14], RS analysis to detect different kind of flow obstruction has drawn great attention recently. Therefore, being the major constituents of RS analysis, characterization and reliable feature extraction are necessary to the discrimination and classification of different types of RS for diagnostic purpose.

1.2 Objectives

Motivated by the importance of computerized RS analysis, the works dedicated in this thesis focus on three main categories of RS analysis from the perspective of advanced signal processing towards the following three objectives:

- **Heart Sound Cancellation:** to propose an automatic and accurate HS cancellation method which is effective for different types of RS. In order to overcome the restrictions of the existing HS reduction or cancellation schemes, variations in characteristics among various types of RS signals besides those differences between normal RS and HS have to be considered and investigated to find a generic solution.
- **Respiratory Phase Detection:** to propose a fully automatic respiratory phase segmentation and identification method which is robust for single-channel RS recordings of different types. Different types of RS should be accurately segmented into consecutive inspiratory phase, inspiratory pause, expiratory phase, and expiratory pause without being affected by the presence of adventitious sounds. The feature extracted for respiratory phase identification should be independent from amplitude and spectral changes to remove the limitations of the existing phase detection methods.
- **Feature Extraction and Classification of RS:** to develop new feature extraction methods that produce more discriminative features for reliable RS classification under different data acquisition condition at low flow rate. The characteristics of RS signal should be analyzed, and the proposed feature extraction methods should be able to produce large cluster separability to ensure high classification accuracies.

The objective of introducing such methodologies is not only to reveal the importance of RS as indicator of respiratory condition, but also to shed light on the inherent characteristics of the advanced signal processing techniques that manage to adapt to the specific properties of RS, providing a new standpoint in the evaluation of respiratory acoustics.

1.3 Thesis Contribution

The original contributions of this thesis include the followings:

- The proposal of a two-stage HS cancellation method which is robust for different types of RS. A wavelet based HS localization scheme has been firstly introduced in **Chapter 3**. The changing RS characteristics have been taken into consideration to ensure the robustness of the method. Also, a semi-blind single-channel source extraction algorithm is then introduced in **Chapter 4** to extract the underlying pure RS from the HS corrupted noisy input signals.
- The introduction of a two-stage respiratory phase detection method resistant to RS characteristics changes. The respiratory phase segmentation method proposed in **Chapter 5** enables the estimation of the number of respiratory phases in RS signal as well as the phase boundaries localization. A reliable phase identification index has then been introduced in **Chapter 6** which is effective with the amplitude changes imposed by the presence of adventitious sounds.
- New methods for the tracking and quantification of nonlinearity in different types of RS have been introduced in **Chapter 7**. The tracked nonlinearity has been then analyzed in the TF domain and the relationships between the adventitious sounds types and the degrees of nonlinearity have been investigated and analyzed.

- Novel RS feature extraction methods have been proposed in **Chapter 8** for RS classification. A new set of features based on the temporal characteristics of the filtered narrowband signal have been suggested to facilitate accurate RS classification into normal and continuous adventitious types. The extracted features are compared with existing features by a proposed evaluation method, and the improvement of classification performance has been shown.

1.4 Thesis Organization

Chapter 1 introduces the motivation and objectives of the thesis. Chapter 2 presents the research background which gives an overview of RS analysis from both medical as well as signal processing point of view. The data acquisition procedure adopted in this thesis for RS recording has also been described. The main works of the thesis are organized into three parts: *Heart Sound Cancellation From Respiratory Sounds* (Part I), *Acoustical Respiratory Phase Detection* (Part II), and *Time-Frequency Analysis and Classification of Respiratory Sounds* (Part III). The summary of the thesis, the future works, as well as the concluding remarks are presented in Chapter 9.

Part I consists of Chapter 3 and Chapter 4. Chapter 3 proposes a new HS localization method by introducing a wavelet based localization scheme based on multiscale decomposition and the cumulative sums of likelihood ratios. In Chapter 4, a novel semi-blind single-channel source extraction algorithm is further proposed for HS removal based on the HS localization results. Part II has two chapters: Chapter 5 and Chapter 6. A new respiratory phase segmentation method based on sample entropy and multi-population genetic algorithm is presented in Chapter 5. An effective respiratory phase identification index is then proposed in Chapter 6 to identify the respective inspiratory/expiratory phases for the segmented RS. Part III consists of Chapter 7

and Chapter 8 which demonstrates the nonlinearity of RS signals and proposes several feature extraction methods for RS classification, respectively.

Chapter 2

Research Background

2.1 Historical Overview of Respiratory Sound Analysis

Direct auscultation was already known to Hippocrates, who advised application of the ear to the patient's chest in order to hear sounds transmitted through the chest wall. However, from the time of ancient Greeks and their doctrine of medical experimentation until at least the 1950s, respiratory sounds (RSs) were considered as the sounds originating from within the thorax and they were justified mainly on the basis of their acoustic impression [4]. René Theophil Laënnec made the important contribution to the qualitative appreciation of RS by inventing the stethoscope in 1816 and publishing his treatise on auscultation in 1819. Descriptions of the acoustic events generated by ventilation of lungs and the systematical correlation between these events with anatomical and pathological findings after autopsy have been carefully elaborated.

At the beginning of the 19th century, clinical examination of pulmonary function was divided into four separate stages: questioning (i.e. asking the patient's history), inspection, palpation and percussion [19]. Direct percussion had only been introduced shortly before palpation, which was serving to evaluate transmission of vocalized vibrations from the trachea to the chest wall.

Laënnec later on found that pleural effusion and pulmonary consolidation prevented the passage of sound waves, whereas the transmission has increased when he placed his stethoscope over tuberculosis cavitations or areas of underlying bronchiectasis. Major cavitations resulted in identical vocalized sounds (pectoriloquy) heard through the chest wall and through the larynx where the high-pitched sounds travelled best. He also proposed that healthy subjects generated vesicular RS (which was described as “murmur of respiration”), while patients with pathological conditions produced adventitious RSs which were “sufficiently distinctive to permit identification of most organic disorders of the chest”.

We then had to wait until the 1950s to observe the development of modern devices and methods of recording and signal processing, which allowed objective studies of RS in time and frequency domains. In 1955, McKusick *et al.* employed condenser-type microphones and recorded the resulting electrical signals on a magnetic disc, where the signals were analyzed at different frequency with a variable filter [20]. Hannon and Lyman [21] then firstly introduced mechanico-electrical transducer to detect RS. They captured lung sounds (LSs) with a microphone linked through filters to a string oscilloscope and then to a graphic recording device.

Computerized recording and analysis of RS from the chest and trachea of normal and pathological subjects therefore become a major preoccupation of many clinical and engineering teams [22]. Most of the subsequent studies had followed a similar general pattern of steps: detection, preliminary processing, recording, and final processing of

recorded signals. The main objective of such studies was to understand the fundamental mechanisms for breathing sounds production [23][24][25][26][27]. In this way, compared to wholly subjective procedures of conventional auscultation using stethoscope, a more accurate and reliable computerized diagnosis of pathological signs has been enabled [28][29]. However, the large variations in equipments and techniques used prevent any meaningful comparison of results.

2.2 Anatomy and Physiology of Respiratory System

The respiratory system is situated in the thorax, and is responsible for gaseous exchange between the circulatory system and the outside world. Air is taken in via the upper airways (the nasal cavity, pharynx and larynx), through the lower airways (trachea, primary bronchi and bronchial tree), and into small bronchioles and alveoli within the lung tissue. The lungs are divided into lobes: the left lung is composed of the upper lobe, the lower lobe, and the lingula (a small remnant next to the apex of the heart); the right lung is composed of the upper, the middle and the lower lobes. The anatomy of the respiratory system is shown in Fig. 2.1.

The primary function of the respiratory system is to supply oxygen to the blood and to expel waste gases (of which carbon dioxide is the main constituent from the body). To take a breath in, the external intercostal muscles contract, moving the ribcage up and out. The diaphragm moves down at the same time, creating negative pressure within the thorax. The lungs are held to the thoracic wall by the pleural membranes, and expand outwards as well. This creates negative pressure within the lungs so that air is able to enter into the upper and lower airways. Expiration is mainly due to the

Figure 2.1: Anatomy of the respiratory system and the zoomed in picture of a bronchiolus branch and alveolar ducts (reprint from [30]).

natural elasticity of the lungs, which tend to collapse if they are not held against the thoracic wall.

Many researchers have worked on modeling the RS production mechanism to understand how a biological system works and also for its application in diagnosis. The RS originates in the larger airways within the lung. It arises distal to the trachea and proximal to the alveoli. Factors which increase flow velocity of air (tracheobronchial narrowing, increased ventilation) would increase turbulence and thus augment the intensity of RS [31].

The combination of the vocal tract and the subglottal airways including lungs form the respiratory tract, has highly unique acoustic properties. The acoustic characteristics of the vocal tract and the subglottal airways have therefore been modelled and investigated with the motivation to assess the relationship between the structure and the acoustic properties of the respiratory tract in health and pathological individuals [32][33][34]. Several models have been proposed for RS generation and transmission [34][35].

In [34], the acoustic properties of the respiratory tract are predicted and verified experimentally by modeling the respiratory tract and cylindrical sound source entering a homogenous mixture of air bubbles in water with thermal losses that represented lung parenchyma. Each airway segment is modelled by a T-equivalent electrical circuit similar to that of the vocal tract but with the addition of another shunt admittance to represent the acoustic properties of the airway walls. This model provides a functional correlation between the sound speed and the size of alveoli. It suggests the possibility of identifying collapsed areas of lungs (due to pulmonary diseases) by measuring the sound speed.

The model in [34] together with other numerous studies show either theoretically or experimentally that the increase in the lung volume results in attenuation in the sound acceleration. However, none of these sound transmission models explicitly predicts attenuation at particular lung volumes. Furthermore, considerable debates and discussions have been raised regarding the way of sound transmission from the major airways to the chest wall.

2.3 Respiratory Sound Characteristics and Categorization

Sounds generated in healthy lungs and airways by normal breathing are different according to their pick-up locations [19]. Although the origin of the sounds generated by ventilation is not completely clear, breath sounds are probably induced by turbulence of the air at the level of lobar or segmental bronchi [36]. The resulting noise, coming from the larger airways, therefore has a wide frequency spectrum. However, after filtering by the lungs and the chest wall which together act acoustically as a lowpass filter,

the normal RS signals captured over the lungs (named as lung sounds (LSs)) have their main frequency up to 200 – 250 Hz (as shown in Fig. 2.2). Normal LS has acoustically a soft character with expiration being nearly silent. The inspiratory phase is longer than the expiratory phase, which a ratio inspiration/expiration of about 2-to-1 during tidal breathing [37].

On the other hand, RS captured over the trachea (named as tracheal breath sound (TBS)) is not filtered and therefore has a frequency spectrum containing higher frequency components as high as 1200 Hz. The main energy of TBS extends up to 1000 Hz due to the influence by the high tracheal resonance frequencies [38] (as shown in Fig. 2.2). It can be seen that normal TBS is much louder than that of LS, but its difference in inspiration and expiration power varies among the subjects greatly.

Figure 2.2: A typical LS signal spectrogram (left) and a typical TBS spectrogram (right) (reprint from [39]).

The principal characteristics for RS categorization include: frequency, intensity, duration, and quality (timbre or texture). Based on these characteristics, RS can be categorized into two main categories of normal and abnormal. The normal RSs are described above, while the abnormal RSs can be further classified into continuous ad-

ventitious sounds and discontinuous adventitious sounds, each consist of the following types [36]:

Continuous adventitious sounds (CASs):

- Wheezes are musical continuous adventitious sounds which are superimposed on the normal breath sounds. The waveform of wheeze resembles that of a sinusoidal sound. It occurs mainly in expiration and is associated with airway obstruction. The dominant frequency of a wheeze is usually > 100 Hz and the duration is > 100 ms (as shown in Fig. 2.3). They can be monophonic when only one pitch is heard, or polyphonic when multiple frequencies are simultaneously perceived.

Figure 2.3: Waveform and spectrogram of a typical wheeze sound (reprint from [39]).

- Stridors are very loud wheezes, which are the consequence of a morphologic or dynamic partial obstruction in larynx or trachea. They are usually characterized by a prominent peak < 200 Hz in their frequency spectrum for adults (as shown in Fig. 2.4).
- Rhonchi are low-pitched wheezes containing rapidly damping periodic waveforms with a duration longer than 100 ms and frequency below 300 Hz (as shown in Fig. 2.5). They occur predominantly in expiration and are associated with chronic bronchitis and bronchiectasis than with asthma.

Discontinuous adventitious sounds (DASs):

Figure 2.4: Waveform and spectrogram of a typical stridor sound (reprint from [39]).

Figure 2.5: Waveform and spectrogram of the typical rhonchi sound (reprint from [39]).

- Squawks are short, inspiratory wheezes that usually appear in allergic alveolitis and interstitial fibrosis, predominantly initiated with a crackle, and caused by the explosive opening and fluttering of the unstable airway. The duration of squawks may vary between 50 and 400 ms (as shown in Fig. 2.6).

Figure 2.6: Waveform and spectrogram of typical squawk (reprint from [39]).

- Pleural rub sounds are DASs localized to the area overlying the involved pleura and occur in inspiration and expiration when roughened pleural surfaces rub together (as shown in Fig. 2.7).

Figure 2.7: Waveform and spectrogram of typical pleural sound (reprint from [39]).

- Crackles are discrete, explosive, non-musical DASs occurring usually during inspiration. They can be further classified based on the waveform, duration, and timing in respiratory cycle into fine (short duration, high-pitched, exclusively inspiratory) and coarse (long duration, low-pitched, usually inspiratory but occasionally expiratory). They are transient in nature, with duration < 20 ms and frequency components ranging from 100 to 2000 Hz (as shown in Fig. 2.8).

Figure 2.8: Waveform and spectrogram of typical fine crackle (top) and coarse crackle (bottom) (reprint from [39]).

The varieties in the RS categorization imply changes in the acoustic characteristics either of the source or the transmission path of RS due to the effects of pulmonary pathologies. Detailed description of each type of RS is elaborated in [14].

2.4 Existing Techniques for Respiratory Sound

Analysis

A few existing signal processing techniques which are commonly used in computerized RS analysis are described here. More specifically, the choice of the existing methods being adopted in this thesis for performance comparison on HS cancellation, respiratory phase detection, as well as RS feature extraction are presented briefly.

2.4.1 Heart Sound Cancellation Methods for Comparison

HS produces an intermittent noise during RS recording that influences the clinical auscultative interpretation of RS [4]. The presence of HS interferes the computerized analysis of RS by modifying the energy distribution in the spectra of RS as well as introducing undesired pseudo-periodicities. Therefore, proper cancellation of HS is necessary to facilitate successful RS analysis. Since a two-steps scheme for HS cancellation consists of HS localization and HS removal has been adopted frequently, amongst all the recent HS cancellation techniques that overcome the limitation by highpass filtering, the following methods have been selected in this thesis for comparison.

As reported in [39], the entropy based [40] and multiscale product based [7] HS localization methods as well as the HS removal methods by linear prediction [7] have been chosen here for comparison due to their high accuracy. The summary of the existing HS localization and removal methods has been elaborated in Chapter 3 and Chapter 4 respectively.

- **Shannon Entropy Based HS Localization Method:**

In [40], Shannon entropy is used to identify HS included segments from noisy RS recording. Since entropy involves computation of probability density function

(PDF), it reveals more information on the statistics of the signal when compared to average power. On finding the entropy of the segments in a RS recording, it has been observed that the HS corrupted segments have larger entropy than those HS void segments. This property has been utilized for HS localization.

Shannon entropy for a set of events $\{x_j|_{j=1,\dots,J}\}$ with probabilities $\{p_j|_{j=1,\dots,J}\}$, can be defined by

$$ShEn(x) = - \sum_{j=1}^J p(x_j) \log p(x_j) \quad (2.1)$$

where $p(x_j) \log p(x_j)$ converges to 0. This leads to the minimum $ShEn$ when the probability of an event is 0 or 1, while random signals with uniform PDFs such as pure noise give the maximum $ShEn$. The threshold is then computed as the mean plus standard deviation of the computed entropy for each segment to localize the HS events.

For J independent observations $\{X_j|_{j=1,\dots,J}\}$ of the random variable x , the PDF $f(x)$ has been estimated here using the nonparametric kernel density (KD) estimator which is defined as

$$\hat{f}(x) = \frac{1}{L_w J} \sum_{j=1}^J K\left(\frac{x - X_j}{L_w}\right) \quad (2.2)$$

where L_w is the window width or ‘smoothing parameter’ and $\hat{f}(x)$ is the estimation of the PDF $f(x)$. In practice, large number of observations are required for KD estimator and the kernel function K has been realized by a biased Normal kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (2.3)$$

with the bias and the variance of the estimator being proportional to L_w^2 and $\frac{1}{JL_w}$ respectively. The optimum value of $L_w = 1.06\sigma(x)J^{-0.2}$ is obtained from minimizing the mean square error between the actual and the estimated PDFs, where $\sigma(x)$ is the standard deviation of the input observation and J is the number

of observation samples. Fig. 2.9 illustrates PDFs of the HS contaminated and the HS void RS segments.

Figure 2.9: Histograms of the RS segments including HS (solid curve) and void of HS (dashed curve) (reprint from [40]).

- **Multiscale Product Based HS Localization Method:**

The method in [7] is based on the product of wavelet coefficients obtained by static wavelet transform (SWT) at different scales. The block representation of a 3 level SWT decomposition of $x(n)$ is illustrated in Fig. 2.10. Two sets of coefficients: approximation coefficients and detail coefficients, could be obtained by convolving the input signal $x(n)$ of length N samples with a lowpass filter $g_{\{l|l=1,\dots,3\}}$ and a highpass filter $h_{\{l|l=1,\dots,3\}}$ respectively. The approximate coefficients are then split into two parts with the same set of filters obtained by upsampling the filters in the previous stage. In [7], the compactly supported Symlet (order 5) wavelet with least asymmetry is used for SWT of the input signal, and the signal is decomposed into 3 scales. The multiscale product of wavelet coefficients is then calculated as

$$P_{j,q}(n) = \prod_{l=q}^{q+j} W_l\{x(n)\}. \quad (2.4)$$

$W_l\{\cdot\}$ is the SWT coefficient at the l th level, and $P_{j,q}$ is the multiscale product

of order j starting from the q th level with $q \geq 0$ and $j = 3$ being the desired number of levels to be multiplied.

Figure 2.10: Block representations of a 3 level SWT decomposition of $x(n)$ (above) and the upsampling process of the SWT filters (below).

Since the product of two adjacent decomposition bands present different behaviors for signal and noise, Lipschitz regularity is then used to distinguish between stationary and nonstationary parts of the signal by having 1 for a step function and -1 for a delta function. As magnitude of the singularities increases along the scales and vice versa, multiplication of the SWT coefficients between the decomposition levels leads to noise reduction. An example of multiscale product using a typical normal RS recording is shown in Fig. 2.11.

- **Linear Prediction Based HS Removal Method:**

After the multiscale product based HS localization as described in [7], the located HS corrupted segments are removed and the missing data are estimated using an autoregressive (AR) or a moving average (MA) model. Since it is not appropriate to assume RS to be stationary during the entire duration of the respiratory cycle

Figure 2.11: (a) Waveform of the original RS recording. (b)-(d) Approximation coefficients, first to third scales. (e) Multiscale products of first and second scales. (f) Multiscale products of first through third scales (reprint from [7]).

especially in the vicinity of breath onsets, the choice of AR or MA model is crucial. A MA model is therefore chosen at the vicinity of breath onset where most of the information is present in the next segment, whereas an AR model is chosen elsewhere. The optimal model order is selected as the order at which the energy difference between the original and the extrapolated data is minimum.

Linear predictive coding (LPC) method has been adopted to estimate the AR model coefficients. This involves the estimation of the autocorrelation to find a biased estimate of the input segment, followed by Levinson-Durbin algorithm to estimate the model coefficients. Levinson solves the symmetric Toeplitz system

of linear equations given by

$$\begin{bmatrix} r(1) & r(2)^* & \cdots & r(n)^* \\ r(2) & r(1)^* & \cdots & r(n-1)^* \\ \vdots & \ddots & \ddots & \vdots \\ r(n) & \cdots & r(2) & r(1) \end{bmatrix} \begin{bmatrix} a(2) \\ a(3) \\ \vdots \\ a(n+1) \end{bmatrix} = \begin{bmatrix} -r(2) \\ -r(3) \\ \vdots \\ -r(n+1) \end{bmatrix} \quad (2.5)$$

where $r = [r(1), r(2), \dots, r(n+1)]$ is the input autocorrelation vector, and $r(n)^*$ denotes the complex conjugate of $r(n)$. The parameters of the filter are given by $a(2), a(3), \dots, a(n+1)$. Levinson-Durbin algorithm therefore estimates the coefficients of a n th order AR process given by

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + a(2)z^{-1} + \dots + a(p+1)z^{-p}}. \quad (2.6)$$

2.4.2 Respiratory Phase Detection Methods for Comparison

It is of great importance to extract appropriate features from RS signals to indicate the precise changes at the respiratory phase boundaries. Therefore conventionally, respiratory phase detection depends dually on the feature extracted. Hard thresholding by local mean or detected peaks [41][42][43] has been then applied on the features to detect phases. However, the use of fixed thresholds becomes a major source of detection errors. An unsupervised and threshold independent detection has therefore been proposed in [44]. Triplet Markov chain has been adapted using the *priors* on the respiratory cycle structure to improve the detection accuracy. Two of the above mentioned methods have been adopted in this thesis for comparison:

- **Variance Fractal Dimension Based Phase Localization Method:**

In [43], it is postulated that RS signal has a chaotic feature during the short period of time between the respiratory phases (i.e. inspiratory pauses and expiratory pauses). The variance fractal dimension of RS has been then hypothesized to

have peaks at the breath onsets which would lead to automated detection of the breath onsets by acoustical means. The variance fractal dimension is defined as

$$D_\sigma = D_E - 1 + H_v \quad (2.7)$$

where $D_E = 1$ is the embedding dimension for 1-D signal and

$$H_v = \lim_{\delta t \rightarrow 0} \frac{\log(\text{Var}(\delta s)_{\delta t})}{2\log(\delta t)} \quad (2.8)$$

with $x(n)$ being the input signal and $(\delta x)_{\delta t} = x(t_2) - x(t_1)$ being the variation of RS in $\delta t = |t_2 - t_1|$. An illustrative plot showing the measurement scales $\delta t = 32, 64, 128$ for D_σ calculation is listed in Fig. 2.12.

Figure 2.12: The measurement scales $\delta t = 32, 64, 128$ for D_σ calculation (reprint from [43]).

- **Unsupervised Statistical Approach for Phase Detection:**

Triplet Markov chain has been performed in wavelet packet domain to improve the phase detection accuracy in [44]. For a given level j , the wavelet packet transform (WPT) decomposes RS signal $x(n)$ with $1 \leq n \leq N$ into 2^j subbands corresponding to wavelet packet coefficient set $w_k^j(m)$. It defines the m th coefficient of the k th subband where $1 \leq m \leq N/2^j$ and $1 \leq k \leq 2^j$. The highest decomposition level $j=6$ and Daubechies wavelets [45] with length of 20 have been chosen. In order to apply statistical approach instead of thresholding

for phase detection, the statistical characteristics of the wavelet coefficients have been adopted. It is assumed that being a broad spectrum noise, the coefficients within each respiratory phases follow an independent normal distribution. The extracted feature Y is therefore calculated as the envelope of $w^{\{j|j=6\}}$ by summing the squared coefficients in each frame.

Within each respiratory phase, the observed time-scaled features with enhanced independency and Gaussianity is approximated as a χ^2 distribution and thus parameterized by a set of parameters θ . By considering the correct respiratory phase sequence as a hidden Markovian process $X=(X_n)_{n \in N}$ taking values from $\Delta=\{\delta_1, \delta_2, \delta_3\}$ (i.e. pause, expiration, inspiration), a constrained version of triplet Markov chain (TMC) model is adopted by introducing a stochastic process $U=(U_n)_{n \in N}$ taking values from the positive integer set $\Lambda=\{\lambda_1, \lambda_2, \dots\}$. The auxiliary process U denotes the sojourn time of the chain X in a particular state, and the couple $V = (X, U)$ is a stationary Markov chain. Therefore, $T = (V, Y)$ could be used as a classical hidden Markov chain with the marginal $p(x_i, y) = \sum_{u_i \in \Lambda} p(v_i, y)$. The size of Λ has been fixed as the estimated duration for respiratory cycle to ensure the homogeneity of X . Modified unsupervised stochastic EM [46] framework has been adopted for the estimation of $p(u, x)$ and Bayesian maximum posterior mode, $\hat{x}_n = \arg \max_{x_n \in \Delta} p(x_n | y)$ has been lastly applied to obtain X .

2.4.3 RS Feature Extraction Methods for Comparison

The well known feature extraction methods for RS classification are based on expanded time waveform (for crackles), waveform fractal dimension analysis (e.g. variance fractal dimension (VFD) and Katz fractal dimension (KFD) [47]), spectral analysis (e.g.

Mel-Frequency Cepstral Coefficients (MFCCs) [48] and signal coherence [49]), and parametric spectral analysis (AR coefficients [50]). All the above mentioned methods have been compared in Chapter 8. Since the theory for VFD and AR coefficients based approaches have been covered in Sections 2.4.1 and 2.4.2 respectively, the rest of the methods are described below:

- **KFD Based Approach:**

Considering the morphological properties of RS signals, it is apparent that in order to scale them, different scaling factor is required for each axis, which indicates that the waveforms are not self-similar but might be self-affine. The algorithm adopted in [47] for obtaining KFD values is defined as

$$KFD = \frac{\log_{10}(\Delta n)}{\log_{10}(\Delta n) + \log_{10}(d/N_l)}. \quad (2.9)$$

Δn is the number of increments between samples of the RS signal over which KFD is calculated (length of the windowed segments), N_l is the sum of all distances between successive increments, and d is the value of the maximum distance measured from the beginning of the first increment.

- **MFCC Based Approach:**

Since the MFCCs are able to provide important features used in various kinds of speech applications, they have been adopted in [48] for wheeze detection. The calculations of the MFCC are as follows: The short-term power spectrum $P(\omega)$ of the windowed segment have been firstly obtained by discrete Fourier transform (DFT) and then warped along its frequency axis into the mel-frequency axis. The warped power spectrum $P(\omega_M)$ with ω_M being the mel-frequency index is then convolved with the triangular bandpass filters into $\theta(\omega_M)$ to produce L outputs of $X(i) = \ln(\theta(\omega_M^i))$ with $i = 1, \dots, L$. The MFCC is finally computed for $b = 1, \dots, L$

as

$$MFCC(b) = \sum_{i=1}^L X(i) \cos\left[\frac{\pi b}{L}(i - 0.5)\right]. \quad (2.10)$$

- **Signal Coherence Based Approach:**

Signal coherence is a measure that indicates the amount of random variation in each Fourier component of the signal. Since variations in the characteristics of RSs give distinctive information about their conditions, signal coherence has been adopted in [49] as a feature to discriminate between normal and abnormal RS. The signal $X(k)$ is decomposed into a deterministic part $\lambda(k)$ with perfect periodicity and a zero mean nonstationary process $U_m(k)$ as the random part with k being the frequency index. The variability of the k th frequency component of the randomly modulated signal is therefore defined in terms of $\lambda(k)$ and the variance of the DFT of the random component $\sigma_U^2(k)$ as

$$\gamma_X(k) = \sqrt{\frac{|\lambda(k)|^2}{|\lambda(k)|^2 + \sigma_U^2(k)}}. \quad (2.11)$$

2.5 Respiratory Sounds Recording

Since the invention of stethoscope, auscultation has been the primary assessment technique for physicians. Listening to RS by means of a stethoscope involves several physical phenomena, such as vibrations of skin that are converted into pressure variations of air in the stethoscope, which are then transmitted to the eardrum. However, since the use of stethoscope is not objectively tested, calibrated, or compared, despite the high cost of many modern stethoscopes (including digital stethoscopes), their use is restricted to auscultation only [39]. In addition, they are not able to provide a flat frequency spectrum of the sounds due to the selective amplification or attenuation within the spectrum of clinical interest [51].

On the other hand, digital signal recording which provides a faithful representation of sounds, enables the quantitative analysis of this bioacoustic signal. Environmental and subject conditions, as well as breathing manoeuvres, may also have marked influences on different variables of RS [18]. Therefore, basic standards are needed for these items to enable the comparison of results between different tests.

2.5.1 Subject Conditions and Procedures

The RS signal has a bandwidth of [100 2000] Hz while being captured over the chest and a bandwidth of [100 6000] Hz while being captured over the trachea [52]. The sampling frequency F_s above 10 kHz is therefore sufficient to capture RS signals of different types. For the acquisition of usable data, an analog system consists of a sensor, an amplifier, and signal conditioning filters (as a combination of lowpass filters and highpass filters) is always necessary.

RSs are usually captured either by electret microphones or sensitive contact accelerometers. Although both types of devices are displacement receivers, the waveforms that they deliver are different because of the coupling differences. Selection criteria of a device should also include size, average lifetime, and maintenance cost. Both sensors have drawbacks that the microphones require mounting elements that change the characteristics of sound transduction [52], whereas accelerometers are very sensitive to movement artifacts [22].

The guidelines concerning preparation of the subjects before short-term recording of RS are the same of those recommended by the European Respiratory Society to be used before lung function tests [53]. Table 2.1 tabulates the recommendations by [18] regarding the short-time recording of RS.

2.6 Recording Setup

Fig. 2.13 illustrates the schematic of the RS recording adopted for our data acquisition in this thesis.

Figure 2.13: Apparatus setup for RS recording.

Real recordings have been carried out in audio laboratory with the subjects in sitting position. Single electret condenser microphone (ECM-77B, Sony Inc., Japan) has been inserted into a hemispherical rubber chamber of 2 cm in diameter, and placed over suprasternal notch. The recording environment and equipments have been chosen based on the standard given by [18]. The choice of the microphone together with the recording condition, are able to suppress the environmental noises to the largest extent. Recording software WAVEPAD (V3.05, NCH Swift Sound Software) has been used and the RS recordings have been saved as mono-channel '*.wav' files. Test subjects have been asked to hold the breath for 10 seconds then breathe normally with no targeted flow, and 600 seconds recording has been saved for each subject.

Table 2.1: Summary of recommendations for short-time RS recording [18]

Sensor Specification	
Frequency Response	Flat in the frequency range of RS, maximum deviation < 6 dB
Dynamic Range	> 60 dB
Sensitivity	Independent of frequency, static pressure, and sound direction
Signal-To-Noise Ratio	> 60 dB ($S = 1$ mV/Pa)
Directional Characteristics	Omnidirectional
Coupling	Conical shape with depth of $2.5 - 5$ mm and diameter of $10 - 25$ mm at skin; vented
Fixing Method	Elastic belt or adhesive ring
Noise and Interference	Shielded microphones; protection from mechanical vibrations
Amplifier	
Frequency Response	Constant gain and linear phase in the band of interest
Dynamic Range	> 60 dB
Filtering	
Highpass Filtering	Cutoff: 60 Hz; linear phase with minimized ripple; roll-off > 18 dB/octave
Lowpass Filtering	Cutoff: above higher frequency of RS; minimized ripple; roll-off > 24 dB/octave
Subject Conditions and Procedures	
Body Posture	Subject in sitting position
Microphone locations	Trachea: on the trachea at the suprasternal notch; Chest: right and left posterior and basal area of the chest usually 5 cm laterally from the paravertabral line and 7 cm below the scapular angle
Environmental Conditions	
Environmental Noise	Background noise level < 45 dB; minimum ambient noise from environment
Other Room Conditions	Comfortable room temperature, lighting, ventilation

Part I

Heart Sound Cancellation From Respiratory Sounds

Auscultation of respiratory sounds (RSs), as the process of listening to RS produced in the body, is the most important clinical technique in the evaluation and diagnosis of respiratory disorders. However, in sound acquisition for computerized respiratory analysis, RS recordings are strongly interfered by the presence of regularly occurred heart sound (HS), which is an unavoidable source of background noise. The complete removal of HS corrupted segments from RS recordings makes the processed RS signal unsuitable for its further analysis due to the artifacts generated. The need to eliminate noise interferences prior to signal analysis has thus been established [22].

The HS falls in the low frequency range from a few Hertz to over 200 Hz when picked up over the chest which has lowpass filtering effect [36]. This signal occupies even a higher frequency when it is picked up at the suprasternal notch. Since it cannot be suppressed by simply adjusting recording techniques, this leads to three basic approaches for HS reduction including low-order highpass filtering [52], notch filtering [54], and adaptive filtering [55]. As HS interferes with RS features in both temporal domain and spectral domain [56], the conventional filtering [52][54] is not effective enough to minimize the interference without the loss of useful signal information [57]. The third approach, proposed by [58], adaptively filters the input RS signal with a reference signal which is highly correlated to HS and is derived from modified electrocardiogram (ECG) signal. An improved technique has been developed by [59] to remove the extra reference input but the HS could only be reduced by 24 – 49%.

In order to achieve the complete HS cancellation while preserving the RS signal components, HS cancellation algorithms have then been developed to process only the HS contaminated RS segments. Therefore, HS cancellation has been achieved in Part I of the thesis in two steps: a wavelet based approach as proposed by Chapter 3 to locate the HS contaminated RS segments; followed by a source extraction method as described in Chapter 4 for HS removal based on the HS localization results.

Chapter 3

Wavelet Based Approach for Heart Sound Localization

3.1 Introduction

Localization of HS segment in normal RS is often achieved through thresholding of the calculated feature values based on certain selected properties of RS and HS. Adaptive thresholding has been the most practical approach as it does not require any *a priori* information of the signal. The accuracy of HS segment localization depends on both the thresholds as well as the effectiveness of the extracted features in discriminating HS from RS. In literature, the thresholds have been calculated either globally or segment-wise based on discriminating features as: average power [60][59][5], lower order statistics of the original signal [7], and lower order statistics of feature domain values [40]. On the other hand, the features to differentiate HS and RS include: time-frequency domain power differences [60], signal singularities [7], physiological dynamics [61], signal distribution [40] and signal complexity [62]. The HS localization methods based on the above mentioned discriminating features have been summarized in [63] and some

comparison results are listed in [39].

All the above mentioned techniques have been developed for RS captured over the chest (i.e. lung sound (LS)). However, besides LS, tracheal breath sound (TBS) being the RS signal detected over the extra thoracic part of trachea, is more appropriate for the analysis of pathological sounds originating in larynx or trachea (such as stridor). Therefore, real TBS recordings are also considered in this chapter for achieving more robust HS localization method. Although TBS is characterized by a broader spectrum of noise containing higher frequency components compared to LS [36], compared to HS obtained over the chest, the spectra of HS signals captured over the suprasternal notch are also broadened. Since we incorporate HS localization of different types of continuous adventitious sounds beyond normal breath sound in this chapter, the large variations in the signal nature and amplitude downgrade the performance of the existing methods which are developed based on the power, dynamics, complexity and distribution of the signal.

Hence, the purpose of this chapter is to develop an automatic and accurate HS localization method which is effective for different types of RS. Third-order cumulant of the original signal is applied to investigate the nonlinearity and identify the nonlinear parts of the signal. Then, an envelope of the localized HS segments is calculated separately for the identified linear/nonlinear parts by incorporating multiscale decomposition and instantaneous phase of the signal in order to remove the effect of adventitious sounds. Since RS and HS have different behaviors in wavelet domain, wavelet based multiscale decomposition is applied first to exploit this difference in signal nature. The highly dynamic and nonstationary nature of the HS signal and the relatively slow-varying nature of RS of various types are then captured by the decision function based on cumulative sum of the likelihood ratios. By processing the linear/nonlinear parts of the signal separately, the same thresholding strategy can be applied on the detection function to

obtain accurate HS locations. The recently proposed Shannon entropy based method in [40] has been chosen for comparison due to its reliable performance under different amplitude ratios of HS and LS. The two methods have been compared in terms of detection rate and boundary localization accuracy for the synthesized as well as the real recorded RS signals.

3.2 Methodology

An overall block diagram of the proposed HS localization method is shown in Fig. 3.1(a). The input noisy RS sequence is indicated by $\mathbf{y}$ and the output sequence of HS segment locations is denoted by $\mathbf{A}$. The nonlinearity of the input signal has firstly been examined by the calculated third-order cumulant in Section 3.2.3. The identified g th linear and nonlinear part of $\mathbf{y}$ are then denoted by $\mathbf{y}_g^L$ and $\mathbf{y}_g^{NL}$ respectively. In this context, $g = 1, \dots, \mathbb{G}$ where $\mathbb{G} = \{G_1, G_2\}$ with G_1 and G_2 being the total number of the linear and nonlinear parts detected for each $\mathbf{y}$. A proposed wavelet based localization scheme as described in Section 3.2.4 has then been applied to every identified parts to produce the HS segment location sequences $\mathbf{A}_g^L$ and $\mathbf{A}_g^{NL}$ that correspond to $\mathbf{y}_g^L$ and $\mathbf{y}_g^{NL}$ respectively. The output HS segment location sequence is obtained by concatenation of the $\mathbf{A}_g^L$ and $\mathbf{A}_g^{NL}$.

3.2.1 Data

The synthesized data has been generated by superimposing pure HS recording onto various types of standard RS of same length as provided by [37] and [64]. The total of 5 different types of RS signals including normal LS for adults and infants, expiratory mild wheeze, inspiratory stridor and expiratory moderate wheeze, monophonic wheeze, and polyphonic wheeze, each of 10 seconds duration, have been used. Pure

Figure 3.1: (a) An overall block scheme of the proposed HS localization method; (b) The detailed steps involved in the wavelet based localization scheme for linear/nonlinear parts of the noisy RS signal.

HS recordings (which is phonocardiogram (PCG) signal) together with the synchronized electrocardiogram (ECG) signal for healthy subjects have been obtained from the Department of Cardiology, Lund University Hospital, Lund, Sweden, with permission.

Real recordings were done following the procedures as described by Section 2.6 with sampling frequency $F_s = 44.1$ kHz. In this study, the real recording dataset consists of TBSs from 7 healthy and 10 subjects with different degrees of airway obstruction (8 males/9 females, 15 ± 9 years old). The characteristics due to sex, age, weight were not taken into consideration.

3.2.2 Signal Model

In particular, RS heard over the large airways is primarily related to the vibrations of the upper airway walls and turbulent airflow, while HS occurs mainly due to the valvular activity of the heart. The hypothetical sound sources of HS and RS can be approximately considered as point sources [65] which are assumed to be mutually uncorrelated. However, the assumption no longer hold at the recording position since RS and HS are transmitted through partially similar pathes before reaching the microphone placed at the suprasternal notch.

The noisy RS signal $\mathbf{y}$ is considered to be a continuous RS signal of different types contaminated by discontinuous HS signal. The corresponding noisy segment as corrupted by the j th HS event is denoted by $\mathbf{y}_j$. It can be expressed by

$$\mathbf{y}_j = \Upsilon(\mathbf{s}_j + \mathbf{v}_j), \quad 1 \leq j \leq J \quad (3.1)$$

where $\mathbf{s}_j$ and $\mathbf{v}_j$ are the $(N \times 1)$ vectors representing the underlying RS component of different types and HS component respectively. The total number of HS events is denoted by J . HS signal $\mathbf{v}_j$, which has transient waveform, is superimposed onto uncorrelated RS signal $\mathbf{s}_j$ while $\Upsilon(\cdot)$ represents the effect of transmission path through trachea and skin. We assume an additive model at the sources of RS and HS here. According to the nature of sound production mentioned in the previous paragraph, possibilities of multiplicative or even more complex interaction are not considered here.

3.2.3 Calculation of Cumulant

We have identified the possible nonlinearity of the input signal based on third-order cumulant. In this chapter, the nonlinear region refers to the part of the signal with asymmetric distribution and broader spectrum compared to the linear region. There-

fore, the third-order cumulant is chosen instead of the fourth-order cumulant to identify this asymmetry in distribution using high values.

A short-time estimate of the third-order cumulant of an arbitrary zero-mean signal $z(n)$ calculated from the observed signal $\mathbf{y}$ is given by [66]

$$\hat{C}_{3,z}(a_1, a_2; b) = \frac{1}{N_{12}} \sum_{n=N_1}^{N_2} z_b(n) z_b(n + a_1) z_b(n + a_2), \quad (3.2)$$

$$z(n) = y(n) - \frac{1}{N} \sum_{n=1}^N y(n),$$

$$N_1 = \max\{b - B, b - B - a_1, b - B - a_2\},$$

$$N_2 = \min\{b + B, b + B - a_1, b + B - a_2\},$$

$$N_{12} = N_2 - N_1 + 1,$$

$$z_b(n) = \begin{cases} z(n)w(n - b) & b - B \leq n \leq b + B \\ 0 & \text{otherwise.} \end{cases}$$

where N is the length of the observed signal $\mathbf{y}$. B is the duration in time with b being the time-lag, whereas the length of the rectangular window $w(b)$ in the above is $2B + 1$.

The estimates $\hat{C}_3(a_1, a_2; b)$ are reduced to one-dimensional functions by taking $a_1 = -a$ and $a_2 = a$ for nonlinearity identification. Since frequency interaction and nonlinear phase coupling are not considered here, the diagonal cumulant slices $\hat{C}_3(-a, a; b)$ is sufficient to extract the useful auto-terms information while keep modest computational complexity. However, the cumulants decrease rapidly with increasing a . Thus, the proposed nonlinear identification function

$$\rho_3(b) = \hat{C}_3(0, 0; b) - \hat{C}_3(-1, 1; b) \quad (3.3)$$

is used for detecting the nonlinear parts of the input signal. By taking the identification function as the difference between $\hat{C}_3(0, 0; b)$ and $\hat{C}_3(-1, 1; b)$, the third harmonics in the signal are subtracted, and the aliasing effect for the broadband signals is therefore reduced.

The nonlinear regions of the input signal are extracted leaving the rest being linear parts.

$$\mathbf{y} = \parallel \sum_{g=1 \dots \mathbb{G}} \{\mathbf{y}_g^L, \mathbf{y}_g^{NL}\} \quad (3.4)$$

where $\mathbf{y}_g^L$ and $\mathbf{y}_g^{NL}$ refer to the g th linear and nonlinear parts of $\mathbf{y}$, with $\mathbb{G} = \{G_1, G_2\}$ indicates the total number of the detected linear and nonlinear parts for each noisy RS signal. If a noisy RS signal has $\max(\rho_3) \leq 10^{-8}$, then $G = 0$ and the signal is considered to be linear throughout. The symbol $\parallel \sum$ indicates the process of concatenating the identified linear/nonlinear parts alternatively by following their indices (e.g. $\mathbf{y} = \parallel \sum_{g=1, \dots, \mathbb{G}} \{\mathbf{y}_g^L, \mathbf{y}_g^{NL}\} = [\mathbf{y}_1^L; \mathbf{y}_1^{NL}; \dots; \mathbf{y}_g^L; \mathbf{y}_g^{NL}; \dots; \mathbf{y}_G^{NL}; \mathbf{y}_G^L]$ if the signal starts and ends with linear parts). Fig. 3.2 shows the illustrative plots of the third-order cumulant and the identified linear/nonlinear parts of a real RS signal. Similarly, the two HS location sequences $\mathbf{A}_g^L$ and $\mathbf{A}_g^{NL}$ as obtained in the following Section 3.2.4 are also concatenated as shown in (3.4) with $\mathbf{y}$ being replaced by $\mathbf{A}$.

By identifying and segmenting the linear/nonlinear parts of the observed signal, different thresholds can be employed for HS localization in the next step (i.e. the wavelet based localization scheme in Section 3.2.4). Since a global threshold is calculated based on the input signal of the HS localization scheme in the following section, the presence of nonlinearity tends to raise the threshold level when the entire signal $\mathbf{y}$ is used as the input of the scheme. In this way, HS events occurred during nonlinear parts would have high probability to be missed out by the localization scheme with $\mathbf{y}$ being the input. Therefore, by feeding in the identified linear/nonlinear parts separately, the characteristic difference between HS and the underlying RS in each part is enhanced which improves the HS localization accuracy.

Since our method is applied to detect broadband nonlinear signals, the choice of window length for the cumulant estimators is crucial. The window length selected

Figure 3.2: (a) The proposed nonlinear identification function for a real RS signal; (b) Original waveform of the RS signal with its respective linear (dashed line)/ nonlinear (solid line) parts being labeled manually.

should not be large as the amplitude of the output signals would become zero at some points for a broadband signal. A proper window $w(b)$ should be able to track the fast variations due to nonlinearity but ignore the relatively slower variations caused by HS. The window length $2B + 1$ is therefore chosen to be much shorter than the minimum variation durations of HS. Since the length of $w(b)$ is chosen to be smaller than $1/20$ of the minimum duration for HS (which is 20 ms [67]), $10 \leq B \leq 21$ is therefore the acceptable range at $F_s = 44.1$ kHz, with $B \leq 10$ produces very noisy results.

3.2.4 The Wavelet Based Localization Scheme

The proposed HS signal localization scheme is based on the time-scale characteristic of wavelet transform. The block diagram summarizing the steps involved in the proposed scheme is illustrated by Fig. 3.1(b). The choice of the wavelets is application dependent and it depends upon the requirement of the time-frequency/ scale resolutions. Among different choices in the wavelet family, Morlet wavelet [68] is employed here because of its ability to provide different window lengths for signals composed of different frequencies/scales. Also, the duration-bandwidth product is minimal that thereby ensures a good time and frequency resolution. Furthermore, the relative bandwidth, which is the ratio of the bandwidth to the center-frequency, is constant for all frequencies. This is ideal for nonstationary signals that require good time resolution. It contains low frequencies components with longer duration and high frequencies components with shorter duration. The Morlet wavelets form non-orthogonal basis functions with considerable spectral overlap among the basis functions and it can be expressed by the wavelet functions $\{\psi(n, m) = \pi^{-1/4} e^{j2\pi f_0 n/m} e^{-n^2/2m^2}\}$, with $n = 0, \dots, N_p - 1$ being the time indices and $m = 0, \dots, M - 1$ being the scale indices. N_p refers to the total number of samples of the noisy signal $\mathbf{y}_g^{NL}$ or $\mathbf{y}_g^L$.

The affine time-scale transformation of the identified $\mathbf{y}_g^{NL}$ or $\mathbf{y}_g^{NL}$ then take the following matrix form:

$$\mathbf{y}_n = \mathbf{W}_n^H \mathbf{y}^p \quad (3.5)$$

where H is conjugate transpose. $\mathbf{y}_n$ is with size $(M \times 1)$, $\mathbf{y}^p$ refers to $\mathbf{y}_g^L$ or $\mathbf{y}_g^{NL}$ with size $(N_p \times 1)$, and $\mathbf{W}_n$ is a $(N_p \times M)$ matrix given by

$$\mathbf{y}^p = [y(n) \ y(n-1) \cdots y(n-N_p+1)]^T$$

$$\mathbf{W}_n = \begin{bmatrix} \psi(n, 0) & \cdots & \psi(n, M-1) \\ \psi(n-1, 0) & \cdots & \psi(n-1, M-1) \\ \vdots & \cdots & \vdots \\ \psi(n-N_p+1, 0) & \cdots & \psi(n-N_p+1, M-1) \end{bmatrix}$$

with $(\bullet)^T$ denoting transposition and $y(n, m)$ representing the wavelet transform coefficient of $\mathbf{y}^p$ at the m th scale.

In compared to discrete wavelet transform, the continuous wavelet transform being shift invariant, is able to ensure high HS localization accuracy. The smallest scale of 2 is chosen here which corresponds to the frequency range of $[0 \text{ } 1378]$ Hz. This covers the whole frequency range of pure HS signal. Such a choice necessarily depends on the central frequency, the orthogonality, and the mother wavelet, to provide the reliable HS localization. Since the scale selection depends very much on the choice of wavelet, we select the scales based on the normalized global wavelet power of the available real signals. Since RS components are always confined in the finer scales, the localization accuracy for HS contaminated signals with high amplitude RS components can be improved by reducing the number of selected scales. The total number of dyadic scales (which is used as 24 here) depends on the length of the HS events.

Decision Rule

The localization of HS signal can be realized as a binary hypothesis testing between the two hypothesis, H_0 and H_1 :

$$\begin{aligned} H_0 : \mathbf{y}_n &= \mathbf{s}_n \\ H_1 : \mathbf{y}_n &= \mathbf{s}_n + \mathbf{v}_n \end{aligned} \tag{3.6}$$

where $\mathbf{s}_n$, $\mathbf{v}_n$ and $\mathbf{y}_n$ are the wavelet transform coefficients of $s(n)$, $v(n)$ and $y^p(n)$, respectively. At a given time instance n , we have to decide if $\mathbf{y}_n$ contains the HS component $\mathbf{v}_n$, or it contains only the RS component $\mathbf{s}_n$.

We employ the maximum likelihood approach, which provides Generalized Likeli-

hood Ratio Test (GLRT) for the noisy signal $\mathbf{y}_n$ and the underlying RS component $\mathbf{s}_n$ with unknown covariances given by [69]

$$L(\mathbf{y}_n) = \ln p(\mathbf{y}_n, \hat{\mathbf{R}}_{yy}; H_1) - \ln p(\mathbf{y}_n, \hat{\mathbf{R}}_{ss}; H_0) \begin{cases} > \gamma \text{ for } H_1 \\ \leq \gamma \text{ for } H_0 \end{cases} \quad (3.7)$$

where $L(\cdot)$ is the log likelihood ratio with $\hat{\mathbf{R}}_{yy}$, $\hat{\mathbf{R}}_{ss}$ as the maximum likelihood estimates of the covariance under hypothesis H_1 and H_0 , respectively. $\mathbf{s}_n$ and $\mathbf{v}_n$ are uncorrelated in the hypothesis (3.6) which may not be consistent with the signal model proposed in (3.1). The signal transmission model $\Upsilon(\cdot)$ might introduce coupling between the two uncorrelated sources $\mathbf{s}_n$ and $\mathbf{v}_n$. However, these source signals with additive model can be decorrelated through the implementation of GLRT.

This GLRT is implemented by a simple and efficient scheme based on principal component analysis (PCA). The method is based on orthogonal transform in the wavelet domain. The extraction of the principal components is performed by using singular-value-decomposition (SVD) which provides the orthogonality of the estimated eigenvectors. The principal eigenvectors are computed using SVD of the covariance matrix. For this, the covariance matrix of $\mathbf{y}_n$, $\mathbf{R}_{yy}$, under hypothesis H_1 is given by

$$\mathbf{R}_{yy} = E[\mathbf{y}_n \mathbf{y}_n^H] \quad (3.8)$$

with $E[\cdot]$ being the expectation operator. Using SVD of the measurement matrix $\mathbf{R}_{yy}$, we get

$$\mathbf{R}_{yy} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T, \quad \mathbf{\Lambda} = \text{diag}[\sigma_1^2 \cdots \sigma_I^2] \quad (3.9)$$

where $\mathbf{U}$ is the eigenvector matrix of $\mathbf{R}_{yy}$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_I$ are the corresponding singular values with I denotes the total number of singular values.

By applying the subspace concept in [70] together with SVD of the correlation

matrix $\mathbf{R}_{yy}$, the test statistics can be approximated as

$$\hat{T}(\mathbf{y}_n) = \sum_{i=1}^r |\hat{U}_i \mathbf{y}_n|^2 \quad (3.10)$$

where $\hat{\mathbf{U}} = [\hat{U}_1 \cdots \hat{U}_r]$ contains r orthogonal eigenvectors having the r largest eigenvalues of $\mathbf{R}_{yy}$, whereas the remaining $(I-r)$ eigenvectors correspond to the eigenvalues (power spectral densities) of the RS component (see Appendix A for more detail).

Our localization approach is based on cumulative sums $\mathbf{CS}$ of likelihood ratios which capture the dynamic behaviour using N_d samples before the time n as

$$CS(n) = \sum_{n=N_d-1}^n L(\mathbf{y}_n). \quad (3.11)$$

$L(\mathbf{y}_n)$ is calculated using (3.7), where the log likelihood of HS, $\ln p(\mathbf{y}_n, \hat{\mathbf{R}}_{yy}; H_1)$, is calculated using (3.8)-(3.10) with the coefficients $\mathbf{y}_n$ from scale 1 to the smallest scale 2. Similarly, the log likelihood of the underlying RS, $\ln p(\mathbf{y}_n, \hat{\mathbf{R}}_{ss}; H_0)$, is found by using $\mathbf{y}_n$ from scales 3 to 24. Since scale 2 covers the full range of the HS component, $\mathbf{y}_n$ for the rest of the scales can be considered as pure RS. The log likelihood ratio $L(\mathbf{y}_n)$ is high in the presence of the HS component $\mathbf{v}_n$ (under the hypothesis H_1) and it is much smaller under the null hypothesis H_0 . Finally, the decision function $\mathbf{d}$ used to estimate the change over time n is expressed as

$$d(n) = \max_{n-N_d-1 \leq n_s \leq n} \{CS(n_s)\} - CS(n). \quad (3.12)$$

As cumulative sum of the log likelihood ratios instead of instantaneous values is used for HS localization, the decision function seems less sensitive to the choice of the scale.

The choice of N_d has an influence on the localization process as large N_d leads to low probability of localizing short events and small N_d tends to produce false alarms. The length of the shortest HS events being detected has been taken into consideration for the selection of N_d . Since the frequency range of the signal is divided equally into

8 subbands followed by decomposing each subband into 3 octave bands, the minimum duration of HS is therefore scaled from 20 ms to 5 ms. $N_d \leq 200$ is therefore the acceptable range of N_d corresponds to the minimum HS duration at $F_s=44.1$ kHz, and $N_d = 200$ is set here to detect the shortest events.

The envelope $\mathbf{e}$ of the localized HS segments is obtained by

$$e(n) = d(n) \cdot \phi(n) \quad (3.13)$$

where $\phi(n)$ is the instantaneous phase of $y^p(n)$. Since the localization result does not depend much on the accuracy of ϕ , a simple approach based on fast Fourier transform (FFT) is therefore chosen to calculate $\phi(n)$: The FFT is first applied to the TBS signals using a 50% overlapping Hamming window of 512 samples, and $\phi(n)$ is then obtained by phase unwrapping. A moving average (MA) filter with length 512 samples (corresponding to bandwidth of 86 Hz) is finally used to capture the trend of $\phi(n)$ while removing the spikes. Besides MA filter, any other linear spike removal techniques rather than nonlinear smoothing filters can also be applied to capture the trend of $\phi(n)$.

$\phi(n)$ is less sensitive to the choice of window length due to the short durations of HS. Moreover, since $\phi(n)$ depends on the distribution of the signal energy within the window rather than its amplitude or total energy, the estimated envelope $\mathbf{e}$ is insensitive to the abrupt amplitude changes imposed by the presence of adventitious sounds. On the other hand, $\mathbf{e}$ also reduces the HS amplitude variations due to the difference between the first HS (namely S1) and the second HS (namely S2) (as illustrated by Fig. 3.3(b)).

Since the envelope captures all the detailed changes over time corresponding to TBS signal, a threshold γ_e

$$\gamma_e = \frac{1}{N_p} \sum_{n=1}^{N_p} e(n) \quad (3.14)$$

is applied on the envelope to eliminate the effect of noise and therefore to localize the HS segments with N_p being the total number of samples in the noisy signal $\mathbf{y}^p$. Finally,

the identified HS segments with durations smaller than 20 ms are discarded first. The remaining neighboring HS segments with separation less than 50 ms are then combined together. These correspond to the shortest duration of HS and the largest split normal HS interval [67]. The overall structure of the localization scheme for linear/nonlinear parts of signal has been shown in Fig. 3.1(b), and Fig. 3.3 illustrates the steps involved in the HS localization scheme for linear RS recording. The HS localization result for a RS recording containing nonlinear parts is also illustrated in Fig. 3.4 where the resulting HS locations of the linear/nonlinear parts ($\mathbf{A}_g^L/\mathbf{A}_g^{NL}$) before and after concatenation are demonstrated.

Figure 3.3: (a) Decision function $\mathbf{d}$ for the noisy signal $\mathbf{y}$ displayed in (c); (b) Envelope $\mathbf{e}$ of the same signal; (c) Noisy RS signal with HS segment locations $\mathbf{A}$.

The performance of the proposed HS localization approach is measured as a percentage of “true” S1 and S2 activities that have been “accurately” located. A fundamental

Figure 3.4: (a) Noisy RS signal with nonlinear parts located by third-order cumulant (solid line); (b) Linear parts of the noisy RS signal with HS segment locations $\mathbf{A}_g^L$; (c) Nonlinear parts of the noisy RS signal with HS segment locations $\mathbf{A}_g^{NL}$; (d) Noisy RS signal with the output HS segment locations $\mathbf{A}$.

activity S1 has been counted as accurately located with estimation error $\epsilon = 0\%$ if the labeled S1 region of the synthesized signal coincides with the peak in the QRS complex in its synchronously recorded ECG signal. On the other hand, correct S2 locations with estimation error $\epsilon = 0\%$ are decided by the experienced doctors through listening and labeling of the pure HS signals. Any misalignment in the simulated signal increases the estimation error ϵ as defined by

$$\begin{cases} \epsilon_j &= \frac{1}{2} \left\{ \left| \frac{\hat{P}_j^S - P_j^S}{D_j} \right| + \left| \frac{\hat{P}_j^E - P_j^E}{D_j} \right| \right\} \\ \mu &= \sum_{j=1}^J \epsilon_j \\ \epsilon &= \mu \pm \sqrt{\sum_{j=1}^J (\epsilon_j - \mu)^2} \end{cases} \quad (3.15)$$

ϵ_j is the percentage error of the j th located HS segments, with $\hat{P}_j^S$ and $\hat{P}_j^E$ being the estimated starting and end positions of the j th HS segment obtained from the HS location sequence $\mathbf{A}$. P_j^S and P_j^E are the benchmark starting and end positions of the j th HS segment with D_j being the duration of the j th HS segment obtained based on the doctors' decision. J represents the total number of HS segments.

3.3 Results and Discussion

3.3.1 Results on Synthesized Data

Fig. 3.5 illustrates the HS localization results on a synthesized noisy mild wheeze signal by the proposed wavelet based localization method and the entropy based method in [40]. Synchronized ECG signal is also displayed in Fig. 3.5(c) for reference. The performance comparison of these two methods on different types of HS contaminated RS signals is summarized in Table 3.1. Mean and standard deviation of the estimation error between the actual HS locations (based on synchronized ECG signals and doctors' decision) and the estimated HS locations for the synthesized noisy RS signals have been calculated. For performance evaluation on the synthesized signals, the ϵ is calculated for each subject using (3.15) followed by averaging over the subjects.

The proposed HS localization approach gives an overall estimation error as low as $(0.1 \pm 2.50)\%$ for the pure HS signal as indicated in Table 3.1. Furthermore, due to the different signal characteristics, the performance of the presented method is significantly better for wheeze and stridor than that for normal LS. By definition in [36], both stridor and wheeze are classified as continuous adventitious sounds (CAS) which are characterized by their periodic waveforms with a dominant frequency over 100 Hz, while normal LS is characterized by a broad spectral noise. Therefore, in comparison

Figure 3.5: (a) HS localization results for the pure HS signal using the proposed method (solid line) and the entropy based method in [40] (dotted line); (b) HS localization results for the synthesized noisy mild wheeze signal using the proposed method (solid line) and the entropy based method (dotted line); (c) The synchronized ECG signal.

to normal LS, the dynamic changes over time are much smaller for CAS. This results in larger CS difference between the HS void CAS segments and those HS contaminated segments with transient nature. This implies that the localization of HS segments is more accurate for slow-varying RS signals.

Besides the HS void segments in moderate wheeze signals, all the HS including or void segments in other types of signals are having distributions with similar standard deviations. Since the similarity in distribution produces similar entropy for the segments including or void of HS, the performance of the entropy based method in [40]

is degraded. Therefore, the proposed method outperforms the entropy based method for all types of signals except the moderate wheeze signals as depicted in Table 3.1. However, in terms of false detection rate (i.e. false negative and false positive) for HS event, both methods achieve 0.0% due to the relatively high amplitude ratio of HS over LS.

Table 3.1: Comparison of the localization accuracy for different HS localization methods on synthesized data in terms of the mean and standard deviation error ϵ (%)

Type of Signal	The proposed method	The method in [40]
Pure HS	0.1 ± 1.08	2.1 ± 1.26
Normal Adult LS	1.8 ± 1.04	2.0 ± 0.98
Normal Infant LS	1.9 ± 1.33	2.3 ± 1.96
Stridor with Mild Wheeze	0.1 ± 1.30	1.7 ± 2.16
Stridor with Moderate Wheeze	1.7 ± 1.02	0.8 ± 1.02
Polyphonic Wheeze	1.2 ± 1.43	2.6 ± 2.76
Monophonic Wheeze	0.6 ± 1.03	0.6 ± 1.10

3.3.2 Results on Real Data

The real recording dataset consists of TBS recordings obtained as described in Section 3.2.1. Since the patients are having different degrees of airway obstruction, each recording is a different mixture of various CAS. Illustrative results on the real normal and abnormal TBS signals using the proposed methods are shown in Figs. 3.3(c) and 3.4(d). Types of adventitious sound within each mixture as well as HS events locations are manually verified by experienced doctors. Number of HS attacks and starting positions of each HS segments are manually identified as the standard to examine the performance of the proposed method. The performance are summarized in Table 3.2 in terms of errors in correct HS segment detection rate of the proposed wavelet

based method on different types of real TBS signals. The types of various adventitious sounds mixture include: pure normal TBS, pure wheeze, pure stridor, wheeze with stridor, wheeze with rhonchi, and wheeze with squawk. The performance of the localization method is related to the nature of the signals. The more stationary the signal with homogeneous nature (i.e. no abrupt appearance of different adventitious sound), the better the performance of the method. Therefore, the proposed method is very effective for single type signals (i.e. normal TBS, pure wheeze, and pure stridor) with maximum error of $1.8 \pm 4.2\%$.

Table 3.2: The performance of the proposed method on different types of real recorded TBS signals in terms of the mean and standard deviation error ϵ (%)

Type of Signal	False Positive	False Negative	Overall
Normal RS	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Pure Wheeze	1.0 ± 1.0	0.8 ± 1.7	1.8 ± 2.7
Pure Stridor	0.3 ± 0.9	0.2 ± 1.2	0.5 ± 2.1
Wheeze with Stridor	0.2 ± 1.9	0.7 ± 1.3	0.9 ± 3.2
Wheeze with Rhonchi	0.8 ± 1.0	0.9 ± 1.6	1.7 ± 2.6
Wheeze with Squawk	1.2 ± 1.0	1.6 ± 1.8	2.8 ± 2.8

Since we adopt additive model for data synthesis, the signal coupling between different sources is not considered. Therefore, compared to the results on synthesized data, real data has higher error detection rate due to the complexity of the actual signal model. In practice, we assume that the SVD orthogonalization during the implementation of GRLT is able to remove the correlation between $\mathbf{s}_n$ and $\mathbf{v}_n$ (i.e. HS) for real data. The small deviation from the assumption mentioned above downgrades the performance slightly.

The performance of the entropy based method on real data is not as satisfactory as that on synthesized data due to the irregular histograms and standard deviations

Figure 3.6: Normalized histograms of a HS including segment (solid line) and a HS void segment (dashed line) of the signal in Fig. 3.4.

of the segments with and without HS in real data. In compared to the HS including segment, the HS void segment has similarly irregular histogram and even higher standard deviation as indicated by Fig. 3.6. This similarity in distribution implies that the entropy of HS void segment is indistinguishable from that of HS including segment. The HS localization performance is thus not satisfactory enough.

3.4 Chapter Conclusion

This chapter presents a novel algorithm for HS localization from single-channel RS signals. Time scale characteristics of Morlet wavelets has been used to exploit the different characteristics of HS and RS. It is followed by a decision function based on cumulative sum of likelihood ratios capturing the dynamic changes of the signal over time for HS localization. By incorporating the third-order cumulant and the

instantaneous phase, the proposed method seems to be less sensitive to the abrupt changes of signal amplitude caused by the varying nature.

The performance of the proposed method has been evaluated using both synthesized LS signals and real TBS recordings. The slight degradation of the localization performance on real RS recordings compared to the synthesized LS signals verifies the characteristics difference between LS and TBS as well as the inherent correlation between HS and RS due to the sharing of transmission path. The results with real RS recordings also show that higher stationarity in RS signals leads to higher localization accuracy.

Since HS localization depends on the difference between HS and RS in terms of their dynamic behaviors over time, the challenges remain for the localization of abnormal HSs (e.g. murmurs) with different dynamics as well as for HS localization in the presence of discontinuous adventitious sounds (such as crackles) which have similar transient nature as the HS signal. However, experimental results for synthesized LS and real TBS data validate the robustness and the high accuracy of the proposed method for various types of continuous adventitious sounds which are the core in RS analysis. Since none of the existing HS localization methods works well for different types of RS, this proposed method is a successful first attempt to open up the exploration into a greater field.

Chapter 4

Filter Bank Based Source Extraction for Heart Sound Removal

4.1 Introduction

Several filtering techniques have been adopted for HS removal from the localized HS contaminated RS segments. These include wavelet transform [71][72] which have been summarized in [73] and linear adaptive filtering [59][74][5][75][6][60][7]. Since the previous attempts in [74][71][5] result in reduced but still audible HS, [61][60][7] provide complete HS removal through cutting off the HS containing segments followed by interpolation of the missing data. A detailed description and comparison of the existing HS removal methods have been presented in [63]. However, none of the methods except [71] has attempted to remove the HS from pathological RS other than normal

RS. In fact, the dynamic signal characteristics together with the frequency overlapping due to the presence of low-pitched adventitious sounds (e.g. wheeze and stridor) pose problems for the existing HS removal methods. Therefore, an improved method for HS removal with high accuracy is required to be developed to accommodate various types of RS.

In this chapter, a source extraction algorithm is proposed by generating basis functions using filter banks in order to extract the underlying RS from each located HS corrupted segment of an input noisy RS signal. By using a constructed RS template containing *a priori* information as input, the desired basis functions are obtained as outputs of the filter bank. Each basis function of a selected set is then enlarged by an FIR filter matched with the observed signal by the mean-square error criteria. Thereby, the desired RS signals can be extracted without the loss of signal information even in the presence of adventitious sounds.

4.2 Methodology

4.2.1 Signal Characteristics

Both RS signal as broad spectrum noise [36] and HS signal which has impulsive transient nature show nonstationary behavior. On the other hand, as illustrated in Fig. 4.1(a), the impulsive transient HS has larger amplitude variation and shorter duration compared to RS that is slow-varying due to the low amplitude variability. Moreover, the high zero-crossing rate of RS implies the presence of high frequency components, while the smoother waveform of HS with lower zero-crossing rate indicates the lower frequency range of HS. However, this lower frequency range of HS as depicted by the spectra in Fig. 4.1(b) does not imply the slow-varying character that

depends solely on the amplitude variations in time domain.

Figure 4.1: (a) Waveforms of pure HS and RS recordings; (b) Spectra of the HS and RS signals in (a).

4.2.2 Basic Concept

The source extraction algorithm proposed here can be developed in the following way.

A set of M linearly independent basis functions $\underline{\phi}_{\{i|i=1,\dots,M\}}$ is first considered where

$$\underline{\phi}_i = [\phi_i(1) \cdots \phi_i(N)]^T. \quad (4.1)$$

Each of the basis functions is processed by an adaptive FIR filter, the parameter vector of which is $\mathbf{b}_{\{i|i=1,\dots,M\}}$ of length $(N_b + 1)$.

The corresponding output is the sequence

$$\hat{\mathbf{s}}_i = \sum_{k=0}^{N_b} b_i(k) \phi_i(n - k) = \Phi_i \mathbf{b}_i \quad i = 1, \dots, M \quad (4.2)$$

where

$$\Phi_i = \begin{bmatrix} \phi_i(1) & 0 & \cdots & 0 \\ \phi_i(2) & \phi_i(1) & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_i(N_b + 1) & \phi_i(N_b) & \vdots & \phi_i(1) \\ \vdots & \vdots & \cdots & \vdots \\ \phi_i(N) & \phi_i(N - 1) & \cdots & \phi_i(N - N_b) \\ 0 & \phi_i(N) & \cdots & \phi_i(N - N_b + 1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_i(N) \end{bmatrix} \quad (4.3)$$

$$\mathbf{b}_i = \begin{bmatrix} b_i(0) & b_i(1) & \cdots & b_i(N_b) \end{bmatrix}^T. \quad (4.4)$$

The matrix $\Phi_{\{i|i=1,\dots,M\}}$, has a size of $(N + N_b) \times (N_b + 1)$. The sum of the M sequences provides $\hat{\mathbf{s}}$, which is the estimate of a given signal $\mathbf{s}$ that belongs to the signal space $\mathbf{s}_{\{j|j=1,\dots,J\}}$,

$$\hat{\mathbf{s}} = \sum_{i=1}^M \hat{\mathbf{s}}_i. \quad (4.5)$$

Using (4.2), the estimated signal can be expressed as

$$\hat{\mathbf{s}} = (\Phi_1 \cdots \Phi_M)(\mathbf{b}_1 \cdots \mathbf{b}_M)^T = \Phi \mathbf{b}. \quad (4.6)$$

In this way, the set of basis functions is enlarged by producing $(N_b + 1)$ delayed copies of the original set. The unknown parameter vector $\mathbf{b}$ is determined in a way which minimizes the weighted square of the Euclidean norm

$$J(\Phi, \mathbf{b}) = \|\sqrt{\mathbf{W}}(\mathbf{y} - \hat{\mathbf{s}})\|^2 = (\mathbf{y} - \Phi \mathbf{b})^T \mathbf{W}(\mathbf{y} - \Phi \mathbf{b}) \quad (4.7)$$

with the weighting matrix $\mathbf{W}$ being a non-singular symmetric diagonal matrix and $\mathbf{y}$ as the observed signal of length N consists of signal plus noise. The optimum parameter vector $\mathbf{b}_{opt}$ is given by [69]

$$\mathbf{b}_{opt} = (\Phi^T \mathbf{W} \Phi)^{-1} \Phi^T \mathbf{W} \mathbf{y}. \quad (4.8)$$

Since both Φ and $\mathbf{W}$ have full rank, the matrix $(\Phi^T \mathbf{W} \Phi)^{-1}$ exists and the corresponding minimum norm

$$\begin{aligned} J(\Phi, \mathbf{b}_{opt}) &= \|\sqrt{\mathbf{W}}\mathbf{y} - \Phi \mathbf{b}_{opt}\|^2 \\ &= \|\mathbf{y}\|^2 - \mathbf{y}^T \mathbf{W} \Phi (\Phi^T \mathbf{W} \Phi)^{-1} \Phi^T \mathbf{W} \mathbf{y} \end{aligned} \quad (4.9)$$

is dependent on the choice of basis functions. Basically, the basis functions decompose a noisy signal into signal space and noise space. Compared to fixed basis set such as discrete cosine transform (DCT), the signal dependent basis functions used in (4.8) are more suitable for calculating the coefficients of the time-varying filter $\mathbf{b}_{opt}$. These basis functions are generated by a template signal through a bandpass filter bank, as elaborated in Section 4.3.4.

4.3 Algorithm Implementation

The basic principle of the proposed source extraction algorithm presented here is the matching of the low frequency components of the observed signal $y_j(n)$ with a template signal $x_j(n)$. The wavelet based method in Chapter 3 is applied here to locate the HS corrupted segments of the noisy RS signals.

4.3.1 Outline of the Proposed Algorithm

As shown in Fig. 4.2, the proposed semi-blind, single-channel source extraction algorithm for HS removal consists of four parts: A localization scheme, a template design scheme, a high-frequency shelving filter $\mathbf{H}_{LP}$, and a novel source extraction scheme. Our source extraction scheme consists of a pair of identical filter banks and adaptive FIR filters for template matching. The HS segments are localized and extracted from the noisy signal first by employing a suitable HS localization scheme. The respective

HS-free RS segment prior to each located HS segments is then extracted to generate the reference template by a template design scheme. After template design, the low-frequency components from both the designed templates containing *a priori* knowledge of the underlying RS, as well as the HS corrupted segments are extracted by a shelving filter. In this way, the underlying RS signals can be obtained as $\hat{\mathbf{s}}_{\{ij|i=1,\dots,L\}}$ based on these low-frequency components through the proposed source extraction scheme, with L indicating the total number of subfilters in the filter bank. Finally, the unaltered HS-free high-frequency components of the located segments are added back to the outputs of the source extraction scheme.

Figure 4.2: An overall block diagram of the proposed semi-blind single-channel source extraction algorithm.

4.3.2 The Proposed Template Design Scheme

The preliminary template $y_j^t(n)$ is first selected based on the input segment prior to the localized $y_j(n)$. Time scaling is then applied on the selected $y_j^t(n)$ to produce the

template $y_j^T(n)$ which has the same length as its respective $y_j(n)$. Our template design scheme is based on the signal character that the amplitude variation of the pure RS signal is not as large as that of HS signal over a short time interval. The presented template design scheme is illustrated by the flow-diagram in Fig. 4.3, with the detailed steps involved being described as below.

Figure 4.3: A block diagram of the proposed template design scheme.

Template Selection

The input signal $\mathbf{y}$ is first normalized by its mean value to locate the zero-crossing points. The temporal locations of these zero-crossing points are then labeled by a set of nodes $A_{\{a|1 \leq a \leq N_c\}}$ with $A_{N_c} - A_1 = N$ where N_c indicates the nearest zero-crossing point before the starting point of $\mathbf{y}_j$. The segments in between the consecutive nodes A_a and A_{a-1} are denoted by a set of $\mathbf{y}_{\{a|1 \leq a \leq N_c-1\}}$.

Let $w_a > 0$ be the score of $\mathbf{y}_a$ which indicates the peakedness of the segment, a high value of w_a should be induced by rapid changes (i.e. presence of impulses) in $\mathbf{y}_a$. A suitable definition of w_a is therefore

$$w_a = \frac{\max(|\mathbf{y}_a(n)|)}{A_a - A_{a-1}}. \quad (4.10)$$

The score w_a is calculated for each segment $\mathbf{y}_a$ with a values descending from $a = N_c$ until $a = N_w$ where $w_{\{a|a=N_w \geq 1\}} > 1$. The selected template $\mathbf{y}_j^t$ is therefore the part of the signal from A_{N_w} to A_{N_c} . The starting point for $\mathbf{y}_j$ is also adjusted to N_c in order to maintain the continuity. If the selected template $\mathbf{y}_j^t$ is shorter than $\mathbf{y}_j$, then $\mathbf{y}_j^t$ is time scaled to construct $\mathbf{y}_j^T$ which is of the same length as $\mathbf{y}_j$ using the time scaling technique described below. Otherwise, $\mathbf{y}_j^t$ is directly used as the template $\mathbf{y}_j^T$ for signal extraction.

Time Scaling

Time scaling techniques are usually employed when we want to modify the length of an audio signal without losing the perceptual quality by expanding the signal through time scale expansion. A phase vocoder based time scaling approach is employed here to expand the obtained template signal $\mathbf{y}_j^t$ with a sequence of analysis, modification, and resynthesis.

The time stretching algorithm adopted preserves the magnitude as well as the instantaneous frequencies, and the signal can be then reconstructed as the weighted sum of cosines using a filter bank approach or an IFFT approach. The detailed steps involved to scale $\mathbf{y}_j^t$ to $\mathbf{y}_j^T$ are elaborated in Appendix B.

4.3.3 The Implementation of Shelving Filter

Since HS signals interfere with RS signals only at low frequencies up to 500 Hz of the recorded RS, a high-frequency shelving filter implemented in [76] is applied on the observed input signal prior to the source extraction scheme (see Fig. 4.2). The bandwidth of the linear FIR filter $\mathbf{H}_{LP}$ is set to be [20 800] Hz with a corresponding high frequency gain $G_h = 1$. Only low frequency components of the observed signal are therefore input to the source extraction, while the high frequency components are added back unaltered during signal synthesis. The low-frequency components of $y_j(n)$ are then passed through an identical filter bank to produce outputs $\mathbf{z}_{\{ij|i=1,\dots,L\}}$. At the same time, the low-frequency components of $y_j^T(n)$ are put through the same set of identical filter banks that allow us to use the decimation procedure.

4.3.4 The Proposed Signal Extraction Scheme

A linear perfect reconstruction bandpass filter bank has been designed based on [77][78] for the proposed signal extraction scheme. The identical filter bank therefore consists of $L = 32$ FIR filters, each with length of 256 samples. The magnitude response of the filter bank is illustrated in Fig. 4.4.

The basis functions are obtained as

$$\phi_i(n) = h_i(n) * y^T(n), \quad 1 \leq i \leq L, \quad i \in I \quad (4.11)$$

Figure 4.4: Magnitude responses of (a) a prototype lowpass filter; and (b) the linear reconstruction bandpass filter bank used in Fig. 4.2 with $L = 32$ and filter length of 256.

where ‘*’ denotes the convolution operator and $h_i(n)$ is the impulse response of the i th subfilter in the filter bank. The basis set I of size M in (4.11) is chosen according to

$$\min_I \left\{ \frac{\sum_{i \in I} E_i}{\sum_{i=1}^L E_i} \right\} \geq \gamma \quad (4.12)$$

with $E_i = \sum_{n=1}^N \phi_i^2(n)$ being the energy of the i th basis function and $\gamma < 1$ is a given parameter.

The FIR filter coefficients are calculated according to (4.8) by minimizing the squared error $\|\mathbf{e}_{ij}\|^2 = \|\mathbf{z}_{ij} - \Phi_i \mathbf{b}_{ij}\|^2$ during the implementation with Φ_i given by (4.3), for each channel separately. The optimum parameter vector $\mathbf{b}_{ijopt}$ is thus

$$\mathbf{b}_{ijopt} = (\Phi_i^T \Phi_i)^{-1} \Phi_i^T \mathbf{z}_{ij}. \quad (4.13)$$

When Φ_i is ill conditioned, it is decomposed by SVD according to [69]

$$\Phi_i = \mathbf{U}_i \mathbf{S}_i \mathbf{V}_i^T. \quad (4.14)$$

The matrices $\mathbf{U}_i$ and $\mathbf{V}_i$ are unitary, and matrix $\mathbf{S}_i$ is written as

$$\mathbf{S}_i = \begin{bmatrix} \sum_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (4.15)$$

where $\sum_i = \text{diag}(\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{i\alpha 1})$. $\alpha 1$ is the rank of the matrix Φ_i in (4.14), and $\mathbf{0}$ is the null matrix. Inserting (4.14) into (4.13), we obtain [79]

$$\mathbf{b}_{ijopt} = \sum_{l=1}^{\alpha 1} \frac{(\mathbf{u}_{il}^T \mathbf{z}_{ij})}{\sigma_{il}} \mathbf{v}_{il}. \quad (4.16)$$

The extracted signal in the i th band at the j th epoch then becomes

$$\hat{\mathbf{s}}_{ij} = \mathbf{y}_j^T * \mathbf{h}_{ij} * \mathbf{b}_{ijopt}, \quad i \in I. \quad (4.17)$$

Finally, the extracted signal is obtained as

$$\hat{\mathbf{s}}_j = \sum_{i \in I} \hat{\mathbf{s}}_{ij}. \quad (4.18)$$

The detail steps for the selection of γ_{opt} and the optimal order of $\mathbf{b}_{iopt}$ are discussed below:

Selection of γ_{opt}

The selection of basis set I of size M and thus the number of required sub-bands/subfilters are determined by γ_{opt} . The selected set I in the numerator of (4.12) corresponds to the filters with the highest output energies, where the sum of all these energies should not be less than $\gamma_{opt}\%$ of the total energy.

The power spectral density (PSD) analysis is used to determine the value of γ_{opt} . The average PSDs of the HS corrupted segments are found over the selected frequency

bands. The average power difference between the input HS corrupted segments and the output reconstructed RS segments are therefore calculated over each frequency band for various types of noisy RS recordings. The γ_{opt} value is then calculated based on the averaged input SNR of the localized HS segments as shown by

$$\gamma_{opt} = \sum_{k=1}^{K_{opt}} \left(\frac{SNR^k}{\sum_{k=1}^{K_{opt}} SNR^k} \right) \gamma_{opt}^k. \quad (4.19)$$

In (4.19), K_{opt} corresponds to the total number of frequency bands selected with $K_{opt} = M$ here and γ_{opt}^k refers to the γ value that produces minimum average power difference at the k th frequency band. SNR^k representing the averaged input SNR at the k th frequency band is defined as

$$SNR^k = \frac{1}{J} \sum_{j=1}^J SNR_j^k = \frac{1}{J} \sum_{j=1}^J \left(\frac{P_{xj}^k}{P_{yj}^k} \right) \quad (4.20)$$

where P_{xj}^k/P_{yj}^k refers to the power of the j th template/HS containing segment, respectively.

Order of the Optimum Filters

Pure RS being a relatively slow time-varying signal, is able to provide higher temporal continuity compared to HS. The optimum $\mathbf{b}_{iopt}$ filters are therefore selected by (4.16) to capture the signals having temporal continuity (i.e. having small changes over time).

The temporal continuity ζ is measured as

$$\zeta = \sum_{i=1}^I \sum_{n=2}^N (g_{n,i} - g_{n-1,i})^2 \quad (4.21)$$

with $\mathbf{g}_i = [g_{i,1} \cdots g_{i,N}]^T$, $i = 1, \dots, I$ and $\mathbf{g}_i = \mathbf{h}_i * \mathbf{b}_{iopt}$ (T and $*$ stand for transpose and convolution). The effect of $\mathbf{b}_{i|i=1,\dots,I}$ filters can be shown by the different ζ values for $\mathbf{g}_i = \mathbf{h}_i * \mathbf{b}_{iopt}$ and $\mathbf{g}_i = \mathbf{h}_i$. It is found that ζ varies for different $\mathbf{y}_j$ segments and ζ is reduced by the use of $\mathbf{b}_{iopt}$ filters that favor the temporal continuity (see Fig. 4.5).

In order to keep the temporal continuity (i.e. small ζ value) for all the HS corrupted segments, the minimum order of the optimum filters $\mathbf{b}_{iopt}$ is set to be $N_b = 5$ in this chapter.

Figure 4.5: The temporal continuity ζ with, order of the optimum filters $\mathbf{b}_{iopt}$, $N_b = 5$.

4.4 Experimental Results and Discussion

4.4.1 Experimental Dataset

The same experimental dataset used has been obtained from Section 3.2.1. Four different sets of experimental data have been used in this chapter for parameter selection as well as quantitative analysis:

Set #1: 10 simulated HS contaminated RS signals including 5 normal LS, 3 pure wheeze signals, and 2 signals with inspiratory stridor and expiratory wheeze, each of length 6s.

Set #2: 5 simulated HS contaminated normal RS signals, each of length 6s.

Set #3: 15 real recorded normal RS signals, each of length 6s.

Set #4: 10 real recorded wheeze signals, 10 real recorded stridor signals, 20 real recorded signals as mixtures of stridor and wheeze, each of length 6s.

4.4.2 Analysis of γ_{opt}

The steps described in Section 4.3.4 are carried out on our experiment dataset. Average power difference between the output RS segments and their corresponding template segments are calculated by varying γ value between $[0.5 \ 1]$ with a step size of 0.05 over each of the 6 selected frequency bands: 20 – 40 Hz ($k = 1$), 40 – 70 Hz ($k = 2$), 70 – 150 Hz ($k = 3$), 150 – 300 Hz ($k = 4$), 300 – 500 Hz ($k = 5$), 500 – 800 Hz ($k = 6$). A Hamming window of length 23.2 ms with an overlap of 1.45 ms between successive windows is adopted to calculate the PSDs of the corresponding segments. The PSDs for different frequency bands obtained by Welch method are averaged over the segments for different signals and the average power differences are then computed using the method in [80].

Fig. 4.6 indicates how the average power difference changes by varying γ in each frequency band. The γ_{opt}^k is selected to be the γ^k value with the minimum power difference in the k th frequency band. The selected γ_{opt}^k values together with their respective SNR^k values are listed in Table 4.1. As the value of γ_{opt} depends largely on SNR^k of the input signal, an ascending trend of γ_{opt}^k with increasing SNR^k has been observed. $\gamma_{opt} = 0.70$ is then calculated according to (4.19) using Set #1 of the experimental dataset. Since the dataset contains a wide range of continuous adventitious sounds (CAS) from different subjects, $\gamma_{opt}=0.70$ is then adopted as a reliable parameter for HS cancellation of other RS signals in Sets #2, #3, #4.

Figure 4.6: The average power differences with changing γ in different frequency bands.Table 4.1: SNR^k and γ_{opt}^k for different frequency bands on experimental dataset #1

Frequency Bands k	SNR^k (dB)	γ_{opt}^k
1 (20 – 40 Hz)	−14.4135	0.65
2 (40 – 70 Hz)	−12.0985	0.65
3 (70 – 150 Hz)	−10.9169	0.75
4 (150 – 300 Hz)	−11.0289	0.65
5 (300 – 500 Hz)	−10.3683	0.80
6 (500 – 800 Hz)	−0.6604	0.99

4.4.3 Auditory Test

A subjective test by the skilled clinicians have been performed on the reconstructed HS-free RS signals (from both real TBS recordings and LS in [37][64]). By listening to the output signals, no abrupt changes or additional noises in the HS contaminated segments have been reported. This shows the ability of the proposed algorithm to produce HS-free RS signals with high sound quality. Some sound clips of the reconstructed RS signals can be found together with their corresponding HS corrupted inputs at http://www.ntu.edu.sg/home/jinfeng/hs_cancellation.htm for demonstration purpose.

4.4.4 Quantitative Analysis

Fig. 4.7 illustrates the waveforms and the respective spectrograms of a real recorded wheeze signal as well as its reconstructed HS-free RS signal obtained by the proposed method. These plots verify that the proposed HS removal method has the ability to completely remove the presence of HSs. In addition, Fig. 4.8 presents the signal waveform and the spectrogram of the reconstructed RS signal obtained by using the noisy input segment prior to $y_j(n)$ as template without applying our template design scheme. By comparing the spectrograms of the reconstructed RS signal in Figs. 4.7 and 4.8, it is clearly seen that the inclusion of the template design scheme successfully prevents the power leakage and improves the signal continuity by providing template segments $\mathbf{y}_j^T$ with the closest proximity to the underlying RS in $\mathbf{y}_j$ segments.

Figure 4.7: The waveforms and spectrograms of a real recorded wheeze signal (top two) and its reconstructed HS-free RS signal (bottom two).

Figure 4.8: The waveform (top) and spectrogram (bottom) of the reconstructed HS-free RS signal for the same noisy wheeze recording in Fig. 4.7 without template design scheme.

Besides visual and auditory observations, the PSDs of all $\mathbf{y}_j^T$ segments as well as their respective reconstructed $\hat{\mathbf{s}}_j$ segments are computed for every signal in the experimental dataset. Fig. 4.9 shows the average PSDs of the simulated noisy RS signals before and after HS removal by the presented method using Set #2. Also, the performances of our method as well as the linear prediction method in [7] for real normal RS recordings from Set #3 are illustrated in Fig. 4.10 in terms of their average PSDs. For an effective HS removal method, the average PSD of the reconstructed RS signal should coincide or be close to that of the original HS-free RS signal. Therefore, comparing the reconstructed RS signal obtained by using these two methods, the presented method is shown to be more effective by producing a closer approximation to the template PSD.

Figure 4.9: Average PSDs of the original HS-free normal LS signal, the simulated HS corrupted LS signal, and the reconstructed LS signal using the proposed method.

Figure 4.10: Average PSDs of the original HS-free RS signal, the HS corrupted RS signal, and the reconstructed RS signal using the proposed method and the linear prediction method in [7].

Table 4.2: Comparison of the average power difference ($\mu \pm \sigma$)(dB) between the PSDs of the original and the reconstructed normal RS by different HS cancellation methods

Frequency Bands k	The Proposed Method	The Method in [7]
1 (20 – 40 Hz)	2.1977 ± 1.4349	4.6484 ± 2.4532
2 (40 – 70 Hz)	1.6662 ± 0.8754	2.0415 ± 0.6773
3 (70 – 150 Hz)	1.2567 ± 0.8732	1.0437 ± 0.7683
4 (150 – 300 Hz)	3.6688 ± 3.0105	8.6670 ± 6.5433
5 (300 – 500 Hz)	0.1806 ± 0.1694	4.6102 ± 1.2344
6 (500 – 800 Hz)	1.7534 ± 1.6443	1.8345 ± 1.7247

The average power difference between the original and the reconstructed RS signals is further calculated for qualitative evaluation. Mean (μ) and standard deviation (σ) of the average power differences evaluated over different frequency bands by different HS cancelation methods for normal RS (Set #3), are presented in Table 4.2. It can be seen that the proposed method outperforms the method in [7] for all frequency bands except 70–150 Hz. The presented method has the advantage of producing more accurate localization for HS and template segments to ensure accurate reconstruction. Furthermore, the method is not limited by the types of RS signals used. On the other hand, the method in [7] fails in accurate HS localization and therefore is not able to provide high performance. In this way, it verifies the speculation that the RS reconstruction accuracy depends on its HS localization accuracy.

Moreover, the performance of the proposed method on different types of CAS in Set #4, is listed by Table 4.3 in terms of average power difference. It can be observed that the presented algorithm is more effective for the wheeze signals than those stridor signals. This is due to the fact that HS interference occurs at low frequencies, and stridor being a low-pitched wheeze, has more prominent low frequency components

Table 4.3: The average power difference ($\mu \pm \sigma$)(dB) between PSDs of the original and the reconstructed RS of different types by the proposed method

Frequency Bands k	Wheeze	Stridor	Stridor and Wheeze Mixture
1 (20 – 40 Hz)	1.0936 ± 1.0044	2.2824 ± 1.2439	2.0364 ± 2.0023
2 (40 – 70 Hz)	0.7440 ± 0.6872	1.7952 ± 0.9763	1.5698 ± 1.5682
3 (70 – 150 Hz)	0.2775 ± 0.1345	1.2683 ± 0.7742	1.0329 ± 1.4592
4 (150 – 300 Hz)	0.5808 ± 0.3268	0.5407 ± 0.4537	0.5532 ± 0.5526
5 (300 – 500 Hz)	1.7644 ± 0.6734	0.5409 ± 0.4364	1.5964 ± 1.3965
6 (500 – 800 Hz)	2.8707 ± 0.9875	1.0263 ± 1.0031	2.1349 ± 1.6829

than wheeze. The HS interference is therefore more prominent for stridors, which gives slightly less accurate results. While for those signals containing mixture of wheeze and stridor, the performance lies in between that for pure wheeze or pure stridor. Generally speaking, the proposed method performs better on wheeze and stridor than on normal RS. This is due to the relatively slow-varying character of CAS. Although the presence of adventitious sounds poses problems for HS localization, it does not affect the performance of the proposed HS cancelation method by incorporating an effective HS localization scheme.

4.5 Chapter Conclusion

This chapter proposes a novel method for HS removal for various types of single source RS. The presented nonparametric algorithm is able to perform accurate single trial RS extraction and HS removal without the aid of any training dataset. In compared to the existing HS cancellation method [81] based on parametric approach, the accuracy of the reconstructed HS-free signal for the proposed algorithm is not limited by the size and location of the HS segments. The algorithm incorporates a high-frequency shelving filter and an appropriate template design scheme to solve the problems of

power leakage in high frequency bands and signal discontinuity of the reconstructed RS. This also ensures the algorithm to perform quite well for different types of RS.

Both subjective listening test by the experienced clinical doctors, as well as quantitative analysis of the reconstructed RS signals have been performed. The results confirm that the proposed method is able to remove completely the HS artifacts, while maintaining the quality of the RS signals for the single-channel RS recordings. As the presented method does not involve any prediction scheme, it leads to a more accurate RS signal analysis such as wheeze detection based on the output RS signals.

Part II

Acoustical Respiratory Phase Detection

As respiratory sounds (RSs) are primarily related to vibrations of the airway walls and the turbulent airflow [36], breath monitoring is essential in numerous clinical situations. RS can be segmented into four successive phases: inspiratory phase, end-inspiratory pause, expiratory phase, and end-expiratory pause. Precise timing of the respiratory phases is crucial in many situations such as heart flow studies [82], MRI investigations [83], as well as computerized pulmonary auscultation. As indications of infectious and respiratory diseases, many pathological RS are to be clinically characterized by their duration in respiratory cycle and their relationship to the phase of respiration [84][85][86]. Thus it is necessary in adventitious sound characterization to segment the RS into individual respiratory cycles and then their respective respiratory phases.

Generally, phonopneumography or spirometer together with sound recording devices are used in RS analysis, where the waveform of the signal is displayed simultaneously with the airflow as a function of time. Signals can be segmented into inspiratory and expiratory phases according to the Forced Expiratory Volume (FEV) readings [87][88] measured. However, they have suffered from the limitation that the accurate flow measurement depends on mechanisms which affect the natural breathing pattern [89]. This makes spirometric test failed in many situations, especially for patients with high obstruction in trachea [88] or with neurological impairments [10]. Attempts have therefore been made to relate flow with acoustical RS. Airflow has been estimated using RS by the application of different models together with training dataset [90][91][10]. A relatively high estimation accuracy has been achieved in [10]. The usage of Shannon entropy makes the approach insensitive to changes of RS amplitudes but the predefined linear model used does not support flow estimation for different types of RS other than normal RS. Furthermore, training dataset is required to enable the flow estimations.

An accurate respiratory phase detection has therefore been achieved in Part II of the thesis through two stages: phase segmentation using genetic approach as introduced in Chapter 5 to locate the exact timing of each respiratory phases; and phase identification based on an annotating index as proposed in Chapter 6 to correctly annotate inspiratory, expiratory, and pause phases from the segmented phases.

Chapter 5

Respiratory Phase Segmentation Using Genetic Approach

5.1 Introduction

Due to the incapability of flow estimation, indirect phase segmentation methods by acoustical analysis of RS have been recently presented as an alternative solution for respiratory phase monitoring. Approaches based on spectral and temporal analysis of the transformed RS signals have been suggested for respiratory phase segmentation. Combined investigations of signal envelope, frequency content as well as disturbance characteristics have been applied for respiratory rate (**RR**) monitoring [41] and phase segmentation [9]. In [8], quasi-periodicity of short-term signal energy has been deployed for phase onset detection of normal lung sounds (LSs). Since the existing methods depend on either spectral content or short-term energy of the signal, the phase segmentation results are strongly influenced by the amplitude of the input signal. To avoid such problems, variance fractal dimension being independent of signal power has been adopted in [43] and an accuracy of 40 ± 9 ms has been reported for normal RS.

Moreover, a nomenclature of respiratory cycle from a modeling perspective has been introduced in [44] by adding a transitional phase between the inspiratory/expiratory phase. Triplet Markov chain has been performed in wavelet packet domain to improve the segmentation accuracy in [44]. However, *priors* on the respiratory cycle structure for normal RS has been exploited for chain adaptation which makes the method not robust for adventitious sounds.

Hence, the aim of this chapter is to propose a fully automatic respiratory phase segmentation method which is robust for different types of RS using single-channel recordings. Noise spectral estimation is exploited to calculate the total number of respective respiratory phases present in RS first. Respiratory phase segmentation is then proposed based on a stochastic global search method. In order to guide the search space of the multi-population genetic algorithm (MPGA) and thus to locate the phase boundaries, an evaluation function combining sample entropy (*SampEn*) calculation and heterogeneity measure is introduced. The performance of the proposed method is compared with that of the fractal dimension method in [43] for different types of real RS recordings.

5.2 Methodology

The overall structure of our respiratory phase segmentation method is shown in Fig. 5.1. The input noisy RS sequence is indicated by $y(n)$, and $x(n)$ denotes the preprocessed clean RS. $|\hat{e}_N(n)|$ is the estimated noise envelope for respiratory rate (**RR**) estimation with $|\hat{e}_N^m(n)|$ being its nonlinear mapped version and **RR** indicating the estimated total number of respiratory cycles per minute. The output of the algorithm $\mathbf{A}_r$ is the sequence containing locations of all respiratory phase boundaries in $y(n)$.

Figure 5.1: A general block scheme of the proposed respiratory phase segmentation method.

5.2.1 Data Acquisition

Real recordings were done following the procedures as described by Section 2.6 with $F_s = 11.025$ kHz. In this study, the real recorded dataset consists of TBS from 7 healthy subjects and 14 subjects with different degrees of airway obstruction (8 males/13 females, 15 ± 9 years old). The characteristics due to sex, age, weight have not been taken into consideration. Also, 10 preprocessed LS recordings from [37][64] have been used to investigate the effect of the preprocessing scheme proposed in Section 5.2.2.

Reference signals were obtained from doctors by listening to the RS recordings. Each observer pressed one button at inspiratory phases and another button at expiratory phases while he/she examined the recordings according to [9]. The correct locations for respiratory phases boundaries have been decided by averaging the synchronized reference signals obtained from four well experienced doctors in National University Hospital, Singapore. According to the reference signals, RS signals have been segmented into four successive phases: inspiratory phase, end-inspiratory pause,

expiratory phase, and end-expiratory pause.

5.2.2 Signal Preprocessing

Spike Removal

The original RS recording $y(n)$ is first bandpass filtered in the range of 300–1000 Hz to remove the effects of heart sounds (HSs) and high-frequency noise [10]. Nonlinear energy operator (NEO) is then applied on the resulting bandpass filtered signal to detect the remaining spikes for removal. Spike as a reflection of the transient noises, is mainly due to contact noise in this chapter. The transient noise is characterized by a short burst of acoustic energy of either a single impulse or a series of impulses, with a wide spectral bandwidth [92]. It therefore exists over the bandpass filtered RS signal ranging with 300–1000 Hz. Fig. 5.2 shows the power spectral density (PSD) of a spike-contaminated RS segment and that of a spike-free RS segment of the same length. The broadband nature of the spike can be observed clearly with an increased power below 300 Hz and a higher power over the range 300–1000 Hz compared to the spike-free RS signal.

In order to remove the spikes caused by transient noises, NEO is considered to be more suitable than those average signal energy based estimators. The spike detection ability of NEO is independent to the frequency contents of the outburst energy and it uses quadratic filter which can better reflect the local activity. NEO is defined in [93] as

$$\Psi[y_f(n)] = [y_f(n)]^2 - y_f(n+1)y_f(n-1) \quad (5.1)$$

with $y_f(n)$ as the bandpass filtered input signal having spikes. The peaks can be further

identified by using a threshold γ_Ψ chosen as

$$\gamma_\Psi = \frac{c}{N_f} \sum_{n=1}^{N_f} \{\Psi[y_f(n)] * w(n)\} \quad (5.2)$$

where ‘ $*$ ’ denotes convolution and $w(n)$ is a 6-point Bartlett window. N_f is the length of $y_f(n)$ and c is a scaling factor for spike detection. Comparing with other windows such as Hamming and triangular windows, Bartlett window has narrow main lobe which provides good resolution and it has small support that ensures low complexity [94]. The use of Bartlett window ensures sufficient reduction of noise interference while keeping low complexity for the algorithm. The values of the located spikes are assigned to be 1 and the detection process is iterated for 10 times to detect the adjacent spikes. A global mean of $y_f(n)$ is then used to replace the detected spikes, producing a spike-free signal $y_{sf}(n)$. The dotted line in Fig 5.2 shows a spike-contaminated RS segment after spike removal using NEO, and Fig. 5.3 depicts the illustrative results of bandpass filtering and spike removal for a real recorded stridor signal.

Figure 5.2: PSDs of a spike-contaminated RS segment (solid line), a spike-free RS segment of same length (dash line), and a spike-contaminated RS segment after spike removal using NEO (dotted line).

Figure 5.3: (a) Original waveform $y(n)$ of a real recorded stridor signal; (b) The band-pass filtered signal $y_f(n)$; (c) The resulting spike-free signal $y_{sf}(n)$ after spike removal.

Sparseness Reduction

In this chapter, a sparse sequence refers to one with most samples close to zero while only a few have prominent non-zero values. As a sparse signal contains only a few representative samples [95], the resulting *SampEn* decreases abnormally at those sparse parts and thus diverges the following GA search for the localization of phase boundaries. Therefore, sparse frames which have been identified by the following measurement are to be replaced by sparseness reduction.

The sparseness measure, $\Gamma(\cdot)$, based on the relationship between L_1 and L_2 norms of the windowed signal is defined here as

$$\Gamma(y_{sf}^w) = \frac{\sqrt{N_w} - \overbrace{\left(\sum_{n=1}^{N_w} |y_{sf}^w(n)| \right)}^{L_1} / \overbrace{\left(\sum_{n=1}^{N_w} (y_{sf}^w(n))^2 \right)}^{L_2}}{\sqrt{N_w} - 1} \quad (5.3)$$

where N_w is the length of the w th windowed frame $y_{sf}^w(n)$. The $\sum_{n=1}^{N_w} |y_{sf}^w(n)|$ term refers to the L_1 norm while $\sqrt{\sum_{n=1}^{N_w} (y_{sf}^w(n))^2}$ refers to the L_2 norm. $\Gamma(y_{sf}^w)$ is calculated for each $y_{sf}^w(n)$ obtained by using a non-overlapping rectangular window. The measure evaluates to unity if and only if $y_{sf}^w(n)$ contains only a single non-zero component, and takes a value of zero if and only if all components are equal.

Figure 5.4: (a) The resulting $y_{sf}(n)$ after bandpass filtering and spike removal of a real recorded wheeze signal; (b) The corresponding $\Gamma(y_{sf}^w)$ displayed together with threshold $\bar{\Gamma}$.

The sparse frames in $y_{sf}(n)$ are located by thresholding the measured Γ sequence using its global mean $\bar{\Gamma}$. Fig. 5.4 shows an example of the resulting Γ and its threshold $\bar{\Gamma}$ for a typical wheeze signal. The sparseness of the identified frames is then reduced by replacing the frames with interpolated values using cubic spline interpolation [96] to obtain the preprocessed signal $x(n)$. The samples in the identified sparse frames are replaced by the interpolated values on a one-to-one basis without increasing the number of data points in those identified sparse frames. Thereby, the sparseness reduction scheme proposed does not affect the segmentation accuracy by generating $x(n)$ with

the same length as the observed $y(n)$. Subsequently, an increment in $SampEn$ as a result of sparseness reduction has been illustrated in Fig. 5.5.

Figure 5.5: $SampEn$ of $y_{sf}(n)$ in Fig. 5.4 (dotted line) and $SampEn$ of the corresponding input signal $x(n)$ after sparseness reduction (solid line).

5.2.3 Respiratory Rate Estimation

Noise Estimation

Normal RS is defined as a breath sound detected over trachea/chest and is characterized by a wide band noisy sound. On the other hand, adventitious sounds of various types are considered to be additional RS that are superimposed on normal RS [97]. Therefore, the preprocessed RS signal $x(n)$ can be interpreted as adventitious sound being corrupted by an additive noise which is normal RS. A model can thus be proposed as

$$x(n) = e_s(n) + e_N(n) \quad (5.4)$$

where $e_s(n)$ represents the signal (i.e. the adventitious sound) which may or may not be present and $e_N(n)$ is the additive noise (i.e. the underlying normal RS). The

concept of generalized spectral subtraction [98] has been adopted here for the noise estimation in order to obtain correct **RR**. The quasi-stationarity of $e_s(n)$ as well as the uncorrelatedness between $e_s(n)$ and $e_N(n)$ are assumed.

The signal spectrum relation is thus

$$X(k) = E_s(k) + E_N(k) \quad (5.5)$$

with k varying within $[0 \ N - 1]$ as a discrete variable that denotes the frequency bins. The magnitude spectrum $|E_s(k)|$ can be then estimated by

$$|\hat{E}_s(k)| = |X(k)| - |\hat{E}_N(k)| \quad (5.6)$$

with $\hat{E}_N(k)$ being the estimation of $E_N(k)$. Since $E_N(k)$ represents the spectrum of normal RS which is nonstationary, the transfer function $H(k) = \frac{\hat{E}_s(k)}{X(k)}$ is therefore updated recursively in the following way [99]:

$$H(k) = \begin{cases} \left[\frac{|X(k)|^a - \kappa |\hat{E}_N(k)|^a}{|X(k)|^a} \right]^\beta & \text{for } |X(k)|^a - \kappa |\hat{E}_N(k)|^a > 0 \\ \mu & \text{otherwise} \end{cases} \quad (5.7)$$

$$\begin{aligned} \hat{H}(k) &= \gamma_f \hat{H}^{old}(k) + (1 - \gamma_f) H(k) \\ |\hat{E}_N(k)|^a &= \gamma_d |\hat{E}_N^{old}(k)|^a + (1 - \gamma_d) |X(k)|^a. \end{aligned}$$

$\hat{H}^{old}(k)$ and $|\hat{E}_N^{old}(k)|$ refer to the estimations from the previous frames, and the coefficient κ is defined as a function of a signal-to-noise ratio SNR_{tot} where

$$SNR_{tot} = 10 \log_{10} \left[\frac{\sum_{k=0}^{N-1} |X(k)|^2}{\sum_{k=0}^{N-1} |\hat{E}_N(k)|^2} \right]. \quad (5.8)$$

The estimated time domain noise envelope $|\hat{e}_N(n)|$ (underlying normal RS) is lastly obtained using the overlap-add method.

Nonlinear Mapping

The obtained envelope $|\hat{e}_N(n)|$ of the underlying normal RS is nonlinearly mapped by

$$|\hat{e}_N^m(n)| = \begin{cases} |\hat{e}_N(n)| & \text{for } |\hat{e}_N(n)| < \tilde{\mathbf{e}}_N \\ \tilde{\mathbf{e}}_N & \text{for } |\hat{e}_N(n)| > \tilde{\mathbf{e}}_N \end{cases} \quad (5.9)$$

where $\tilde{\mathbf{e}}_N$ is the median value of the $|\hat{e}_N(n)|$ sequence. Nonlinear mapping is applied here to suppress the high variances in $|\hat{e}_N(n)|$. R_t as the total number of inspiration and expiration segments, is then obtained by thresholding $|\hat{e}_N^m(n)|$ with its global mean $\bar{e}_N^m$. Thereby, $\mathbf{RR}$ is estimated as $\mathbf{RR} = \frac{R_t}{T}$ with T representing the signal duration and R_t being half of the number of points crossing the threshold $\bar{e}_N^m$. An illustrative plot for $|\hat{e}_N(n)|$, $|\hat{e}_N^m(n)|$, and the corresponding threshold $\bar{e}_N^m$ is shown in Fig. 5.6.

5.2.4 Sample Entropy

In the proposed GA based segmentation method, sample entropy $SampEn(m, r, N_s)$ as a similarity measure of the preprocessed RS signal $x(n)$, is employed to determine the boundaries of respiratory segments. $SampEn(m, r, N_s)$ is chosen as it does not count for self-matches of the time series and thus ensures the consistency of the measurement and reduces the dependency on the signal length. It is defined in [100] as the negative natural logarithm of the conditional probability that a data set of length N_s , having repeated itself within a tolerance r for m points, will also repeat itself for $m+1$ points, without allowing self-matches. The detailed equations are shown in Appendix C.

$SampEn(m, r, N_s)$ being a measure of regularity, reflects a high degree of self-similarity in time series with low values, and an increasing irregularity by showing high values. Therefore, the dynamics of $x(n)$ can be captured by the $SampEn$ sequence with its increased values for respiratory segments and decreased values for pause segments. Hence, signal conditioning is done on the $SampEn$ sequence by taking the natural

logarithmic scale to reduce its dynamic range and thereby to homogenize its samples density. The discriminations by the evaluation function (in Section 5.2.5) based on heterogeneity measure of *SampEn* between respiratory phases is enhanced. It thus further favors the GA search which is piloted by the function output.

Figure 5.6: (a) Original waveform $y(n)$ of a real recorded signal as mixture of real wheeze and pleural sound (WNP) recording; (b) The preprocessed clean RS signal $x(n)$; (c) $|\hat{e}_N(n)|$, $|\hat{e}_N^m(n)|$, and $\bar{e}_N^m$ of $x(n)$ with $R_t = 6$; (d) Time-frequency plot of the signal in (a).

5.2.5 Multi-population Genetic Algorithm (MPGA)

Genetic algorithms (GAs) are numerical optimization algorithms inspired by both natural selection and natural genetics [101]. GAs operate on a population of strings, that is, a group of potential solutions of a problem. Fitness of each string is calculated in decoded form (solution vector) by applying an evaluation function to measure how good or bad the solutions within the population. At each generation, a new set of solutions are produced by selecting the fittest strings in the problem domain and through the application of the genetic operators such as crossover and mutation. A review for the fundamental operations of a simple GA can be found in [102]. The procedure of a simple GA is described as follows, where the population of candidate solutions at time n is represented by $\mathfrak{R}(n)$:

```

begin
     $n = 0$ ;
    initialize  $\mathfrak{R}(n)$ ;
    while not termination criteria do
        begin
             $n = n + 1$ ;
            select  $\mathfrak{R}(n)$  from  $\mathfrak{R}(n - 1)$ ;
            reproduce pairs in  $\mathfrak{R}(n)$ ;
            evaluate  $\mathfrak{R}(n)$ ;
        end
    end
end

```

Initial Population

In order to detect both start and end locations of each segment, a population of GA is generated with the strings whose length is twice the total number of segments R_t as obtained in Section 5.2.3. A string is a real-valued number sequence representing the locations of the candidate segment boundaries in increasing order. Although the binary-coded GAs are the most commonly used representation, a more natural real-valued representation is used in this scheme to increase the efficiency of the GA. By

using the real-valued strings, there is no need to convert strings to solution vectors to evaluate their fitness.

Evaluation Function

An evaluation function is designed using the heterogeneity measure [103] and *SampEn* in order to obtain accurate boundaries of each segment. It quantifies the optimality of the individual in GA by scoring the candidate segmentations using the homogeneity within the segments and the heterogeneity among different segments based on the calculated *SampEn*. Accurate segmentation is thus facilitated by optimizing this function through the evolution procedure.

In this way, *SampEn* of the original segmenting signal is calculated first to investigate the dynamics. Let H_w be the total within-segment heterogeneity and H_b denotes the total between-segment heterogeneity, a segmentation evaluation function is therefore defined as

$$H = \frac{H_b + 1}{H_b + H_w + 1} \quad (5.10)$$

with total within-segment heterogeneity H_w being defined as

$$H_w = \frac{\sum_{u=1}^U N_u \sigma_u^2}{N}. \quad (5.11)$$

Here, N is the length of the segmented signal, N_u is the length of the u th segment, σ_u^2 is the variance of the *SampEn* of the u th segment and U is the total number of segments in the segmented signal. The between-segment heterogeneity, H_b , is defined as the average Euclidean distance between the mean value of the *SampEn* of any two adjacent segments.

$$H_b = \frac{\sum_{(u,v) \in \text{adjacent}, u \neq v} \|\mu_u - \mu_v\|^2}{ns} \quad (5.12)$$

where ns is the total number of the adjacent segments in the segmented signal, μ_u and

μ_v are the mean values of the *SampEn* for the u th and v th segments. H_b becomes one when the internals of all segmented RS signals are completely homogeneous.

Evolution Procedure

In order to effectively search the solution space, and to take advantage of the parallelism of GAs, the proposed algorithm applies the multiple subpopulations approach provided by [104] for the evolutionary process. By using multiple populations, the quality of the results obtained can be improved compared to GAs with single population. This approach divides the population into subpopulations where each of them can evolve independently using parallel processing technique. It can search in parallel different subspaces of the search space and is thus less likely to be trapped by low-quality subspaces. MPGA is a widely used parallel GA model where multiple subpopulations evolve independently toward different optima. More diverse subpopulations can be maintained by exchanging genetic materials between subpopulations. The premature convergence effect of simple GA can then be mitigated by this approach. To reduce the required computational time, it is implemented through the use of high-level genetic operator functions and exchanging individuals between subpopulations.

Over generations, each subpopulation is evolved as in traditional simple GA using the basic operators: *crossover* and *mutation*. Depending on the migration interval (i.e. the number of generations between successive migrations) as well as the migration rate (i.e. the number of individuals to be migrated from one subpopulation to another), individuals from one subpopulation migrate to another from time to time. The initial population is created using subpopulations containing number of individuals each. At each generation, certain portion of the individuals with higher fitness values within each subpopulation are selected for breeding using a *stochastic universal sampling* function which has minimum spread and zero bias.

The recombination operator is used to produce the new offsprings. By applying *discrete recombination crossover* which is a uniform crossover for real-valued representation, the new offsprings within each subpopulation are produced. Normally, offsprings are mutated after recombination to prevent the population from converging to local minima, and the new possible solutions can be introduced to the population by mutating the offsprings. The mutation rate is set to be $1/nvar$ with $nvar$ being the length of an individual.

When the offsprings produced are less than the size of the original population, the new offsprings have to be reinserted into the population to maintain the size of the original population. Similarly, when not all the offsprings are to be used at each generation, or if the offsprings produced are more than necessary, a reinsertion scheme must be used. This scheme determines which individuals should be replaced by the offsprings produced and which individuals should be inserted into the new population.

In this segmentation method, offsprings are inserted into the appropriate subpopulations depending on *fitness-based reinsertion*. The migration of individuals between subpopulations is performed with certain migration rate. After GA iterates for *maxgen* times, the evolution of this GA stops. The best individual with the maximum fitness value presents the optimized solution for the boundaries of the segments of RS signals.

5.2.6 Parameter Setting

Signal Preprocessing

In any spike detection algorithm, the threshold is optimized to minimize the missing true peaks while keeping a low false detection rate. Since the calculation of *SampEn* is not affected much by the presence of spikes, the scaling factor c in (5.2) is therefore chosen experimentally according to [93]. As the choice of c depends on the type of spikes

being detected, a simple criterion based on the kurtosis which measures the peakedness of the sequence is used here to select the appropriate value of c . The averaged kurtosis value of y_{sf} after spike removal with different values of c has been computed using four RS signals randomly chosen from the experimental dataset. A reference line is found as the averaged kurtosis of the neighboring spike-free segments in y_{sf} with the same length as those spike-contaminated segments. A low kurtosis reflects low peakedness and thereby high signal-to-spike ratio. The proper choice of c is therefore depicted in Fig. 5.7 by having averaged kurtosis below the reference line (solid line). As shown by Fig. 5.7, c value can be selected within the range of [15 55]. $c = 35$ yielding y_{sf} with minimum averaged kurtosis has therefore been chosen here and satisfactory results are obtained.

Figure 5.7: Averaged kurtosis of the spike-contaminated parts of $y_{sf}(n)$ by varying c values (solid line with indents) and the averaged kurtosis of their neighboring spike-free parts of $y_{sf}(n)$ (solid line).

The dimension of the windowed frames for sparseness reduction is chosen to be $N_w = 100$. It is set to be the same as that for the *SampEn* calculation to prevent the generation of extra data points during interpolation, and the segmentation accuracy is thus ensured.

Noise Estimation

In this chapter, FFT with radix-2 algorithm and a 75% overlapped Hann window of length 32 ms have been applied to obtain the spectra. The signal is therefore upsampled to $F_s = 16$ kHz to have the window size of power of 2. As a result, the corresponding spectrum becomes flattened due to the effects of bandwidth extension [105] and the redistribution of spectral power caused by upsampling. Since nonlinear distortion and therefore extra frequency components are produced by the nonlinear preprocessing steps involved (i.e. spike removal and sparseness reduction) [106], the lifted noise floor level of the underlying normal RS (i.e. the noise $e_N(n)$ in (5.4)) can also be suppressed by upsampling.

For the updating process in (5.7), the power exponent $a = 2$ for power spectral subtraction [107] and the noise subtraction factor $\kappa = [1.3 \quad 2]$ value which varies with the estimated SNR_{tot} , are suggested to better cope with a range of input SNRs. The noise gain function $G_N(k) = \left[\frac{\gamma^{\frac{a}{2}}(k)+1}{\gamma^{\frac{a}{2}}(k)+1-\kappa} \right]^{2\beta}$ with $\gamma(k) = |E_S|^2(k)/|E_N|^2(k)$ [99] is adopted for the selection of β . As depicted by Fig. 5.8, at $a = 2$ and $\kappa = 2$, high value of β gives much higher noise gain that implies a more prominent noise envelope. Therefore, $\beta = 1.5$ is chosen here giving distinct noise envelope for accurate **RR** estimation. A small spectral floor $\mu = 0.006$ is chosen to enable an accurate tracking with fairly fast convergence. At the same time, the forgetting factor $\gamma_f = 0.5$ is selected together with $\mu = 0.006$ considering the tracking of fast variations in noise level for nonstationary noise (i.e. underlying normal RS). The initial value of γ_d is chosen to be 0.8 and is updated recursively by spectral discrepancy measurement. Also, the parameter κ based on SNR_{tot} in (5.8) is defined as

$$\kappa = \begin{cases} 2 & \text{for } SNR_{tot} > 8 \\ 0.175 \times SNR_{tot} + 0.6 & \text{for } 4 \leq SNR_{tot} \leq 8 \\ 1.3 & \text{for } SNR_{tot} < 4. \end{cases} \quad (5.13)$$

Figure 5.8: Performance of the noise gain function $G_N(k)$ with varying β at $a = 2$ and $\kappa = 2$.

SampEn Calculation

In order to achieve feasible computational time and to make the proposed algorithm tractable, $SampEn(m, r, N_s)$ of each input signal is calculated using the parameter values chosen from the suggested ranges in [108] as $N_s = 100$, $m = 1$, and $r = 0.2\sigma$ with σ being the standard deviation of the signal.

We calculate the *SampEn* for each window of 9 ms length with $F_s = 11025$ Hz. $N_s = 100$ is chosen due to the fast dynamic nature of the RS signal used, and more importantly, it ensures the feasibility for implementation as well as produces higher phase boundaries localization accuracy. In fact, the F_s increases as the sampling interval decreases while the samples become increasingly correlated. Therefore, although reducing m is required for increasing F_s to have the same *SampEn*, the corresponding *SampEn* value becomes insensitive with respect to m when F_s is very high. This

has been illustrated in Fig. 5.9(b) by having the same $SampEn$ with varying m . Thus $m = 1$ is arbitrarily chosen here for the high $F_s = 11.025$ kHz. On the other hand, although r increases with smaller N_s , as indicated by the experiments conducted in [109], the $SampEn$ and its sensitivity to the choice of noise filtering parameter r decreases with increasing F_s . Fig. 5.9(c) shows the changes of $SampEn$ with varying r . Since the optimization process using MPGA does not depend on the amplitude of $SampEn$, all three choices of $r = 0.1\sigma, 0.2\sigma, 0.3\sigma$ produce accurate segmentation. Because the effect of r is not significant for high F_s , $r = 0.2\sigma$ is arbitrarily chosen.

Figure 5.9: (a) Waveform of a complete respiratory cycle (inspiration-pause-expiration-pause) from a preprocessed stridor signal; (b) Effects of m on $SampEn$ with $r = 0.2\sigma$ for the signal in (a); (c) Effects of r on $SampEn$ with $m = 1$.

Genetic Algorithm

The parameters of the genetic algorithm are selected as evolved in traditional simple GA using the basic operators of *crossover*, and *mutation* [110] and the parameter values we used in this chapter, are taken from [111]. The number of subpopulation

should be large enough to ensure the speed of the GA and the number of individuals should be chosen based on the overall size of the population (i.e. signal length N_1). As the minimum length of the RS signals used is 6 s (which corresponds to $N_1 \geq 66150$ samples at $F_s = 11.025$ kHz), the initial population is therefore created using 8 subpopulations containing 60 individuals each. Top 90% of the individuals in terms of high fitness values are selected within each subpopulation for breeding. Offsprings are inserted into the appropriate subpopulations depending on *fitness-based reinsertion* with a rate of 0.9, and migration of individuals between subpopulations is performed at every 20 generations with a migration rate of 0.2. The evolution stops after the algorithm iterates for *maxgen* times (here *maxgen*=80). The best individual with the maximum fitness value presents the optimized solution for the boundaries of the segments of the segmented signal. Fig. 5.10 indicates how the mean segmentation errors calculated using (3.15) evolve through the generations. It further verifies the choice of *maxgen* = 80 since all the curves become constant after *maxgen* = 70.

5.3 Results and Discussion

5.3.1 Respiratory Rate Estimation

The definition of adventitious sounds, being additional respiratory sounds superimposed on normal breath sound [36], motivates the use of signal plus noise model in (5.4) for the original RS recordings $y(n)$. Since the bandpass filtering is a linear process, and the nonlinear steps (spike removal and sparseness reduction) involved to obtain the preprocessed RS signal $x(n)$ from $y(n)$ are only applied to very short intervals of $y(n)$, the same additive model is retained for $x(n)$. If the preprocessing steps are to be replaced, the alternative steps are required to be linear or at least linear for

Figure 5.10: Mean segmentation error of the MPGA with different *maxgen* for TBS recordings of different types.

most parts of the signal in order to ensure the validity of the model in (5.4).

The proposed method uses recursive noise estimation to track accurately the underlying normal RS which is nonstationary. In addition, the signal power leakage corresponding to discontinuous adventitious sounds (DAS) can be minimized with the aid of nonlinear mapping. Therefore, the introduction of the noise estimation step followed by the nonlinear mapping produces smooth envelope with least distortion due to adventitious sounds. This ensures the ease of hard thresholding and therefore the correct estimation of **RR**.

The performance of the proposed method in terms of **RR** estimation accuracy (%) has been compared with that of based on signal autocorrelation in [41] for different types of respiratory sounds. The **RR** estimation accuracy is calculated as $Accuracy = \left(\frac{1}{G} \sum_{g=1}^G \frac{\text{Estimated } \mathbf{RR} \text{ for } g\text{th signal}}{\text{Actual } \mathbf{RR} \text{ for } g\text{th signal}} \right) \times 100\%$ with G being the total number of signals of the same type, and the actual **RR** is identified by highly experienced doctors.

Table 5.1: RR estimation accuracies (%) of different methods on different types of RS

Signal Type	RR Estimation Accuracy (%)	
	Method in [41]	The Proposed Method
Normal RS	100	100
Stridor	90	100
Wheeze	93.3	100
WNP	NA	100

The corresponding results obtained are listed in Table 5.1 for all available data including clean LS signals drawn from [37][64] and real TBS recordings of different types. The types of RS used include 15 normal RS (10 real TBS recordings), 10 pure wheezes (7 real recordings), 8 pure stridors (5 real recordings), and 15 mixture of wheeze and pleural sounds (WNP) (15 real recordings). As shown in Table 5.1, the proposed method achieves 100% accuracy in **RR** estimation for all types of RS which outperforms the autocorrelation based method in [41].

The adventitious sounds can be divided into continuous and discontinuous according to Section 2.3. Continuous adventitious sound (CAS) is quasi-stationary, while DAS is explosive and transient in nature [36]. Fig. 5.6(d) shows the time-frequency representation of the recorded signal in Fig. 5.6(a) obtained by using short-time Fourier transform. The dark vertical lines extended towards the higher frequency range indicate the presence of pleural sound which is of discontinuous adventitious type. The rapid increment in the estimated noise envelope $|\hat{e}_N(n)|$ as shown in Fig. 5.6(c), has therefore occurred at the presence of pleural sounds. Since relatively long time window is used for the tracking of quasi-stationary signal in order to avoid leakage, it is not suitable for the DAS which has transient nature. This implies that the signal power leakage is induced by the nonstationarity of the DAS which defies the quasi-stationarity assumption for $e_s(n)$ in (5.4). Furthermore, since the underlying normal RS is nonstationary,

continuous noise tracking is chosen in Section 5.2.3 to favor the noise estimation for the nonstationary $\hat{e}_N(n)$. It is faster in adaptation time to the changing level of $\hat{e}_N(n)$ [112] and is therefore able to track the respiratory phase changes. However, at the same time, it is also susceptible to the undesired signal power leakage.

Nonlinear mapping has then been incorporated to compensate the deterioration in tracking $\hat{e}_N(n)$ due to signal power leakage caused by the presence of DAS. By mapping $|\hat{e}_N(n)|$ to $|\hat{e}_N^m(n)|$ using (5.9), the low amplitude components can be preserved while the dynamic range of $\hat{e}_N(n)$ can be reduced. In this way, a hard threshold of $\bar{e}_N^m$ can therefore be applied for accurate estimation of **RR**.

5.3.2 Respiratory Phase Segmentation

Table 5.2 presents the performances of the proposed phase segmentation method on the same set of experimental data as used in Section 5.3.1. The overall segmentation accuracy is calculated as the mean and standard deviation ($\mu \pm \sigma$) of estimation error ϵ in ms between the actual respiratory phase locations identified by highly experienced doctors and the estimated phase locations obtained by the proposed method. The ϵ for each RS signal is defined as

$$\left\{ \begin{array}{l} \epsilon_{start} = \frac{1}{J_S} \sum_{j=1}^{J_S} | \hat{P}_j^S - P_j^S | \\ \epsilon_{end} = \frac{1}{J_E} \sum_{j=1}^{J_E} | \hat{P}_j^E - P_j^E | \\ \epsilon = \frac{\epsilon_{start} + \epsilon_{end}}{2}. \end{array} \right. \quad (5.14)$$

where J_S and J_E represent the total number of starting and end points of the respiratory phase segments in every RS signal. Also, $\hat{P}_j^S$ and $\hat{P}_j^E$ are the starting and end positions of the j^{th} respiratory phase segment as obtained in $\mathbf{A}_r$ while P_j^S and P_j^E are those identified by the doctors. Since 4 doctors have been involved, the agreement among doctors are obtained by averaging their judgements on phase boundary locations as described in Section 5.2.1. For performance evaluation on different types of RS signals,

Table 5.2: The estimation errors $(\mu \pm \sigma)$ (ms) of the proposed method on clean LS and real recorded TBS signals of different types

Signal Type	Segmentation Error $(\mu \pm \sigma)$(ms)		
	Clean LS	Real TBS	Overall Accuracy
Normal RS	11.2 ± 5.3	12.8 ± 4.5	12 ± 5
Stridor	20.4 ± 10.4	18.5 ± 8.6	19 ± 9
Wheeze	18.1 ± 8.6	17.9 ± 6.4	18 ± 7
WNP	20.8 ± 10.2	21.5 ± 8.1	21 ± 9

the accuracy is calculated for each RS signal using (5.14) and then averaged over the same type of signals.

According to the values shown in the second and third columns of Table 5.2, only a slight performance degradation (maximum 1.9 ± 1.8 ms) is observed for the proposed method on real TBS recordings compared to that on clean LS from the standard databases [37][64]. This verifies the noise interference removal ability of the preprocessing steps proposed in Section 5.2.2. Since the number of sparse frames in the clean LS signals are much less than that in real recorded TBS signals, the slight degradation is caused by the interpolation used in Section 5.2.2. The sparseness reduction can be further improved by replacing cubic spline interpolation with other higher order polynomial interpolation scheme.

Furthermore, the proposed method gives an overall accuracy as high as 12 ± 5 ms for the phase segmentation of normal RS, 18 ± 7 ms for CAS, and 21 ± 9 ms for mixture of continuous and discontinuous adventitious sounds as indicated in Table 5.2. Due to the absence of adventitious sounds which results in homogeneous signal nature, normal RS signals have the highest overall segmentation accuracy. On the other hand, the performance of this method on adventitious sound is better for wheeze and stridor than that for pleural sounds with different signal nature. By definition in [97], both

Table 5.3: The estimation errors $(\mu \pm \sigma)$ (ms) for different respiratory phase segmentation methods on real recorded TBS signals of different types

Signal Type	Segmentation Error $(\mu \pm \sigma)$ (ms)		
	Method in [43] without GA	Method in [43] with GA	The Proposed Method
Normal TBS	30.2 ± 15.1	20.4 ± 9.5	12.8 ± 4.5
Stridor	Failed	30.6 ± 18.6	18.5 ± 8.6
Wheeze	41.5 ± 25.3	25.8 ± 12.2	17.9 ± 6.4
WNP	Failed	38.6 ± 16.1	21.5 ± 8.1

stridor and wheeze are classified into CAS which are characterized by the periodic waveforms with dominant frequency over 100 Hz, while pleural sounds being DAS are characterized by transient waveforms. Since a higher *SampEn* is obtained with increasing irregularity, the dynamic changes of *SampEn* along time are much larger for DAS than that for CAS. This further reduces the homogeneity of *SampEn* density and thus poses difficulties for its signal conditioning. Hence, the segment boundary localization piloted by evaluation function using heterogeneity of *SampEn* becomes less accurate in the presence of DAS.

When comparing the performance of the presented method with that of the fractal dimension based method in [43], the proposed method achieves much better performance for all types of RS signals, as indicated in Table 5.3. Also, due to the fixed length searching window used, the method in [43] fails for stridor and the mixture of wheeze and pleural sound (as shown in the second column of Table 5.3). This is due to the variability in the **RR** for different subjects. Therefore, the proposed multi-population GA in Section 5.2.5 has been applied on the extracted feature of variance fractal dimension to improve the segmentation accuracy. This significant performance improvement by using GA can be seen in the third column of Table 5.3.

Unlike conventional average power methods, the variance fractal dimension does

not compromise between time and frequency resolutions . This is the reason for the fractal dimension method being robust for different types of RS. It depends solely on its time resolution which varies with the size of window used. However, this dependency on window size makes the method sensitive to outliers and noises. Unlike *SampEn*, fractal dimension does not have any noise filtering. The resulting spiky envelope of fractal dimension deteriorates the effectiveness of GA evaluation function and thereby downgrades the segmentation performance. Therefore, comparing the results of GA using fractal dimension with that using *SampEn*, the latter outperforms the former.

5.4 Chapter Conclusion

An automatic and robust method is proposed in this chapter to locate respective respiratory phases for different types of RS signals. The incorporation of a robust respiratory rate estimation scheme fully automates the proposed method. A 100% accuracy has been achieved in **RR** measurements for all types of RS signals. Heterogeneity measure based on *SampEn* is calculated for the design of GA evaluation function. By using this new feature, the presented method is able to eliminate the dependency of the phase segmentation accuracy on the signal length and amplitude. The introduction of MPGA has removed the requirements of threshold and training dataset for segment boundary localization. This makes the algorithm automatic even in the absence of any *a priori* information of the input signal types.

To our best knowledge, GA based segmentation method for RS signals has not been published earlier. Comparison has been made between using *SampEn* and variance fractal dimension as feature vectors. The comparison results reveal the potential to further improve the method by choosing even more appropriate features.

Chapter 6

Annotating Index for Respiratory Phase Identification

6.1 Introduction

Respiratory phase information is essential for the automation of pathological respiratory signal analysis. In adventitious sounds quantification [84] and the studies of flow in heart, it is crucial to correctly differentiate the individual respiratory phases as inspiration/expiration. Therefore, it is of great importance to extract appropriate features from each segmented respiratory phases for the identification of the respective inspirations and expirations.

Respiratory phase identification methods based on spectral and temporal analysis of the transformed RS have been recently suggested in [8][9][42][113]. As presented in [8] and [42], normal lung sounds (LSs) which have much wider spectra for inspirations than those for expirations, has been used for phase identification. More specifically, [8] deploys the quasi-periodicity of the short-term signal energy and uses a bandpass filter

for phase identification. Also, [44] adopts triplet Markov chain in the wavelet packet domain to improve the identification accuracy by removing its dependency on signal amplitude. However, *priors* on the respiratory cycle structure for normal LS have been exploited for chain adaptation which make this method ineffective for pathological RS. The accuracies of these methods depend solely on the distinct amplitude and spectral differences between inspiration and expiration of LS.

On the other hand, compared to LS, the TBS signal being a more appropriate source for the analysis of those RS analysis related to upper airway constriction, has less prominent amplitude and spectral differences between respective respiratory phases. All the above mentioned phase identification methods except [9] therefore failed. In [9], the respiratory phases are identified based on the summation of the spectral components over a certain frequency range as well as the spectral variations of the normal TBS signals. However, it requires more than one microphone to capture the ambient noise and is therefore sensitive to disturbance.

Since the presence of adventitious sounds further affects the amplitude as well as spectral power distribution of the RS signals to a large extent, the aim of this chapter is to propose a reliable method to identify inspiration and expiration from single-channel pathological RS recordings including TBS. The presented frequency domain mixing index (MI) based on inter-segment similarity measurements, is independent of the amplitude variations between the two phases and able to perform accurate annotation without any *a priori* information of the signal types. Analysis is performed on the signal characteristics of RS as well as the adaptive parameters selection in order to realize the effectiveness of this new index.

6.2 The Proposed Annotating Index

6.2.1 Mixing Index (MI)

This section introduces a frequency domain annotating index MI based on inter-segment similarity measurements for respiratory phase identification. The basic idea behind MI is to compare the signals from the consecutive phase segments in the time-frequency (TF) plane to derive an time-averaged mixing gain associated with each frequency bin. Finally, the MI value for each pair of segments is then obtained by taking the mean over the selected frequency bins. MI therefore has value between $[-1, 1]$ centered at 0 for each frequency bin.

Although RS can be segmented into successive inspiratory phase, end-inspiratory pause, expiratory phase, and end-expiratory pause, only the segmented respiratory phases (inspiratory/expiratory) are involved in phase identification here. The consecutive pair of respiratory phase segments (which consists of $y_1(n)$ as an inspiratory/expiratory phase segment and $y_2(n)$ as the following expiratory/inspiratory segment with n being the time index) are used as inputs for the proposed MI. In this chapter, segmentation method proposed in Chapter 5 has been adopted for extracting the respiratory segments pair due to its high accuracy, although other suitable methods can also be used.

A rectangular window of length N_w is used first to obtain the windowed frames $y_1^w(n)$ and $y_2^w(n)$. The choice of window is to ensure uniform energy distribution within each windowed frame for the accurate measurement of MI based on averaging. Fast Fourier transform (FFT) is then applied to every windowed frame to obtain $Y_1^w(k)$ and $Y_2^w(k)$. $1 \leq w \leq W$ and W indicates the total number of frames of N_w samples. Similarity

between $Y_1^w(k)$ and $Y_2^w(k)$ can therefore be measured as

$$\lambda(k) = \frac{2}{W} \sum_{w=1}^W \frac{|Y_1^w(k)Y_2^w(k)^*|}{|Y_1^w(k)|^2 + |Y_2^w(k)|^2} \quad (6.1)$$

with ‘ $\star$ ’ indicating complex conjugate and k being the frequency bin index that varies between $[0 \ N_w-1]$. Also, we have defined another set of partial similarities based on cross-correlation function of the phase aligned signals to avoid ambiguity:

$$\begin{cases} \lambda_1^w(k) = \frac{|Y_1^w(k)Y_2^w(k)^*|}{|Y_1^w(k)|^2} \\ \lambda_2^w(k) = \frac{|Y_2^w(k)Y_1^w(k)^*|}{|Y_2^w(k)|^2}. \end{cases} \quad (6.2)$$

The annotating index MI can be then obtained as

$$\text{MI}(k) = [1 - \lambda(k)] \times \nu^m(k) \quad (6.3)$$

where $\nu^m(k)$ is a mapped version of the difference $\nu(k) = \frac{1}{W} \sum_{w=1}^W \lambda_1^w(k) - \lambda_2^w(k)$ between the partial similarities

$$\nu^m(k) = \begin{cases} +1 & \text{for } \nu(k) > 0 \\ 0 & \text{for } \nu(k) = 0 \\ -1 & \text{for } \nu(k) < 0. \end{cases} \quad (6.4)$$

Therefore, the overall MI for each pair of consecutive phase segments is defined as

$$\text{MI} = \frac{1}{K2 - K1} \sum_{k=K1}^{K2} \text{MI}(k) \quad (6.5)$$

with $K1$ and $K2$ being the frequency bin indices. As an inherent property of normal RS, the inspiration segments have more frequency components with higher magnitude compared to the expiration segments. Therefore, the partial similarity $\lambda_1^w(k) < \lambda_2^w(k)$ and thus $\text{MI} < 0$ for $y_1^w(n)$ being annotated as inspiration and $y_2^w(n)$ being annotated as expiration. The annotation is vice versa for $\text{MI} > 0$.

The proposed annotating index is useful since MI being a relative index, is not affected by the change of spectral content or amplitude dynamics. Thereby, any changes

of amplitude or even complete modifications of spectra to both respiratory phases (as long as the changes are proportional to the similarity coefficients), do not affect the MI values. In this way, MI is found reliable even at the presence of adventitious sounds which affects the spectra of the input RS. Since the reliability of the MI depends on the appropriate choice of frequency bins ($K1$ and $K2$) and window length (N_w), the corresponding parameter values are adaptively chosen based on the following criteria described in Section 6.2.2 and 6.2.3.

6.2.2 The Choice of Frequency Bins

The presence of HS is mainly dominated within the frequency band of 20-300 Hz for $F_s = 11025$ Hz [36], and would be even higher when being captured over suprasternal notch. The analysis of RS signals is therefore confined above 500 Hz to minimize the effects of HS and the corresponding $K1 = 500 \cdot N_w / F_s$ has been selected for MI calculation.

At the same time, $K2$ can be selected as $K2 = N_w / 4$ (2756 Hz) or $K2 = N_w / 2$ (5512 Hz) based on different RS signal characteristics. Since the MI measurement is computed as an average value across all selected frequency bins $k = [K1 \ K2]$, an unbiased and reliable MI depends on the appropriate choice of $K2$. The choice of $K2$ should be as large as possible (i.e. $K2 = N_w / 2$ which is the Nyquist frequency) to minimize spectral power leakage, but it should also be small enough (i.e. $K2 = N_w / 4$) to avoid ambiguity for narrowband signals. This ambiguity arises when there is close proximity between the spectra of $y_1^w(n)$ and $y_2^w(n)$ for $k = [N_w / 4 \ N_w / 2]$ due to very low power at high frequencies.

Hence, a spectral power ratio $\gamma = \{\gamma_1, \gamma_2\}$ is proposed here for the choice of $K2$ based on PSD. Since MI is a frequency domain index, and the power spectrum of a

signal gives the distribution of the signal power at various frequencies, γ is defined as

$$\gamma = \frac{\sum_{k=N_w/4}^{N_w/2} \hat{P}(k)}{\sum_{k=K_1}^{N_w/2} \hat{P}(k)} \quad (6.6)$$

with $\hat{P} = \{\hat{P}_{y_1}, \hat{P}_{y_2}\}$ representing the PSDs of $y_1^w(n)$ and $y_2^w(n)$. The PSD is estimated by Welch method [114] with N_w samples Hamming window, and K_2 is selected based on the values of γ such that

$$K_2 = \begin{cases} N_w/4, & \text{for } \gamma_1 < 1\% \text{ and } \gamma_2 < 1\% \\ N_w/2, & \text{otherwise} \end{cases} \quad (6.7)$$

where $y_1^w(n)$ and $y_2^w(n)$ are obtained from $y_1(n)$ and $y_2(n)$ using a non-overlapping rectangular window of $N_w = 2048$ (choice of N_w is elaborated in Section 6.2.3). The threshold of $\gamma_{1,2} < 1\%$ can be retained for PSD estimation using Welch method or any other more effective nonparametric PSD estimator.

6.2.3 The Choice of Window Length

The choice of window length N_w is required to be large enough to maintain high frequency resolution for less spectral power leakage. At the same time, it should be small enough to have a relatively smooth spectrum which reduces the inter-frame PSD variance in the case of adventitious sounds. Hence, the window length is adaptively chosen among the three values of $\mathcal{N} = \{512, 1024, 2048\}$. As the shortest respiratory phase segment observed is 464 ms (5120 samples), $N_w = 2048$ is chosen to be the largest allowable window size to ensure that at least 2 frames ($W \geq 2$) are obtained for the calculation of MI. On the other hand, $N_w < 512$ produces over-smoothed spectra with reduced variance but drastically increased bias.

After the proper choice of frequency bins using a fixed $N_w = 2048$, an appropriate

value of N_w is chosen adaptively based on the following condition:

$$N_w = \arg \min_{N_w \in \mathcal{N}} \left| \frac{P\hat{S}R_1 - P\hat{S}R_2}{P\hat{S}R_1 + P\hat{S}R_2} \right| \quad (6.8)$$

where

$$P\hat{S}R_1 = \sum_{k=K1}^{K2} \sum_{w=1}^W \frac{|\hat{P}_{y1}(k) - [\frac{1}{W} \sum_{w=1}^W \hat{P}_{y1}(k)]|^2}{W(K2 - K1)}$$

$$P\hat{S}R_2 = \sum_{k=K1}^{K2} \sum_{w=1}^W \frac{|\hat{P}_{y2}(k) - [\frac{1}{W} \sum_{w=1}^W \hat{P}_{y2}(k)]|^2}{W(K2 - K1)}.$$

Since the presence of adventitious sounds imposes large inter-frame variance for inappropriate N_w , the choice of N_w is to reduce the difference of inter-frame variance between $y_1(n)$ and $y_2(n)$. As MI is calculated in frequency domain, this difference is measured in terms of the averaged distance from the PSD of each windowed frames to their averaged PSDs of all frames.

6.3 Analysis and Discussion

6.3.1 Analysis on the Choice of $K2$

Table 6.1 summarizes the choices of frequency bins $K2$ as obtained by using (6.6)-(6.7) for various types of RS signals. Referring the selection of $K2$ based on the signal characteristics listed in the last column of Table 6.1, $K2 = N_w/2$ is chosen for the transient signals with broad spectra extending beyond 1500 Hz as well as for the pseudo-periodic signals with rich harmonics. On the other hand, $K2 = N_w/4$ is selected for the pseudo-periodic signals without or with a few harmonics as well as for the nonstationary narrowband signals. Therefore, the choice of $K2$ depends solely on the spectral support rather than periodicity or stationarity of the signal.

The spread of the spectrum can be quantified by computing spectral support. Narrow spectrum with most of the energy concentrating at lower frequencies gives a small

spectral support, while the support is large for a flat spectrum with more energy spread towards the high frequencies. Therefore, a large $K2$ should be chosen for signals with large spectral support to avoid power leakage and to include more frequency bins for an unbiased MI calculation. $K2 = N_w/2$ is thus selected for the mixtures of wheeze and pleural sound (WNP) which have broad spectra extending beyond 2500 Hz. As the richness in harmonics broadens the spectra, $K2 = N_w/2$ is also chosen for the snore signals as well as some wheeze and stridor signals which are supposed to have narrow spectra. On the other hand, since the exclusion of the spectrum between $[N_w/4 \ N_w/2]$ reduces the ambiguity during MI calculation, $K2 = N_w/4$ is chosen for signals with small spectral support.

Table 6.1: The resulting choices of $K2$ and N_w for different types of RS signals

Type	N_w	$K2$	Signal Characteristic
Normal RS	1024	$N_w/4$	nonstationary, broad spectrum of noise, spectrum < 1500 Hz
Wheeze	1024	$N_w/4$	pseudo-periodic, dominant frequency > 100Hz,
	512	$N_w/2$	possible harmonics, duration $\geq$ 100ms
Stridor	2048	$N_w/4$	pseudo-periodic, lower dominant frequency
	1024	$N_w/2$	than wheeze, possible harmonics
WNP	512	$N_w/2$	pleural sound is discontinuous, transient, duration > 10ms, broad spectrum > 2500Hz
Snore	512	$N_w/2$	pseudo-periodic, presence of harmonics
	1024	$N_w/2$	

Fig. 6.1 displays the scatter plots of the estimated signal roots for an inspiratory normal RS segment $y_1(n)$ and its consecutive expiratory wheeze segment $y_2(n)$ from the same signal. Here, the signal template used for root estimations in Figs. 6.1(a) and 6.1(c) is calculated by averaging the windowed frames of $y_1(n)$ or $y_2(n)$ with $N_w = 2048$. Linear prediction filter coefficients of order 16 have been calculated based

on the constructed template and the corresponding roots are then estimated from the first 10 coefficients. The distributions of the signal roots in the first-quadrant ($[0 \pi/2]$) and the second-quadrant ($[\pi/2 \pi]$) of the scatter plot verify the proper choice of K_2 by referring the spectral support. Small spectral support is reflected in Fig. 6.1(a) by the absence of significant roots (i.e. closer to unit circle) in the second-quadrant, whereas large spectral support is shown in Fig. 6.1(c) by having more significant roots in the second-quadrant.

Figure 6.1: Scatter plots of the estimated signal roots for an inspiratory normal RS segment $y_1(n)$ of small spectral support with (a) $N_w = 2048$ and (b) $N_w = 1024$; scatter plots for its consecutive expiratory wheeze segment $y_2(n)$ of large spectral support with (c) $N_w = 2048$ and (d) $N_w = 1024$.

6.3.2 Analysis on the Choice of N_w

As listed in Table 6.1, the window length N_w chosen according to (6.8) depends on both the choice of frequency bins and the signal characteristics. Smaller N_w is chosen for the signal with fast variations and therefore prominent high frequency components

(due to the richness in harmonics) or broad spectrum. This can be observed from Table 6.1 that smaller N_w is necessary when $K2 = N_w/2$ is selected for the snore, wheeze, WNP, and stridor signals. On the other hand, larger window length is chosen for signals with smaller spectral support (i.e. those with $K2 = N_w/4$). As stridor signal has lower dominating frequencies and longer duration than wheeze, the largest window length $N_w = 2048$ is allowed in the case of no or less harmonics. Moreover, $N_w = 1024$ samples is considered rather than $N_w = 512$ for snore, wheeze, and stridor due to their pseudo-periodic nature.

The effect of N_w on balancing between spectral support and frequency resolution can also be confirmed by the magnitude difference between the signal roots (as estimated in Section 6.3.1) of $y_1(n)$ and $y_2(n)$. An argument based on root magnitude difference is proposed below to verify the choice of N_w .

$$\arg \min_{N_w \in \mathcal{N}} (|\rho_{N_w, y_1} - \rho_{N_w, y_2}| / |\rho_{N_w, y_1} + \rho_{N_w, y_2}|) \quad (6.9)$$

where ρ_{N_w, y_1} represents the summation of the roots magnitudes of $y_1(n)$ as obtained with frame size N_w . Since significant roots are always present in the first quadrant of the scatter plot, a small value of ρ corresponds to the high-frequency components that are less significant.

As small N_w reduces frequency resolution, it is able to redistribute the spectral power towards higher frequencies. Hence, normal RS in Fig. 6.1(a), as broad spectrum noise below 1500 Hz, has smaller ρ than that in Fig. 6.1(b) with a smaller N_w . At the same time, due to the presence of rich harmonics and thus large spectral support, the effect of N_w on wheeze segment is less significant. Thereby, similarly large ρ is found for wheeze segment with reducing N_w as obtained from Figs. 6.1(c) and (d). In order to minimize the spectral difference between the signals in Figs. 6.1(a) and (c), a small N_w should be chosen according to (6.8). Since the choice of N_w redistributes the

spectral power, $N_w = 1024$ increases ρ_{N_w, y_1} and thus fulfills the argument in (6.9) as reflected in Figs. 6.1(b) and (d). It can be noted that as the values of N_w chosen here refer to $F_s = 11025$ Hz, they should be scaled by a ratio of $F_s/11025$ when a new F_s is used.

6.4 Results and Comparison

6.4.1 Experimental Dataset

In this study, the real recording dataset consists of TBS from 7 healthy subjects and 14 subjects with different degrees of airway obstructions (8 males/13 females, 15 ± 9 years old). The characteristics due to sex, age, weight were not taken into consideration. Also, preclassified and preprocessed normal/wheezing LS recordings were drawn from the databases [37][64]. A total of 60 normal TBS inspiration/expiration (IE) pairs, 40 wheezing IE pairs, 26 stridor IE pairs, 15 snoring IE pairs, and 25 IE pairs from mixture of polyphonic wheeze and pleural sounds (WNP) recordings were available. Furthermore, since normal LS has distinct amplitude and spectral differences between inspiration and expiration that favors the respiratory phase identification, 20 normal LS IE pairs have also been included. The difference between LS and TBS in terms of phase identification results using the proposed and existing features has then been explored.

6.4.2 Comparison With Time-Frequency Domain Features

In the TF domain, the variation in frequency spectra for different periods of the breathing cycle is conventionally used. The identification of the phases can be obtained by comparing the difference in frequency content within a specific frequency range. Power

spectra of the windowed frames are calculated by taking Fourier transform. Spectral features can be extracted as peak frequency component [9] and average power [42]. However, this transform suffers from the tradeoff between time and frequency resolution according to Gabor-Heisenberg uncertainty principle. If high time resolution is obtained to locate spectral peaks or to calculate the average power for accurate phase identification, the corresponding low frequency resolution would result in over-smoothed spectra without significant peaks.

Since phase identification in [42] depends on the higher power for inspiration than expiration in normal LS, it works well only for LS signals with distinct power differences between the respiratory phases. In contrary, phase identification in [9] relies on signal characteristics of the normal TBS such that the summation of spectral components within certain frequency range is higher at the beginning of expiration than that of inspiration. As the presence of adventitious sound alters the signal nature, this identification criteria no longer sustain. Therefore, the proposed method outperforms the method in [9] as shown in Table 6.2, especially in the presence of inspiratory adventitious sounds such as stridor and snore.

Table 6.2: Performance comparison between the proposed MI and the method in [9] on different types of RS signals

Signal Type	Accuracy (%)	
	The proposed MI	Method in [9]
Normal LS	100	50.4
Normal TBS	100	86.8
Stridor	100	20.2
Wheeze	96.8	95.2
WNP	95.5	74.4
Snore	100	40.3

6.4.3 Comparison with Wavelet Domain Feature

Unlike Fourier transform for extracting the spectral information of the signal, the main objective of wavelet transform is to classify the signal information at different scales in a scale-space domain. Wavelet packet transform applies the transform steps to both the lowpass and the highpass outputs at each dyadic scale, while reserving the low- and high-frequency parts by downsampling. More redundant local TF information can be obtained without considering Gabor-Heisenberg uncertainty principle. Therefore, compared to the TF domain features, the features extracted in the wavelet domain are more suitable for nonstationary signals like TBS.

However, as the identification in wavelet domain still depends on the distinct power difference between inspiration and expiration, [44] that based on statistical modeling in wavelet domain only improves the identification accuracy for normal LS. As the probability distribution of the observed process is assumed to follow χ^2 distribution with predefined parameters for each respiratory phase, this method works well only for normal LS where the wavelet packet coefficients within each respiratory phase can be approximated to an independent normal distribution following the assumption. Since the probability distribution for respiratory phases can be easily altered by replacing LS with TBS or at the presence of adventitious sounds, the Markov chain driven by these parameters failed for all types of TBS signals. Illustrative results are shown in Figs. 6.2 and 6.3 where the identification performance of the proposed MI and the method in [44] on preprocessed normal LS and a typical real wheeze recording are compared. It can be seen that the method in [44] works for normal LS signals, while it identifies inspiration and expiration oppositely for the real recorded wheeze signals.

Figure 6.2: (a) Original waveform of a normal LS signal extracted from [37]; (b) The wavelet domain feature (dash-dot) displayed with the phase identification results by [44] (solid line) and the proposed method (dotted line) with 1, 2, 3 correspond to transition, expiration and inspiration.

Figure 6.3: (a) Original waveform of a real recorded wheeze signal; (b) The wavelet domain feature displayed with the phase identification results by [44] (solid line) and the proposed method (dotted line) with 1, 2, 3 correspond to pause, expiration and inspiration.

6.5 Chapter Conclusion

In this chapter, we present a reliable annotating index MI to identify respiratory phases for pathological respiratory sounds. The proposed MI is insensitive to the amplitude variations which assures accurate results without any *a priori* information of the input signal types. Analysis on the choice of parameters has been performed by introducing new criteria based on signal characteristics. Furthermore, the spectral support and spectral content of different RS signals are discussed by investigating the distribution of the estimated signal roots. The performances of the proposed MI on different types of real RS recordings are compared with the existing methods to show its effectiveness.

Part III

Time-Frequency Analysis and Classification of Respiratory Sounds

Changes in the acoustic properties of the pulmonary system reflect its corresponding anatomical or physiological changes and are strongly associated with breath sounds [115]. Adventitious sounds, being the abnormal sounds heard during the respiratory cycle, can therefore be used to represent various pathological pulmonary processes. Since the characteristics of adventitious sounds, such as the occurrence, timing, and pitch, differ according to various pathological conditions, they are thus further classified into continuous (e.g. stridors and wheezes) and discontinuous (e.g. crackles) based on their characteristics, as discussed in Section 2.3 of Chapter 2.

The close connection of different types of RS with the aforementioned pathologies implies the need for efficient discrimination of RS, in order for the physician to better understand the acoustical changes in the pulmonary system due to the associated pathology. Various techniques have been reported in the literature addressing the problem of RS analysis and classification. Conventionally, time-expanded waveform analysis (TEWA) using time domain features (i.e. duration, zero crossing counts, and deflection width) has been used for the discrimination of discontinuous adventitious sounds (DAS) [116][117]. In contrast, spectral analysis that uses peak, median, and maximum frequencies of the signal power spectra is the most straightforward approach proposed for the detection of continuous adventitious sounds (CAS) [14][51]. Furthermore, criteria or rules concerning the amplitude, duration and pitch range of CAS have been combined with the spectra to further improve the reliability [118][119][120]. However, the above mentioned techniques show a high correlation of the number of adventitious sounds detected with sound signal amplitude, and are therefore strongly affected by the sound attenuation. Among the significant scientific effort put into the discriminative analysis of RS, only a limited number of works have investigated the fractality and nonlinearity of different types of RS signals [121][12][87].

Generally, the second-order statistics fail to reveal any characteristics of the nonlin-

ear interaction of the distinct harmonics, as they are reflected in their phase relation. Higher-order statistics, which preserve the phase information of the signal, can thus be used to detect the nonlinearity and deviation from Gaussianity of the signals, and therefore enable the discriminative analysis of RS. Part III of the thesis therefore performs a nonlinear analysis of RS in Chapter 7, followed by proposing several feature extraction methods for RS classification in Chapter 8 based on higher-order statistics.

Chapter 7

Tracking and Analysis of Nonlinearity in Respiratory Sounds

7.1 Introduction

All types of respiratory sounds (RSs) are primarily related to vibrations of the upper airway walls and the turbulent airflow [36]. RS may be adventitious in certain pathological conditions of the airways such as bronchial obstruction and this change of breathing pathway affects the spectral features of these adventitious RS [122]. Nonlinear characteristics of adventitious lung sounds (LSs) such as wheezes [87], crackles [123], and snore sounds [85] as well as the nonlinear dynamics of breathing pattern [124] have been investigated, while justification of nonlinear analysis applications has been provided in [13]. Although the origin and production mechanism are different for tracheal breath sounds (TBSs) and LS, many authors assume that the RS (including LS and TBS) signals are generally nonlinear. However, surrogate data tests could not provide statistically sufficient evidence regarding the data nonlinearity for TBS in [13].

The aim of this chapter is therefore to verify the assumption and to demonstrate the nonlinearity of various types of RS. The signal nonlinearity is first tracked using third-order cumulant. The proposed filter bank based method is then applied on the identified linear and nonlinear parts of the recorded RS data to provide the time-frequency (TF) analysis for nonlinearity. The method produces two high resolution images: weighted and unweighted outputs in TF domain. Similarity of these two images is calculated to justify the inherent nonlinearity of RS for patients with pulmonary dysfunctions.

7.2 Methodology

7.2.1 Experimental Dataset

Real recordings were done following the procedures as described by Section 2.6 with $F_s = 11.025$ kHz. In this study, the real TBS recordings were obtained from 7 healthy and 7 subjects with different degrees of airway obstruction (5 males/9 females, 15 ± 9 years old). Also, preclassified and preprocessed LS recordings were taken from 2 respiratory sound databases [37][31]. A total of 10 normal (including 7 real recordings) and 25 adventitious RS signals (including 20 real recordings) each with 10 seconds duration were used.

7.2.2 Nonlinearity Tracking

The nonlinear signals are referred to signals with wide spectra, significant frequency interactions, and nonlinear phase coupling. The possible nonlinear regions in the input data is tracked by using third-order cumulant. Compared to other higher-order cumulant, third-order cumulant being asymptotically zero for symmetric non-Gaussian

distributions, is more appropriate for the localization of the nonlinear regions with longer durations.

A short-time estimate of the third-order cumulant of an arbitrary zero-mean signal $x(n)$ introduced in (3.2) has been applied here. The length of the window $w(b)$ in (3.2) is $2B + 1$. Since the method is applied to the detection of wide band nonlinear signals, the choice of window length for the cumulant estimators is critical. For a wide band signal, the window length should not be large, since the amplitude of the output signals would become zero at some points. A proper window length of $2B + 1$ should enable the tracking of the cumulant variations with long duration due to nonlinearity. At the same time the relatively shorter variations caused by the heart sound (HS) or other environmental transient interferences should be ignored. This can be achieved by selecting a narrow window based on the HS duration. Therefore, the maximum window length is then chosen here as $B = 5$ at $Fs = 11.025$ kHz which corresponds to be $1/20$ of the minimum HS duration that is 20 ms [67].

The dimension of the estimates $\hat{C}_3(a_1, a_2; b)$ are reduced by calculating $\hat{C}_3(-a, a; b)$. Since the diagonal cumulant slices extract useful information while keeping modest computational complexity [125], it is a more suitable function for nonlinearity detection. As the cumulants decrease with increasing a , the proposed nonlinear function which has been earlier introduced in (3.3) is used in the tracking of nonlinearity. One motive for using the difference between $\hat{C}_3(0, 0; b)$ and $\hat{C}_3(-1, 1; b)$ as an identifying function is that the third harmonics are subtracted, thereby reducing the aliasing effect for the signals with wide frequency bands. The nonlinear region of our third-order cumulant function with high values can then be identified from the linear region with very small cumulant values. A threshold of 10^{-8} is used for the identification.

7.2.3 The Nonlinear Analysis Method

After cumulant calculation, the identified nonlinear/linear parts using (3.3) are extracted and stored separately. The extracted segments are then subjectively tested by experienced doctors for classification into different types of RS. The classified RS signals are then evaluated using the nonlinear analysis method proposed below for the nonlinearity verification.

We present a new nonlinear TF analysis method in this subsection. The method consists of two stages: The first stage is to design a weighted filter bank based on the weights calculated using the least square error criteria. The weighted filter bank then produces two outputs including an unweighted TF output and an optimally weighted averaged Wigner-Ville distribution (WVD) of the subband signals. The second stage is to perform similarity measurement between these two outputs of the filter bank.

Therefore, by given $x(n), 0 \leq n \leq (N - 1)$, with N being the length of the signal, we calculate L different WVDs,

$$W_i(n, k) = \sum_{b=-(K-1)/2}^{(K-1)/2} w(b) z_i(n+b) z_i^*(n-b) e^{-j4\pi b k / K}, i = 1, \dots, L \quad (7.1)$$

where $z_i(n)$ is the complex signal at the output of the i th subfilter in the filter bank and $n, k (0 \leq k \leq K - 1)$, and b are the indices corresponding to discrete time, discrete frequency, and time-lag. $w(b)$ is a real window and $\star'$ denotes the conjugate transpose. $z_i(n)$ is obtained as

$$z_i(n) = x(n) * h_i'(n) \quad (7.2)$$

where $\star$ denotes the convolution operator, and $h_i'(n)$ defined later in (7.5) is the weighted version of $h_i(n)$ with $h_i(n)$ being the impulse response of the i th subfilter of a uniform bandpass filter bank given by

$$h_i(n) = h(n) e^{j2\pi k_i n}, i = 1, \dots, L. \quad (7.3)$$

In (7.3), $h(n)$ with $0 \leq n \leq (l - 1)$ where $l = 256$ being the filter length, is the real-valued impulse response of a prototype lowpass filter and L is the number of subfilters. Each subfilter is obtained by modulating the lowpass filter by a complex exponential with the normalized frequency $k_i = i/L$ for $i = 1, \dots, L$.

Using least square error criterion, the optimum weight vector $\mathbf{c} = [c_1 \cdots c_L]^T$ of the filter bank can be obtained by solving the linear system

$$\mathbf{Q}\mathbf{c} = \mathbf{d}. \quad (7.4)$$

In (7.4), $\mathbf{Q} = \sum_{n=0}^{l-1} \mathbf{h}(n)\mathbf{h}^T(n)$ is a Hermitian and positive definite matrix of size $(L \times L)$ where $\mathbf{h}(n) = [h_1(n) \cdots h_L(n)]^T$ containing the impulse responses of the subfilters. Also, $\mathbf{d} = \sum_{n=0}^{l-1} \mathbf{h}(n)f(n)$ is an L -component column vector with $f(n)$ being the matched filter response which is the time reversal of the input sequence $x(n)$.

The impulse response $h'_i(n)$ of the optimally weighted filter bank will then be

$$h'_i(n) = c_i \underbrace{h(n)e^{j2\pi k_i n}}_{h_i(n)}, i = 1, \dots, L. \quad (7.5)$$

In the second stage, we optimally weight each of the WVDs, $W_{\{i|i=1,\dots,L\}}(n, k)$, by solving the least square (LS) problem

$$\begin{aligned} \min_{\gamma_i} \sum_n \sum_k \psi(n, k) & [R(n, k) - \sum_{i=1}^L \gamma_i W_i(n, k)] \\ & \times [R(n, k) - \sum_{i=1}^L \gamma_i W_i(n, k)]^* \end{aligned} \quad (7.6)$$

where γ_i 's are the weights, and $R(n, k)$ is a reference TF distribution which is the WVD of the $x(n)$ in our case. $\psi(n, k)$ is the real and positive weighting function with $\psi(n, k) = 1$ being the constant weighting in this context.

Using a one-dimensional indexing (via row-wise scanning) of $W_i(n, k)$ and $R(n, k)$, we can write (7.6) in matrix form as

$$\min_{\gamma} [\mathbf{R} - \mathbf{W}\gamma]^* \mathbf{\Psi} [\mathbf{R} - \mathbf{W}\gamma] \quad (7.7)$$

where $\gamma = [\gamma_1 \cdots \gamma_L]^T$ is a column vector and Ψ is the weighting matrix (positive definite and Hermitian). For simplicity, we have chosen the weighting matrix Ψ to be equal to the identity matrix. From (7.7) the optimal weights are then obtained as

$$\gamma^{\text{opt}} = (\mathbf{W}^T \Psi \mathbf{W})^s \mathbf{W}^T \Psi \mathbf{R}. \quad (7.8)$$

In (7.8), $(\mathbf{W}^T \Psi \mathbf{W})^s$ represents the pseudo inverse of the matrix $(\mathbf{W}^T \Psi \mathbf{W})(L \times L)$ which is easily calculated using the Cholesky factorization technique, since the above matrix is a symmetric positive-definite matrix.

Then the nonlinear TF analysis $\hat{\mathbf{R}}_{ls}$ as the weighted least square estimate of $\mathbf{R}$, together with the unweighted TF analysis $\hat{\mathbf{W}}$, are obtained as

$$\hat{\mathbf{R}}_{ls} = \sum_{i=1}^L \gamma_i^{\text{opt}} W_i(n, k) \text{ and } \hat{\mathbf{W}} = \sum_{i=1}^L W_i(n, k). \quad (7.9)$$

Since the two outputs $\hat{\mathbf{R}}_{ls}$ and $\hat{\mathbf{W}}$ are high resolution TF representations of the original signal, eigenvalue-based similarity measurement of images in [126] are adopted. Eigenvalue λ_S adopted as measurement for similarity between the two outputs is defined as

$$\lambda_S = \frac{1}{2} \left[m_{11} + m_{22} - \sqrt{(m_{11} - m_{22})^2 + 4m_{12}^2} \right] \quad (7.10)$$

where

$$m_{11} = \left[\frac{1}{K \times N} \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} \hat{R}_{ls}(n, k) \right] - (\overline{\hat{\mathbf{R}}_{ls}})^2$$

$$m_{22} = \left[\frac{1}{K \times N} \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} \hat{W}(n, k) \right] - (\overline{\hat{\mathbf{W}}})^2$$

$$m_{12} = \left[\frac{1}{K \times N} \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} \hat{R}_{ls}(n, k) \times \hat{W}(n, k) \right] - (\overline{\hat{\mathbf{R}}_{ls}} \times \overline{\hat{\mathbf{W}}})$$

$\overline{\hat{\mathbf{R}}_{ls}}$ and $\overline{\hat{\mathbf{W}}}$ are the mean values of $\hat{\mathbf{R}}_{ls}$ and $\hat{\mathbf{W}}$, respectively. High value of λ_S refers to low similarity between the two outputs.

7.3 Results and Discussion

Fig. 7.1 shows the illustrative plots of the third-order cumulant and linear/ nonlinear parts identified from a real RS. Since higher order statistics preserve the phase information of the signal, they can be used to detect the nonlinearity (quadrature phase coupling) and the deviation from Gaussianity of the signal.

Figure 7.1: (a) Third-order cumulant of a typical real RS with its threshold; (b) Original waveform of the RS with its corresponding nonlinear regions segmented.

From illustrative plots in Fig. 7.2, broad spectra (spreading to 2000 Hz) with at least two dominant frequency components are present for the nonlinear parts, while narrow spectra with no dominating frequency or single dominating frequency have been detected for linear parts. Spectrograms using short-time Fourier transform (STFT) are with either low frequency resolution or low time resolution due to the usage of window, and are therefore not able to resolve closely spaced frequency components. Furthermore, pure WVD suffers from severe cross-terms interference and therefore not able to track the auto-terms which represent the actual frequency components of the signal. By using the proposed analysis method which combines the weighted filter

bank with WVD, the optimally weighted output is able to preserve both the cross-terms and auto-terms with high resolution. Also, the unweighted output retains the auto-terms with higher resolution compared with those analyzed results using STFT. The evolution on nonlinearities with time which has not been taken into consideration by the cumulant detection, is therefore observed in terms of cross-terms.

Figure 7.2: Illustrative TF plots of (a) $\hat{\mathbf{R}}_{ls}$ (b) $\hat{\mathbf{W}}$ for a nonlinear segment and (c) $\hat{\mathbf{R}}_{ls}$ (d) $\hat{\mathbf{W}}$ for a linear segment from heart sound contaminated real recording.

Table 7.1 presents the resulting eigenvalue λ_S for linear and nonlinear segments. According to doctors evaluation, linear segments have been identified as normal RS and monophonic wheeze, while nonlinear segments have been identified as polyphonic

wheeze, stridor, crackles, and snoring. Mean (μ) and standard deviation (σ) of λ_S for each types of signals are displayed together with the nonlinearity detection results. Since nonlinearity implies more frequency interactions and nonlinear phase coupling of frequencies, cross-terms are captured by $\hat{\mathbf{R}}_{ls}$ for the nonlinear segments. The dissimilarity between $\hat{\mathbf{R}}_{ls}$ and $\hat{\mathbf{W}}$ is therefore high by the presence of nonlinearity and thus large λ_S is obtained for nonlinear segments. $\lambda_S < 1$ for linear segments, and it increases along with the increment of frequency components present in different types of RS.

Table 7.1: Nonlinearity detection and λ_S for different types of RS

Types of RS (No. of segments)	Decision	$\lambda_S(\mu \pm \sigma)$
Normal RS (30)	Linear	0.5 ± 0.3
Monophonic wheeze (10)	Linear	0.7 ± 0.2
Stridor (9)	Nonlinear	2.5 ± 2.4
Snoring (15)	Nonlinear	5.1 ± 2.6
Crackle (6)	Nonlinear	8.9 ± 5.2
Polyphonic wheeze (13)	Nonlinear	30.6 ± 10.1

However, high standard deviations have been observed in Table 7.1 for the resulting eigenvalue λ_S , especially for stridor, normal RS, and crackle. In order to improve the results, noise contribution has been removed using the preprocessing method in [54]. Although the appropriate recording condition suppresses the ambient noise significantly, nonlinear energy operator has been used to detect and remove the spikes due to transient environmental noises. Wavelet denoising technique is then applied to reduce the continuous environmental noise to minimum level without having distortion. The resulting eigenvalue λ_S for real RS recordings before and after preprocessing has been listed in Table 7.2. Significant improvement in standard deviation for λ_S is observed

Table 7.2: Eigenvalue λ_S ($\mu \pm \sigma$) for the noisy and preprocessed RS recordings of various types

Types of RS (No. of segments)	Noisy RS	Preprocessed RS
Normal RS (20)	0.6 ± 0.4	0.3 ± 0.1
Monophonic wheeze (6)	0.7 ± 0.3	0.5 ± 0.2
Stridor (6)	2.6 ± 2.7	2.4 ± 1.1
Snoring (15)	5.1 ± 2.6	5.1 ± 1.0
Crackle (4)	8.6 ± 5.8	8.9 ± 2.3
Polyphonic wheeze (10)	30.6 ± 11.2	30.6 ± 4.1

by standardizing RS signals using noise suppression. Furthermore, HS has not been removed since its presence in real recordings does not affect the performance of the tracking and the analysis methods proposed.

7.4 Chapter Conclusion

In this chapter, nonlinearity tracking in time series has been performed using higher-order cumulant. Attempts have been made to introduce a novel time-frequency domain method for nonlinearity analysis of RS. The approach provides some basics for the investigation of underlying pathology for adventitious sounds and their associated nonlinear characteristics. The resulting λ_S can be used as an index indicating degrees of nonlinearity caused by the interactions between different frequency components.

Chapter 8

Feature Extraction Methods for Respiratory Sound Classification

8.1 Introduction

In order to remove the inherent subjectivity caused by auscultation via stethoscope, the necessity of respiratory sound (RS) classification based on the extraction of relevant features has been established. According to the Computerized Respiratory Sound Analysis (CORSA) guidelines [36], RS can be divided into normal breath sounds and adventitious sounds. Normal breath sounds are characterized by broad spectrum of noise, while adventitious sounds are additional RS superimposed on the normal breath sounds. An important category of the latter refers to continuous adventitious sounds (CASs) of pseudo-periodic nature such as wheeze and stridor (i.e. low-frequency wheeze). The presence of such sounds are relevant to airway obstruction and many pulmonary diseases. They have dominant frequencies usually greater than 100 Hz and duration longer than 100 ms. In time-expanded waveform analysis, normal breath sounds signals are characterized as an irregular signal shape without repetitive pattern,

while CAS signals have periodic waveform, either sinusoidal or more complex.

As a crucial step in classification, feature extraction reduces the dimensionality of the pattern vector and provides a set of measurements with more discriminatory information and less redundancy to the classifier [127]. Therefore, for more effective RS classification, various feature extraction methods have been applied in the literature. Parametric approaches such as AR modeling are adopted where AR parameters are estimated using the Levinson-Durbin recursion algorithm in [128][50]. Since CAS can be viewed as narrow peaks in power spectrum, nonparametric frequency domain methods have also been widely used. These include power spectral density (PSD) estimation for percentile frequencies extraction [129], cepstral analysis for the extraction of MFCC [130][48], as well as signal coherence method measuring the spectral stability of signal in terms of magnitude and phase [49]. Furthermore, time domain feature extraction methods such as those based on Katz fractal dimension and variance fractal dimension measuring the signal complexity are presented in [47].

In particular, the classification accuracies by the energy based features including AR parameters, MFCC, signal coherence, and percentile frequencies depend on the amplitude of the CAS signals. These tend to produce less discriminative patterns and therefore the classifier has to be chosen more prudently to achieve reliable classification. On the other hand, features based on waveform fractal dimension that measure the signal complexity are less dependent on the signal energy. However, the signals are analyzed in broadband sense which makes these features less sensitive to intermittent, monophonic mild wheeze.

The purpose of this chapter is to develop new ways to extract discriminative features for more reliable RS classification (i.e. by classify RS signals into normal and CAS signals) with the presence of mild CASs. Time-frequency analysis has been adopted first to produce narrowband signals. Instantaneous kurtosis along frequency are then

recursively measured to select the appropriate narrowband signals to make the presence of CAS signals more prominent. New discriminating features based on the ratio of short-term and long-term AR averaging, and the averaged block-wise instantaneous kurtosis along time are therefore extracted from the narrowband signals at the selected frequency bin. Another new feature set based on sample entropy (*SampEn*) and its histogram distortion between different frequency bins is also proposed here. The incorporation of histogram as a statistical tool removes the dependency of classification accuracy on the adventitious sound intensity [131]. The extracted new feature sets are then evaluated and compared with the conventional features used in RS classification in terms of our modified clustering index, named separability index, by incorporating different distance matrices. By investigating the separability index values and the classification results using k-NN classifier, the proposed features are shown to be quite attractive in terms of their ability to produce large cluster separability and high classification accuracies.

8.2 Methodology

8.2.1 Data Acquisition

Real recordings were done following the procedures as described by Section 2.6 with $F_s = 11.025$ kHz. In this study, the experimental dataset consists of RS from 7 healthy and 17 pathological subjects with different degrees of airway obstruction (8 males/16 females, 15 ± 9 years old). However, the characteristics due to sex, age, weight were not taken into consideration.

The recorded RS signals were first segmented into individual inspiratory/expiratory segments, and manual classification results for these inspiratory/expiratory segments

were obtained respectively from an experienced doctor in Singapore National University Hospital by listening to each individual RS segment. RS segments belong to normal breath sounds were labeled as group 1 and those belong to CAS were labeled as group 2. The segmented and pre-classified inspiratory/expiratory segments were then used as the inputs to extract the proposed feature sets followed by RS classification into normal breath sounds and CAS.

8.2.2 Time-Frequency Analysis

Time-frequency analysis is a very powerful tool in the analysis of nonstationary signals, i.e., signals whose spectral contents vary with time. The purpose of the time-frequency (TF) method is to map a time-domain signal into its time-frequency domain and thereby, shows how the signal components of a time-varying signal are varying with time and frequency [132]. Here, we have used the efficient fast Gabor spectrogram for our time-frequency analysis (see Appendix D for more details).

Fast Gabor time-frequency distribution applied here transforms the input signal into narrowband signals which can be processed separately. This allows us to extract the discriminating features only at the selected frequency bins having more distinct time domain signal characteristics. As the identified peaks due to the presence of CAS (e.g. wheeze) are considered to be part of the same signal when the frequency proximity of the successive peaks (in terms of time-window position) is less than 50 Hz [120], the total number of frequency bins is chosen to be 128 to ensure this frequency resolution (i.e. $F_s/(2 \times 128) = 11025/256 < 50$). The parameters used for Gabor spectrogram, i.e. the variance of the Gaussian analysis window (i.e. α parameter) and the order O balancing the resolution and cross-term interference, have typical values of 160 and 6, respectively. These parameters are selected to satisfy the instantaneous frequency (IF)

law with IF being traditionally defined as the first derivative of the signal phase. In such case, each component of the multicomponent tonal signal (e.g. wheeze) has its energy concentrated near only one frequency each time [133]. Moreover, O is chosen to provide linear signal decomposition with high concentration while removing the noise and interference components. Since wheezes usually have not more than three harmonics [120], $O = 6$ is therefore selected to guarantee the number of peaks coexisting at each time instance to be less than four.

8.2.3 New Feature Set Based on Instantaneous Kurtosis

For RS classification, a new instantaneous kurtosis based on AR averaging (i.e., recursively) is proposed here. The normalized instantaneous kurtosis is measured as for RS classification, a new instantaneous kurtosis based on AR averaging (i.e., recursively) is proposed here. The normalized instantaneous kurtosis is measured as

$$K_k(n) = \frac{E\{|x(n, k)|^4\}}{E^2\{|x(n, k)|^2\}} - 3. \quad (8.1)$$

Here, $x(n, k)$ is the k th frequency output from the Gabor TF transform of the input inspiration/expiration segment $x(n)$, and $E\{\cdot\}$ denotes the expectation operation which is approximated using AR averaging (with details elaborated in Appendix E) as

$$\begin{aligned} E\{|x(n, k)|^4\} &\approx (1 - \alpha_k) \sum_{l=0}^n \alpha_k^{(n-l)} |x(l, k)|^4 \\ &= (1 - \alpha_k) K_k(n-1) + \alpha_k |x(n, k)|^4 \\ E^2\{|x(n, k)|^2\} &\approx [(1 - \alpha_k) \sum_{l=0}^n \alpha_k^{(n-l)} |x(l, k)|^2]^2 \\ &= [(1 - \alpha_k) K_k(n-1) + \alpha_k |x(n, k)|^2]^2 \end{aligned} \quad (8.2)$$

where $\alpha_k = \frac{1}{F_s T}$ is a constant smoothing parameter representing the integration time $T \gg 1/F_s$ [79] for AR averaging, and is able to track variations of the filtered input signal $x(n, k)$. Here, n and k are the sample (time) index and the discrete frequency

bin index respectively and $T = 0.1$ s is chosen corresponds to the minimum duration of CAS.

Mean Instantaneous Kurtosis with Varying Frequency

According to (8.1), the mean instantaneous kurtosis $K_{\text{mean}}(k)$ varying with the k th frequency index, is calculated as

$$K_{\text{mean}}(k) = \frac{1}{N} \sum_{n=1}^N |K_k(n)| \quad (8.3)$$

where N is the length of the signal $x(n)$. As normal breath sound captured over the trachea contains significant components within frequency range of [50 1200] Hz [36], and no CAS has been reported with a pitch of > 1600 Hz [25], mean instantaneous kurtosis are calculated for the narrowband signals of the first 40 frequency bins (i.e. $x(n, k)$, $1 \leq k \leq 40$) to cover the frequency range of [0 1723] Hz.

In probability theory and statistics, kurtosis is a measure of the “peakedness” of the probability distribution of a real-valued random variable. Large kurtosis value implies that the variance is mainly contributed by the infrequent extreme deviations, as opposed to the frequent modestly sized deviations. This concept is adopted here to differentiate the frequency bins interfered by the presence of CAS from those which are not. For narrowband signal at each frequency bin, the magnitude of TF plot along time axis is more uniform for broadband noise like normal breath sound compared to that for CAS, as shown in Figs. 8.2(a)-(c). Hence, the mean kurtosis calculated at each frequency bin is higher at the presence of CAS. The choice of the smoothing parameter $\alpha_k = \frac{1}{F_s T}$ with $T = 0.1$ s enables the tracking of the infrequent variations imposed by CAS but omits the frequent small variations correspond to normal breath sound. The higher order statistics $E\{|x(n, k)|^4\}$ is thus more prominent than $E^2\{|x(n, k)|^2\}$. As the illustrative signal used in Fig. 8.1 is interfered by CAS at frequency bin $k = 8$,

it has a corresponding peak $K_{\text{mean}}(k)$ at $k = 8$ as depicted by Fig. 8.2(d). Since the frequency bin k_{max} with maximum $K_{\text{mean}}(k)$ carries the most significant CAS characteristics, we select the respective narrowband signal $x(n, k_{\text{max}})$ for the following feature extraction steps.

Figure 8.1: Gabor spectrogram of a CAS containing signal with CAS occurred in frequency bin $k = 8$, from 0.4 s to 0.5 s.

Averaged Instantaneous Kurtosis with Fixed Block Size

The averaged instantaneous kurtosis with block index n_i , $\overline{K}_{k_{\text{max}}}^{bl}(n_i)$ is calculated as

$$\overline{K}_k^{bl}(n_i) = \frac{1}{bl} \sum_{n=n_i*bl+1}^{(n_i+1)*bl} |K_{k_{\text{max}}}(n)|. \quad (8.4)$$

The block size of bl is calculated as $bl = \frac{T_{bl}}{1/F_s}$, with the same $T_{bl} = 0.01$ s corresponds to 1/10 of a minimum CAS duration. It can be seen in Fig. 8.3 that the increased value

Figure 8.2: (a)-(c) Waveforms of $x(n, k)$ for $k = 5, 8, 20$ of the CAS signal as used in Fig. 8.1; (d) The resulting mean instantaneous kurtosis $K_{mean}(k)$ for varying frequency bin k of the same signal.

of $\overline{K}_k^{bl}(n_i)$ at $k = k_{max}$ for the same signal as used in Fig. 8.1 refers to the presence of CAS. The block wise averaged instantaneous kurtosis for broadband signal without TF analysis is also illustrated in Fig. 8.3 by a marked solid line. It shows the effect of using narrowband signal to enhance the CAS properties through CAS localization in frequency using (8.3). Since the block size bl is chosen as $1/10$ of the minimum duration of any intermittent CAS, at least ten $\overline{K}_{k=k_{max}}^{bl}(n_i)$ values would be used to signify the presence of CAS signals with raising values. The highest ten values of $\overline{K}_{k_{max}}^{bl}(n_i)$ which correspond to the presence of CAS are therefore selected to form the first feature set based on recursively measured instantaneous kurtosis.

Figure 8.3: The averaged instantaneous kurtosis $\bar{K}_k^{bl}(n_i)$ of the broadband signal used in Fig. 8.1 without TF analysis, and that of the filtered narrowband signals at $k = 5, k_{max} = 8, 10, 20, 30$.

8.2.4 New Feature Set Based on Discriminating Function

The proposed discriminating function based feature set is obtained using time-frequency analysis followed by taking the ratio of short-term and long-term AR averaging. The overall block scheme of the proposed method as well as the detail description of the discriminating function generation are illustrated in Fig. 8.4 with $K = 40$ being the total number of frequency bins of interest. Useful discriminating features are then extracted from the discriminating function output $D(n, k)$ at $k = k_{max}$ which is selected in Section 8.2.3.

Figure 8.4: (a) The block scheme for the proposed discriminating feature extraction method; (b) Generation of the discriminating function for the k th frequency bin.

Discriminating Function Generation

A discriminating function based on the ratio of short-term and long-term AR averaging of the signal envelop is proposed as

$$r_{x,k}(n) = \frac{E_{x,k}(n)}{\bar{E}_{x,k}(n)}. \quad (8.5)$$

Here,

$$\begin{aligned} E_{x,k}(n) &= (1 - \alpha_k) \sum_{l=0}^n \alpha_k^{(n-l)} |x(l, k)| \\ &= (1 - \alpha_k) E_{x,k}(n-1) + \alpha_k |x(n, k)| \end{aligned} \quad (8.6)$$

where $\alpha_k = \alpha$ is a short-term AR averaging parameter that relates to the integration time of the filtered signal envelop. It is sensitive to the rapid changes of the input signal $x(n, k)$, and is chosen to be constant for all k . For simplicity, the x index for

$r_{x,k}(n)$ has been ignored in the following derivation. Similarly,

$$\overline{E}_{x,k}(n) = (1 - \overline{\alpha}_k)\overline{E}_{x,k}(n-1) + \overline{\alpha}_k|x(n, k)| \quad (8.7)$$

where $\overline{\alpha}_k = \overline{\alpha}$ is the long-term AR averaging parameter corresponding to the integration time of the signal envelop which is sensitive to the slow changes of $x(n, k)$.

Since $E_{x,k}(n)$ and $\overline{E}_{x,k}(n)$ are the corresponding envelop values as obtained from the Fast Gabor TF distribution, we can express $10 \log r_k(n)$ as a difference function, $D_{x,k}(n)$, in dB:

$$\begin{aligned} D(n, k) &= 10 \log r_k(n) \\ &= 10 \log E_{x,k}(n) - 10 \log \overline{E}_{x,k}(n). \end{aligned} \quad (8.8)$$

In general, normal breath sounds are characterized as an irregular signal shape, without any repetitive pattern or sudden rapid deflections. In contrast, CAS like wheeze has a periodic waveform, either sinusoidal or more complex [36]. Therefore, short-term prediction of the signal envelop should be applicable to normal breath signal to capture the rapid fluctuation of the signal; while long-term prediction should be used for CAS, to enable the capturing of slow variations in the signal.

Based on the above mentioned rules, T representing the integration time related to short-term AR averaging parameter α , is chosen to be 0.005 s which corresponds to be 1/20 of minimum CAS duration to capture the fast changes of normal breath. On the other hand, the time constant $\overline{T}$ related to long-term parameter $\overline{\alpha}$ is chosen to be 0.5 s, which is 100 times of T but not longer than one respiratory segment. The slow variations of CAS are distinguished by setting $\overline{T} = 0.5$ s, $\alpha = \frac{1}{F_s T}$ and $\overline{\alpha} = \frac{1}{F_s \overline{T}}$ are then calculated.

Fig. 8.5(a) displays the $D(n, k)$ of the narrowband signals at $k = 5, k_{max}, 20$ and that for the broadband signal without time-frequency analysis for the CAS signal shown in Fig. 8.1. The large negative values of $D(n, k_{max})$ indicate the presence of CAS, and

show the effect of time-frequency analysis. While Fig. 8.5(b) demonstrates the same sets of $D(n, k)$ for a normal breath sound signal. It is clear that $D(n, k)$ is having higher values in the absence of CAS, and the differences between $D(n, k)$ at different bins or between narrowband signals and broadband signal are much smaller compared to CAS signal.

Figure 8.5: (a) The discriminating function of the broadband CAS signal used in Fig. 8.1 without TF analysis and that for the filtered narrowband signals at $k = 5, k_{max} = 8, 10, 20, 30$; (b) The discriminating function of a broadband normal breath sound signal without TF analysis, and that for the filtered narrowband signals at $k = 5, k_{max} = 7, 10, 20, 30$.

Center-Surround Contrast

Here, the absolute ratio of the difference function $D(n, k)$ above or below zero is treated as the center-surround contrast of the signal. This contrast is then chosen as the first feature based on the discriminating function.

Due to the different signal characteristics, normal breath sound signal is fast varying with overall mean around zero. Therefore, $E_{x,k}(n)$ as the short-term AR averaging factor, would have much larger values than $\overline{E}_{x,k}(n)$, that is the long-term AR averaging factor. On the other hand, CAS being quasi-stationary, has much higher long-term AR averaging factor $\overline{E}_{x,k}(n)$ than $E_{x,k}(n)$. Using (8.8), we find

$$\begin{aligned} E_{x,k}(n) > \overline{E}_{x,k}(n) &\Rightarrow D(n, k) > 0 \\ \overline{E}_{x,k}(n) > E_{x,k}(n) &\Rightarrow D(n, k) < 0. \end{aligned}$$

Therefore, the gain plot for normal breath sound would have less negative values while that for CAS would have more negative values. The ratio of the area within the gain plot above and below the zero-crossing line is thus an effective feature to discriminate normal breath sound from CAS. As indicated in Figs. 8.5(a) and (b), compared to normal breath sound signal, the CAS signals have more negative $R_{x,k}$ being defined as

$$R_{x,k} = \frac{\sum_{n=1}^N \{D(n, k) \mid D(n, k) > 0\}}{|\sum_{n=1}^N \{D(n, k) \mid D(n, k) < 0\}|} \quad (8.9)$$

where N is the length of $x(n, k)$. Since $D(n, k)$ is calculated based on AR averaging with different integration time, the values of $D(n, k)$ is not able to track the exact time positions of CAS although the presence of CAS decreases the value of $D(n, k)$. Therefore, only the minimum value of $D(n, k_{max})$ (with k_{max} being selected in Section 8.2.3) that reflects the presence of CAS the most, together with the center-surround contrast $R_{x,k}$, compose the second feature set for RS classification.

8.2.5 New Feature Set Based on Sample Entropy (*SampEn*)

The third proposed feature set is based on the calculation of *SampEn*. The extraction method consists of time-frequency analysis followed by the mapping into *SampEn* plane and the calculation of mean distortion of *SampEn* histograms of the selected frequency bins in the mapped plane. The overall block scheme of the presented method can be found in Fig. 8.6. Here n and j are the time index for TF and *SampEn* plane, while k and $K = 40$ denote the frequency bin index and the number of frequency bins of interest respectively.

Figure 8.6: The block scheme for the proposed *SampEn* based feature extraction method.

SampEn Calculation

Sample entropy is a nonlinear signal processing approach used to measure the signal complexity. The origin of sample entropy is the approximate entropy (*ApEn*), which is originally introduced in [108] to measure the regularity in time series. $ApEn(m, r, N)$ is the negative natural logarithm of the conditional probability that a data set of length N , having repeated itself within a tolerance r for m points, will also repeat itself for $m + 1$ points. However, *ApEn* is lack of relative consistency and is heavily dependent on the data length. As it is uniformly lower than expected for short data, *SampEn* that does not count self-matches is developed in [100] to reduce these biases.

$SampEn(m, r, N_s)$ values are computed according to the procedure as described in Appendix C, with parameters chosen as discussed in Section 5.2.6. A *SampEn*

plane is then constructed by calculating $SampEn(j, k)$ values using the filtered narrowband input signals $|x(n, k)|$. $SampEn(j, k)$ denotes the j th normalized $SampEn$ (with $|SampEn| \leq 1$) calculated from the output signal of the k th bandpass filter, where $1 \leq j \leq \lfloor \frac{N}{N_s} \rfloor$ and $1 \leq k \leq 128$. Since the length of each inspiration or expiration segment varies, the value of j changes with the length of each input segment.

8.2.6 Histogram Distortion Calculation

The histogram of each subband $SampEn(k)$ is given by

$$H_{SampEn}^k(z) = \frac{1}{|SampEn(k)|} \sum_j \delta(z - SampEn(j, k)) \quad (8.10)$$

where $1 \leq k \leq K$. Typical shapes of the concatenation representations of the histograms, $H_{SampEn} = (H_{SampEn}^1, \dots, H_{SampEn}^K)$, for the corresponding normal breath sound signal in Fig. 8.5(a) and the CAS signal in Fig. 8.5(b), are shown in Fig. 8.7. The representation is nonparametric in nature and is effective to characterize various $SampEn$ patterns. A statistical measurement which is independent of data length is therefore proposed here based on H_{SampEn}^k instead of $SampEn$.

The mean distortion measure between each H_{SampEn}^k and the mean histogram $\overline{H}_{SampEn}^k$ is

$$\overline{D}_{Hist} = \frac{1}{K} \sum_{k=1}^K \sum_z (H_{SampEn}^k(z) - \overline{H}_{SampEn}^k(z)) \times \log \frac{H_{SampEn}^k(z)}{\overline{H}_{SampEn}^k(z)} \quad (8.11)$$

$$\overline{H}_{SampEn}^k = \frac{1}{K} \sum_{k=1}^K H_{SampEn}^k. \quad (8.12)$$

The calculated $\overline{D}_{Hist}$ varies between $[0 \ 1]$ with high values indicating large distortion and vice versa.

The histograms of the sample entropy is used as a nonlinear separability criterion for differentiating between CAS and normal breath sound, with the linear TF analysis

Figure 8.7: (a) A partial representation with $k = [1 \ 10]$ of a concatenating histogram H_{SampEn} for the normal breath sound used in Fig. 8.5(b); and (b) The same partial H_{SampEn} for the CAS signal used in Fig. 8.5(a).

helps to filter and select the data samples. As shown in Fig. 8.7(a), without the presence of CAS, the broadband nature of normal breath sound produces similar patterns for the $SampEn$ histograms of the narrowband signals in every frequency bin. However, for the CAS signals in Fig. 8.7(b), the dominant spectral peaks causes a change of pattern at those bins where they are present. Therefore, the value of $\overline{D}_{Hist}$ would be higher for CAS containing signals and is lower for normal breath sounds. The value of $\overline{D}_{Hist}$ forms the third feature set for RS classification.

8.2.7 Evaluation of The Proposed Feature Sets

The input signals can be grouped into two clusters according to their respective labels as obtained from manual classification described in Section 8.2.1. Therefore, the extracted features using the proposed methods are evaluated before classification in terms of

cluster separability by a modified cluster index, namely separability index (SI). A large SI value implies a low intra-cluster distance and a high inter-cluster separation, while a low index value implies a high intra-cluster distance with a low inter-cluster distance.

The modified cluster index SI is defined as

$$SI = \frac{S_b}{S_b + S_w} \quad (8.13)$$

with mean intra-cluster distance S_w given in [134] as

$$S_w = \frac{\sqrt{\frac{1}{N_1} \sum_{i=1}^{N_1} d^2(f_i^{c1}, \bar{f}^{c1})} + \sqrt{\frac{1}{N_2} \sum_{i=1}^{N_2} d^2(f_i^{c2}, \bar{f}^{c2})}}{\sqrt{\frac{1}{N} \sum_{i=1}^N d^2(f_i, \bar{f})}} \quad (8.14)$$

where $f = f^1 \cup f^2$ is the overall input feature set as a union of those feature sets belong to cluster $c1$ (normal) and cluster $c2$ (CAS). $N1$ and $N2$ are the number of members in $c1$ and $c2$ with $N = N1 + N2$. The $d(f_i, \bar{f})$ is a distance metric between the feature vectors f_i and the mean $\bar{f} = \frac{1}{N} \sum_i f_i$. On the other hand, the inter-cluster distance S_b is modified from [127] as

$$S_b = \frac{1}{N} \left[N1 \times d^2(\bar{f}^{c1}, \bar{f}) + N2 \times d^2(\bar{f}^{c2}, \bar{f}) \right]. \quad (8.15)$$

S_w evaluates the distributions of the clusters with respect to that of the whole input set in terms of distance measure $d(\cdot, \cdot)$. A smaller value of S_w therefore indicates a higher average cluster compactness, while a larger S_b score indicates a larger overall dissimilarity among the clusters. Therefore, a reliable feature should be able to give a larger SI as defined in (8.13). The resulting SI values obtained by using different distance metrics $d(\cdot, \cdot)$ for the proposed feature sets are listed in Table. 8.2.

8.3 Results and Discussion

8.3.1 Experimental Dataset

A total of 72 expiratory normal breath sound segments, 72 inspiratory normal breath sound segments, 96 expiratory CAS segments, and 99 inspiratory CAS segments were available. For each of the input segment, 3 feature sets are obtained as presented in Sections 8.2.3, 8.2.4, and 8.2.5. These feature sets are used for feature evaluation in Section 8.3.2 and additional 4 feature sets have been generated by the fusion of of these 3 feature sets. The '+' sign denotes feature fusion which is the concatenation of the features. The size and composition of all the 7 feature sets obtained are listed in Table 8.1 and adopted for RS classification in Section 8.3.3.

Table 8.1: Composition and size of the proposed feature sets

Feature Set*	Composition	Size
DP	First 10 maximum values of $\overline{K}_{k_{max}}^{bl}$	1×10
KP	The minimum value of $D(n, k_{max}) + R_{x,k}$	1×2
SEP	$\overline{D}_{Hist}$	1×1
DP+SEP	10 maximum values of $\overline{K}_{k_{max}}^{bl} + \overline{D}_{Hist}$	1×11
SEP+KP	$\overline{D}_{Hist}$ minimum value of $+D(n, k_{max}) + R_{x,k}$	1×3
DP+KP	10 maximum values of $\overline{K}_{k_{max}}^{bl}$ + minimum value of $D(n, k_{max}) + R_{x,k}$	1×12
SEP+DP+KP	$\overline{D}_{Hist}$ + 10 maximum values of $\overline{K}_{k_{max}}^{bl}$ + minimum value of $D(n, k_{max}) + R_{x,k}$	1×13

* '+' indicates direct appending.

8.3.2 Separability Index (SI)

SI is applied to evaluate the proposed features sets DP, KP, SEP referring the extracted features by the methods based on discriminating function (Section 8.2.4), instantaneous kurtosis (Section 8.2.3), and $SampEn$ (Section 8.2.5). Different distance matrices $d(\cdot, \cdot)$ including Euclidean distance, city-block distance, Mahalanobis distance, Minkowski norm with constant exponent chosen as 4 [135], and Chebyshev distance, are used for the SI calculation. Some well-know features for RS classification are implemented for comparison which include: 6th order AR coefficients (AR) [50], Mel-Frequency Cepstral Coefficients (MFCC) [48], signal coherence (SC) [49], variance fractal dimension (VFD) and Katz fractal dimension (KFD) [47].

Table 8.2: Feature evaluation results for different feature sets using SI with different distance metrics

Features Sets	Euclidean	City-Block	Mahalanobis	Minkowski	Chebyshev
KP	0.932	0.7213	0.3323	0.4349	0.3419
DP	0.904	0.6248	0.4958	0.2560	0.2561
SEP	0.8771	0.5916	0.6334	0.2268	0.148
AR	0.6673	0.6573	0.3629	0.1361	0.1057
MFCC	0.6377	0.2021	0.4502	0.1373	0.1415
SC	0.5827	0.2028	0.2664	0.1779	0.2826
VFD	0.8027	0.2454	0.4336	0.3757	0.2738
KFD	0.8581	0.5235	0.2495	0.3677	0.2891

Table. 8.2 illustrates the comparison results for feature evaluation using SI calculated with different distance matrices $d(\cdot, \cdot)$. It can be observed that the energy based features like AR, MFCC, and SC have lower SI for all types of distances used. Since these features are sensitive to the presence of noise, there would be performance degra-

dations when different recording conditions or preprocessing steps (i.e. modification of signal by amplification, attenuation, filtering to improve audibility) are involved. Besides, features of KFD and VFD as measures of the degree of signal complexity, produce clusters with higher separability by giving higher values of SI . As fractal dimension using Katz algorithm is less sensitive to noise than variance fractal dimension, it gives a slightly higher SI compared to VFD .

Since the proposed feature sets have large discriminations for RS of different types, they produce the highest SI compared to other features. These results prove the reliability of the new feature sets and may imply the need of a less complicated classifier (such as nearest neighbor classifier) to achieve satisfiable classification performance. Furthermore, when comparing the SI value of the same feature set using different distance metrics, Euclidean distance and City-block distance give much larger separability than the others. When Euclidean distance is employed, the intra-cluster distance measure S_w becomes coherent to the average cluster scattering index used in [136]. Therefore, Euclidean distance which gives the largest SI for all feature sets, is used in the nearest neighbor classifier as described in the following Section 8.3.3 for RS classification.

8.3.3 Classification Results

Nearest Neighbor (NN) Classifier

The k -NN classifier is a nonparametric decision procedure in which the segment under test is classified by examining the categories of the nearest k neighbours and taking a majority vote [137]. Using the label information of the training sample, an unknown observation is compared with all the cases in the training sample. N distances between a feature vector and all the training features are calculated. The label information

which results in the minimum distance, is then assigned to the incoming feature. The distance function puts data points in order according to their distance to the query, and k determines how many data points are selected and used as neighbors. Classification is usually done by voting among the neighbors. It is common to select small and odd k to break ties (typically 1, 3 or 5) [138]. We adopt $k = 1$ which corresponds to the nearest neighbor classifier (1NN) for RS classification here. Also, the Euclidean distance metric which gives the largest SI in Section 8.3.2 is chosen for the implementation of this distance-based algorithm. The RS signals are classified here into two classes of normal breath sound and CAS.

Classification Performance

Classification performance has been evaluated on the 7 proposed feature sets using the quantitative measures defined by

$$\text{Sensitivity } (SE) = \frac{TP}{TP + FN} \times 100\%$$

$$\text{Specificity } (SP) = \frac{TN}{TN + FP} \times 100\%$$

$$\text{Accuracy}(AC) = \frac{TN + TP}{TN + TP + FP + FN} \times 100\%$$

where TP , TN , FP , and FN are the number of true positive (i.e. accurately classified CAS), true negative (i.e. accurately classified normal breath sound), false positive (i.e. wrongly classified CAS), and false negative (i.e. wrongly classified normal breath sound) classifications, respectively.

Classification experiments are carried out on inspiration and expiration segments separately. 50% of the experimental dataset are randomly chosen in the learning phase for training in order to construct the mean feature vectors from each class. The remaining 50% of the data are tested to obtain the classification accuracy. The learning

and classification phases are repeated for 50 trials by using different randomly selected training and testing dataset. The quantitative measures over the trials are then averaged to give the overall performances.

Fig. 8.8 summarizes the calculated RS classification results using 1NN classifier on the experimental dataset defined in Section 8.3.1. The first three feature sets DP, KP, SEP are defined in Section 8.3.2, while the last four feature sets refer to different combinations of DP, KP, and SEP. By combination, we refer to direct appending of one feature set after another with proper scaling by their respective value ranges [139].

Figure 8.8: Performance of the 1NN classifier using different feature sets obtained by the proposed methods.

Performance of the presented features achieves an accuracy over 90% when individual DP or KP feature set is used for both inspirations and expirations, while SEP

feature set alone produces a slightly lower accuracy and a much lower specificity compared to DP and KP. This might be due to the smaller dimensionality of the SEP feature set where single feature value might not be enough for accurate classification. Therefore the performance improves when combining SEP with the other two feature sets. However, the improvements by combining two feature sets are not significant. This implies that increasing the number of reliable features is not a determining factor for high performance accuracy. On the other hand, by combining all three feature sets, 100% and 95.9% accuracies can be achieved for expiration and inspiration respectively. When cross comparing the three quantitative measures SE , SP , and AC , the proposed features are able to give relatively higher sensitivity than specificity. This means that the features are more prone to wrongly classify normal breath sound into CAS but they are quite reliable in identifying CAS.

Additionally, the RS classification performance is also dependent on the choice of classifier although the choice of classifier should be driven by the factors such as size of the training set, missing values, and probability distributions rather than accuracy [140]. Furthermore, the choice of k depends also on the interactions between individual feature values. Although larger k values help to reduce the effects of noisy features within the training data set [138], smaller $k = 1$ gives a better performance here by having less correlated feature values (from the way they are extracted).

A multilayer perceptron (MLP) with ten hidden layer nodes and one output node has also been implemented as the classifier. Levenberg-Marquardt optimization algorithm is adopted to update the MLP [141]. However, the implemented neural network classifier is not able to give satisfiable results using the same size of training and testing dataset. Compared to k-NN classifier, it is more effective to recognize predictive pattern when a class is associated with a complex and obscured combination of changes of the features being classified. However, this leads to the requirement of a large diversity

of training, and the minimization of over-fitting requires a great deal of computational effort. Furthermore, the relatively large and linear discriminations between the extracted feature values for different classes are observed, the nearest neighbor classifier is therefore more appropriate for the relatively small dataset used here, in terms of processing speed and accuracy.

8.4 Chapter Conclusion

This chapter introduces three novel feature extraction methods based on instantaneous kurtosis, discriminating function, and histogram distortion of sample entropy. TF analysis is applied with proper frequency bin selection to enhance the possible traces of CAS in the input signals. Time domain characteristics of the signal in different frequency bins are adopted for effective feature extraction from the filtered narrowband signals. Compared to the existing features for RS classification, the proposed features being non-energy based, are more resistant to the presence of noise and therefore more persistent under various data acquisition or preprocessing conditions. The feature evaluation using cluster separability measure and the classification results show that the new features extracted using the presented methods are quite reliable for accurate RS classification.

Chapter 9

Conclusion and Recommendations

A brief summary of research contributions and concluding remarks have been given in this chapter, followed by the highlights on some open research areas (which are not addressed in this thesis) for possible extensions of the current work in future.

9.1 Summary and Concluding Remarks

There are large varieties of normal and abnormal RSs with characteristics, which are typical for a disease or a certain pathological change in the respiratory system. The ability to analyze the acoustic patterns of these breathing-induced phenomena would certainly improve the knowledge of the physiology and pathophysiology of respiratory disorders for clinical assessment. This thesis introduces and addresses several technical challenges for the analysis of RS signals. The advanced signal processing methods have been proposed to solve three of the challenges including HS cancellation, respiratory phase detection, as well as RS feature extraction and classification.

Issues of signal characteristics such as nonlinearity, nonstationarity, and non-Gaussianity for RS classification, phase detection, and signal denoising are addressed with new methods proposed. The nonlinearity of the signal has been firstly tracked in Chapter 3 to facility accurate signal denoising (i.e. HS cancellation). Time frequency analysis has then been introduced in Chapter 7 to investigate and quantify the inherent nonlinearity of RS signals as related to different types of pulmonary dysfunctions. The difference in nonstationarities between the HS and RS signals have been investigated in Part I of the thesis for complete HS cancellation. The method produces a maximum PSD difference of 3.67 ± 3.01 dB between the original and reconstructed RS signals as compared to a difference of 8.67 ± 6.54 dB by the existing method in [7]. The nonstationarity difference between different respiratory phases has also been analyzed in Part II for reliable respiratory phase segmentation and identification. An identification accuracy of 95.5% for discontinuous adventitious sounds has been achieved as compared to an accuracy of 74.4% produced by the existing phase detection method in [9]. Furthermore, the non-Gaussianity and nonstationarity differences between different types of RS signals have been examined in Part III of the thesis to introduce new feature extraction methods for reliable RS classification. The proposed feature sets give an highest separability index value of 0.932 as compared to 0.8581 yield by the existing feature introduced in [47], with higher index value indicating higher cluster separability. The proposed feature sets therefore produces an average classification accuracy of 92.2% using k -NN classifier. Based on the performances of the proposed RS processing methods, the fusion of the presented methods could result in more efficient RS analysis.

Although the analytical techniques of signal processing are largely independent of the application, the interpretation of their results on biological data (which is RS in this thesis) still requires substantial understanding of the physiological system involved. Therefore, the RS analysis may be further enhanced with the inclusion of knowledge

from the modeling of cardiovascular system to the accompanying field of sound modeling. In this way, the obtained parameters by signal processing algorithms could be attributed directly with pathophysiological meanings.

9.2 Future Works

The denoising methodology (i.e. HS cancellation) and the new analysis domains (for respiratory phase detection and RS classification) presented in the previous six chapters, respectively, show the potentiality of the advanced signal processing techniques to offer a wide range of new opportunities in RS analysis. The previous chapters show the intrinsic RS characteristics as well as the requirement of appropriate methods to efficiently process RS in a more straightforward manner. Therefore, the proposed methods could be extended in the future as follows:

9.2.1 Issues for HS Cancellation

As described in Chapter 3 and Chapter 4, the HS localization and removal methods proposed are effective for the presence of continuous adventitious sound (CAS) since the localization scheme is based on the characteristic difference between RS (which is relatively slow-varying) and HS (which is transient). Challenging issues are thus raised that is worthy of further investigations: 1) How to generalize the HS localization method to include the presence of discontinuous adventitious sound (DAS) such as crackles which have similar characteristics as HS; 2) How to generalize the HS cancellation method to produce accurate estimation of DAS with very short duration from a RS template which might not contain the same information.

Further investigations on the effect of signal nonlinearity on the generalized like-

likelihood ratio test could also be performed. The possibility of enhancing nonlinearity tracking to make the HS localization method robust against the attack of DAS is therefore worth examining. The possibilities of extending binary hypothesis testing into ternary or N-nary to solve the problem could also be considered.

On the other hand, adaptive weighting matrix might be introduced in template matching for HS cancellation, and the template selection scheme could be improved. Furthermore, since the proposed mixing model for RS and HS is still fuzzy, the effect of the correlation between RS and HS could be further studied to explore the possible improvement in performance by the incorporation of a more accurate signal model.

9.2.2 Issues for Flow Volume Estimation

The existing flow estimation methods are model based and are thus very much interfered by the presence of adventitious sounds. Therefore, the challenge remains with the flow volume estimation for different types of RS. Furthermore, it would also be favorable to remove the need of training data for calibration. Effective features reflecting the characteristics of RS signals, as well as an appropriate and consistent signal model, could be developed to match between the flow and the extracted features. Since a linear model is always preferred, the possibility of the proposed feature (i.e. the sample entropy) could also be studied for automatic adaption of the model parameters at different target flow rates.

9.2.3 Issues for RS Classification

The index λ_S proposed in Chapter 7 could be used to indicate the degrees of nonlinearity. Therefore, it might be further improved and incorporated as a feature for RS classification according to the degrees of interactions between different frequency

components in RS signals. Since the performance of the RS classification proposed in Chapter 8 depends on both the pertinence of the extracted features and the choice of classification method, appropriate classification methods as well as the possible ways for feature dimensionality reduction could also be taken into consideration.

Furthermore, as the proposed feature extraction methods are so far applicable for the classification of CAS, the challenges remain for a more general classification using the proposed features or proposing additional new effective features which enable to count for the characteristic differences among all types of RS signals.

9.2.4 Applications

- The techniques presented in this thesis are all demonstrated using RS signals. The proposed HS localization and removal methods could be generalized to other biomedical signals, such as ECG and electromyogram (EMG). The works presented in this thesis would be further investigated for the extraction of heart related signals or the underlying muscle movements by considering the practical requirements of these signals.
- The applications of the proposed respiratory phase segmentation method could be generalized by modifying the evaluation function in the MPGA algorithm to adapt the segmentation of other signals such as music signals. The annotating index proposed for respiratory phase identification could be possibly extended for the identification of 2D signals.
- The extracted features for RS classification could be further studied in the extraction of adventitious sounds from the abnormal RS signals.
- The practical implementation issues regarding the works presented in the thesis

could also be investigated for further integration into a RS analysis system.

Appendices

Appendix A:

Implementation of GLRT

The sequence $\mathbf{y}_n$ contains signal plus noise under the hypothesis H_1 in (3.6) with its covariance $\mathbf{R}_{yy}$ given by (3.9). The log-conditional probability density function of $\mathbf{y}_n$ under H_1 is

$$\ln p(\mathbf{y}_n; H_1) = -\frac{I}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^I \left(\ln(\sigma_i^2) + \frac{\mathbf{y}_n^H U_i U_i^H \mathbf{y}_n}{\sigma_i^2} \right). \quad (\text{A.1})$$

Furthermore, $\mathbf{y}_n$ contains only noise under the null hypothesis H_0 in (3.6). The log-conditional probability density function of $\mathbf{y}_n$ under this hypothesis becomes

$$\ln p(\mathbf{y}_n; H_0) = -\frac{I}{2} \ln 2\pi - \frac{I}{2} \ln \sigma_s^2 - \frac{\mathbf{y}_n^H \mathbf{y}_n}{\sigma_s^2} \quad (\text{A.2})$$

and the log-likelihood ratio function in (3.7) becomes

$$L(\mathbf{y}_n) = \ln p(\mathbf{y}_n; H_1) - \ln p(\mathbf{y}_n; H_0). \quad (\text{A.3})$$

Using (A.1) and (A.2), we get

$$L(\mathbf{y}_n) = -\frac{1}{2} \sum_{i=1}^I \ln \sigma_i^2 + \frac{I}{2} \ln \sigma_s^2 + \frac{1}{2\sigma_s^2} \sum_{i=1}^I \left(\frac{\sigma_i^2 - \sigma_s^2}{\sigma_i^2} \right) |U_i \mathbf{y}_n|^2. \quad (\text{A.4})$$

The test statistics $T(\mathbf{y}_n)$ for the Generalized Likelihood Ratio Test (GLRT), obtained by combining the non-data dependent terms into the threshold and scaling, is therefore

$$T(\mathbf{y}_n) = \sum_{i=1}^I G_i |U_i \mathbf{y}_n|^2; \text{ where } G_i = \frac{\sigma_{vi}^2}{\sigma_i^2}; \sigma_{vi}^2 = \sigma_i^2 - \sigma_s^2 \quad (\text{A.5})$$

with the singular value σ_{vi} characterizing the signal component. Thus the decision strategy becomes

$$T(\mathbf{y}_n) \begin{cases} > \gamma \text{ decide } H_1 \\ \leq \gamma \text{ decide } H_0 \end{cases}, \gamma = \sigma_s^2 \left(\sum_{i=1}^I \ln \sigma_i^2 - I \ln \sigma_s^2 \right). \quad (\text{A.6})$$

In order to estimate the test statistics $T(\mathbf{y}_n)$ and the threshold γ , the singular matrix U under hypothesis H_1 (i.e. the projection of $\mathbf{y}_n$, $\{U_i, \mathbf{y}_n\}$) and the singular values of the covariance matrices under hypothesis H_0 and H_1 have to be obtained.

Since the covariance matrix can be expressed as

$$\hat{\mathbf{R}}_{yy} = \hat{\mathbf{R}}_{ss} + \hat{\mathbf{R}}_{vv} \quad (\text{A.7})$$

and the singular-value-decomposition (SVD) of the covariance matrix yields

$$\hat{\mathbf{R}}_{yy} = \hat{U} \hat{\Lambda} \hat{U}^H; \hat{\Lambda} = \text{diag}[\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_I^2], \quad (\text{A.8})$$

the SVD of $\hat{\mathbf{R}}_{ss}$ and $\hat{\mathbf{R}}_{vv}$ can therefor be expressed in terms of the singular matrix $\hat{U}$ of $\hat{\mathbf{R}}_{yy}$ as

$$\begin{aligned} \hat{\mathbf{R}}_{vv} &= \hat{U} \hat{\Lambda}_{vv} \hat{U}^H & \hat{\Lambda}_{vv} &= \text{diag}[\hat{\sigma}_{v1}^2, \hat{\sigma}_{v2}^2, \dots, \hat{\sigma}_{vI}^2] \\ \hat{\mathbf{R}}_{ss} &= \hat{U} \hat{\Lambda}_s \hat{U}^H & \hat{\Lambda}_s &= \hat{\sigma}_s^2 I \\ \hat{\Lambda} &= \hat{\Lambda}_v + \hat{\Lambda}_s & \hat{\sigma}_i^2 &= \hat{\sigma}_{vi}^2 + \hat{\sigma}_s^2. \end{aligned} \quad (\text{A.9})$$

By separating the dominant singular values from the rest as

$$\hat{\Lambda} = \text{diag}[\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_r^2, \hat{\sigma}_{r+1}^2, \dots, \hat{\sigma}_I^2] \quad (\text{A.10})$$

with $\hat{\sigma}_{\{i|i=1,2,\dots,r\}}$ being the dominant or principal singular values, the singular matrix $\hat{U}$ can be partitioned into

$$\begin{aligned} \hat{U} &= [\hat{U}^v \ \hat{U}^s], \\ \hat{U}^v &= [\hat{U}_1 \dots \hat{U}_r], \\ \hat{U}^s &= [\hat{U}_{r+1} \dots \hat{U}_I]. \end{aligned}$$

The dominant singular values $\hat{\sigma}_{\{i|i=1,2,\dots,r\}}$ and the corresponding dominant singular matrix $\hat{U}^v$ characterize the signal plus noise components if $\hat{\sigma}_{\{i|i=1,2,\dots,r\}} \gg \hat{\sigma}_s$. The columns of the $\hat{U}^v$ matrix therefore spans the signal plus noise space, while the columns of the $\hat{U}^s$ spans its orthogonal complement, namely the noise space [70]. The singular

matrix $\hat{U}$ under the hypothesis H_1 can be then substituted by its estimate $\hat{U}^v$. At the same time, the singular values $\{\sigma_i\}$ and $\hat{\sigma}_s$ can be substituted by their estimates $\hat{\sigma}_{\{i|1,2,\dots,r\}}$ and $\hat{\sigma}_{\{i|r+1,\dots,I\}}$ respectively. Thus, $\hat{U} \approx \hat{U}^v$, $\hat{\sigma}_i \approx \hat{\sigma}_{\{i|1,2,\dots,r\}}$, and $\hat{\sigma}_s \approx \frac{1}{I-r} \sum_{i=r+1}^I \hat{\sigma}_i$.

The decision strategy is therefore implemented as

$$\hat{T}(\mathbf{y}_n) \begin{cases} > \hat{\gamma} \text{ decide } H_1 \\ \leq \hat{\gamma} \text{ decide } H_0 \end{cases} \quad (\text{A.11})$$

where,

$$\begin{aligned} \hat{T}(\mathbf{y}_n) &= \sum_{i=1}^r \hat{G}_i |\hat{U}_i \mathbf{y}_n|^2; \text{ where } \hat{G}_i = \frac{\hat{\sigma}_{vi}^2}{\hat{\sigma}_i^2}; \hat{\sigma}_{vi}^2 = \hat{\sigma}_i^2 - \hat{\sigma}_s^2 \\ \hat{\gamma} &= \hat{\sigma}_s^2 \left(\sum_{i=1}^r \ln \hat{\sigma}_i^2 - (I-r) \ln \hat{\sigma}_s^2 \right). \end{aligned}$$

Since a) the dominant singular values, $\hat{\sigma}_{\{i|1,2,\dots,r\}}$, correspond to the power spectral density associated with the measurement consisting of signal plus noise, b) the non-dominant singular values, $\hat{\sigma}_{\{i|r+1,r+2,\dots,I\}}$ refer to the spectral densities of the noise [70], and c) the signal dominates the noise in the spectral bandwidth of the signal:

$$\hat{\sigma}_i^2 \approx \begin{cases} \hat{\sigma}_{vi}^2 & \text{for } i = 1, \dots, r \\ \hat{\sigma}_s^2 & \text{for } i = r+1, \dots, I. \end{cases}$$

In view of this, the test statistics can be approximated as in (3.10). However the dominant singular values, and the singular vector characterize the noise components if $\hat{\sigma}_{\{i|1,\dots,r\}} \ll \hat{\sigma}_s$. As a consequence, the decision strategy can be different in the sense that the hypothesis H_1 being replace by H_0 and vice versa.

Appendix B:

Time Scaling for Source Extraction Scheme

This appendix presents the time scaling approach used in our source extraction scheme. This algorithm is adapted from [142][143]. The scaling factor α is firstly defined as

$$\alpha = \frac{R_a}{R_s} \quad (\text{B.1})$$

where analysis hop size R_a and synthesis hop size R_s are the key factors for time scaling.

Here, we modify the signal as a weighted sum of cosines described in [142][143]. This time scaling algorithm first extracts a series of sequential blocks by setting analysis time instants along the input signal for successive integer values. For each block, FFT is computed over a Hann windowed portion of the original signal to obtain a magnitude and phase representation for every R_a samples as follow

$$Y_j^t(t_a^u, \Omega_k) = \sum_{n=-\infty}^{\infty} w(n)y_j^t(t_a^u + n)e^{-j\Omega_k n}. \quad (\text{B.2})$$

Here, STFT representation of the input signal y_j^t , $Y_j^t(t_a^u, \Omega_k)$, is obtained for the analysis time instants t_a^u for set of successive integer values u starting at 0. $w(n)$ is the analysis Hann window, $\Omega_k = \frac{2\pi k}{N_f}$ is the centre frequency of the k th vocoder channel, N_f is the size of discrete Fourier transform and $t_a^u = uR_a$.

At each synthesis time instants t_s^u , a short-time signal $y_j^u(n)$ is obtained for $t_s^u = R_s u$ by taking the inverse Fourier transform of the synthesis STFT as $Y_j^t(t_s^u, \Omega_k)$.

$$y_j^u(n) = \frac{1}{N_f} \sum_{k=0}^{N_f-1} |Y_j^t(t_s^u, \Omega_k)| e^{j\Omega_k n} \quad (\text{B.3})$$

where interpolated magnitude values are the same as the analysis magnitude values

$$|Y_j^t(t_s^u, \Omega_k)| = |Y_j^t(t_a^u, \Omega_k)|. \quad (\text{B.4})$$

When all the short-time signals are concatenated together, the output signal $y_j^T(n)$ is obtained as

$$y_j^T(n) = \sum_{n=-\infty}^{\infty} y_j^u(n - t_s^u). \quad (\text{B.5})$$

In this context, to calculate the phase of $y_j^t(t_s^u, \Omega_k)$, the phase increment per sample is also necessary to be computed as the hop size for the synthesis which is different from the analysis

$$\Delta\Phi_k^u = \angle Y_j^t(t_a^u, \Omega_k) - \angle Y_j^t(t_a^{u-1}, \Omega_k) - R_a \Omega_k. \quad (\text{B.6})$$

The instantaneous frequency $\hat{w}_k(t_a^u)$ is then derived after taking the principal value of the phase increment, $\Delta_p\Phi_k^u$, as

$$\hat{w}_k(t_a^u) = \Omega_k + \frac{1}{R_a} \Delta_p\Phi_k^u. \quad (\text{B.7})$$

After the instantaneous frequency at time t_a^u is estimated, the phase of the time scaled STFT at time t_s^u is set using the following formula

$$\angle Y_j^t(t_s^u, \Omega_k) = \angle Y_j^t(t_s^{u-1}, \Omega_k) - R_s \hat{w}_k(t_a^u). \quad (\text{B.8})$$

Appendix C:

Sample Entropy Calculation

$SampEn(m, r, N_s)$ can be calculated as below:

For an input signal U of length N_s , $\{U(p) : 1 \leq p \leq N_s\}$ forms the $N_s - m + 1$ vectors $x_m(q)$ for $\{q | 1 \leq q \leq N_s - m + 1\}$, where $x_m(q) = \{U(q + b) : 0 \leq b \leq m - 1\}$ is the vector of m data points from $U(q)$ to $U(q + m - 1)$. In this context, only the first $N_s - m$ vectors of length m are considered to ensure that, $x_m(q)$ and $x_{m+1}(q)$ are defined for $1 \leq q \leq N_s - m$. Let $B^m(r)$ be the probability that two sequences will match for m points and $A^m(r)$ is the probability that two sequences will match for $m + 1$ points. $B_q^m(r)$ is defined as $(N_s - m - 1)^{-1}$ times the numbers of vectors $x_m(p)$ within r of $x_m(q)$, where $1 \leq p \leq N_s - m$, and $p \neq q$ to exclude self-matches. Then $B^m(r)$ is defined as

$$B^m(r) = (N_s - m)^{-1} \sum_{q=1}^{N_s-1} B_q^m(r). \quad (C.1)$$

Similarly, $A_q^m(r)$ is defined as $(N_s - m - 1)^{-1}$ times the numbers of vectors $x_{m+1}(p)$ within r of $x_{m+1}(q)$, where $1 \leq p \leq N_s - m$ and $p \neq q$. Then set $A^m(r)$ as

$$A^m(r) = (N_s - m)^{-1} \sum_{q=1}^{N_s-1} A_q^m(r). \quad (C.2)$$

Finally, sample entropy ($SampEn$) is calculated by

$$SampEn(m, r, N_s) = -\ln \frac{A^m(r)}{B^m(r)}. \quad (C.3)$$

Appendix D:

Computation of Gabor Transform

This appendix presents the possible way to compute Gabor transform as suggested in [132]. In general, Gabor transform is known as sampled short time Fourier transform computed as

$$C_{k,n} = STFT(kT, n\Omega) = \int_{-\infty}^{\infty} x(t)w(t - kT)e^{-jn\Omega t} dt \quad (D.1)$$

where Gabor coefficient $C_{k,n}$ describes the signal's behavior in the vicinity of the point $\{kT, n\Omega\}$. Here, $STFT(\cdot)$ represents short time Fourier transform of the signal, where T and Ω denote respectively the time and frequency sampling steps. $w(t) = (\alpha/\pi)^{1/4}e^{-(\alpha/2)t^2}$ with α being adjustable to balance between the time and frequency resolutions of $w(t)$ according to the applications. The analytic expression of the Gabor spectrogram can be found by [144]

$$S(t, \omega) = \sum_{|k-k'|+|n-n'|\leq O} [2C_{k,n}C_{k',n'}^* \exp\{-\alpha^{-1}(t - \frac{k+k'}{2}T)^2 - \alpha(\omega - \frac{n+n'}{2}\Omega)^2\} \exp\{j(n - n')Tt - j(k - k')T(\omega - \frac{n+n'}{2}\Omega)\}] \quad (D.2)$$

with the parameter O denoting the order of the Gabor spectrogram. Gabor spectrogram has a good tradeoff between the short time Fourier transform based spectrogram and Wigner-Ville distribution. The fast Gabor spectrogram realizes the Gabor transform as a sum of separated two-dimensional interpolation filters. This makes the approach not only straightforward, but also fast and therefore, found useful for time-frequency analysis of nonstationary signals such as biomedical signals.

Appendix E:

First-Order AR Averaging

This appendix presents a brief description of the first-order auto-regressive (AR) averaging which is applied in the adaptive function generator. Consider a first-order AR process for an input signal $x(n)$, the output signal $y(n)$ is

$$y(n) = \alpha y(n-1) + \beta x(n) \quad (\text{E.1})$$

where β is a scaling factor [79]. The parameter α ($0 < \alpha < 1$) characterizes the AR averaging by controlling the integration time. It is related to integration time (s) by the expression of $\alpha = (1 - A)^{\frac{1}{2(N+1)}}$ which $A \in [0, 1]$ and $N = F_s T$ (F_s : Sampling frequency in Hz; T : Integration time in s). For $T \gg \frac{1}{2(N+1)}$, α can be approximated as $\frac{1}{F_s T}$. The transfer function and the corresponding impulse response of the AR process in (E.1) is

$$H(z) = \frac{\beta}{1 - \alpha z^{-1}} \quad (\text{E.2})$$

$$h(n) = \beta \alpha^n u(n) \quad (\text{E.3})$$

where $u(n)$ is the Heaviside's unit step function.

Author's Publications

Book Chapter

1. F. Jin and F. Sattar, "Enhancement of Recorded Respiratory Sound Using Signal Processing Techniques", in *Encyclopedia of Information Communication Technology*, A. Cartelli and M. Palma, Eds, Idea Group Inc., USA, 2008.

Journal Papers

1. F. Jin and F. Sattar. Tracking and Time-Frequency Analysis on Nonlinearity of Tracheal Sounds. *Medical & Biological Engineering & Computing*, 47(4), 2009.
2. F. Jin, F. Sattar, and DYT Goh. A Filter Bank Based Source Extraction Algorithm for Heart Sound Removal in Respiratory Sounds. *Computers in Biology and Medicine*, 39(9), 2009.
3. F. Jin, F. Sattar, and DYT Goh. An Acoustical Respiratory Phase Segmentation Algorithm Using Genetic Approach. *Medical & Biological Engineering & Computing*, 47(9), 2009.
4. F. Jin, F. Sattar, and DYT Goh. New Feature Extraction Methods for Respiratory Sound Classification. *IEEE Trans. Biomed. Eng.*, submitted.
5. F. Jin, F. Sattar, S. G. Razul, DYT Goh, and I. M. Louis. Pathological Respiratory Sound Analysis and Phase Identification By A New Annotating Index. *Biomedical Signal Processing and Control*, submitted.
6. F. Jin, F. Sattar, and DYT Goh. Heart Sound Localization From Respiratory Sound Using A Robust Wavelet Based Approach. *Digital Signal Processing*, submitted.
7. F. Jin, F. Sattar, and DYT Goh. A Feature Analysis on Respiratory Sounds for Phase Detection. *IEEE Trans. Info. Tech. Biomed.*, submitted.

Conference Papers

1. F. Jin, F. Sattar, S. G. Razul, and DYT Goh. Discrimination of Respiratory Sounds Based on A New Adaptive Function. *Proc. 6th IEEE Conf. Info. Comm. Signal Proc. (ICICS)*, 2007.

2. F. Jin and F. Sattar. A New Automated Approach for Identification of Respiratory Sounds. *Proc. IEEE Conf. Multimedia Expo (ICME)*, 2007.
3. F. Jin, F. Sattar, and DYT Goh. A Prototype Respiratory Sound Acquisition and Analysis System: Report of Preliminary Work. *National Healthcare Group (NHG) Annual Scientific Congress (ASC)*, 2007.
4. F. Jin, F. Sattar, and M. Pwint. Application of Walsh Transform Based Method on Tracheal Breath Sound Signal Segmentation. *Proc. IEEE-EMBS Conf. Bio-inspired Systems Signal Processing (BIOSIGNALS)*, 2008.
5. F. Jin, F. Sattar, and M. Pwint. Phase Segmentation of Noisy Respiratory Sound Signals Using Genetic Approach. *Proc. IEEE-EMBS Conf. Bio-inspired Systems Signal Processing (BIOSIGNALS)*, 2008.
6. F. Jin and F. Sattar. Heart Sound Cancellation From Respiratory Sound Using A Filter Bank Based Semi-blind Source Extraction Algorithm. *Proc. IEEE Conf. Multimedia Expo (ICME)*, 2008.
7. F. Jin, F. Sattar, S. G. Razul, and DYT Goh. Heart Sound Localization From Respiratory Sound Using A Robust Wavelet Based Approach. *Proc. IEEE Conf. Multimedia Expo (ICME)*, 2008.
8. F. Jin, F. Sattar, and DYT Goh. Automatic Wheeze Detection Using Histograms of Sample Entropy. *Proc. 30th IEEE Eng. Med. Biol. Soc. (EMBS)*, 2008.
9. F. Jin, F. Sattar, DYT Goh, and I. M. Louis. A Robust Respiratory Phase Identification Scheme Based on A New Mixing Index. *Proc. 17th European Signal Processing Conf. (EUSIPCO-2009)*, 2009.
10. F. Jin, F. Sattar, DYT Goh, and I. M. Louis. An Enhanced Respiratory Rate Monitoring Method for Real Tracheal Sounds. *Proc. 17th European Signal Processing Conf. (EUSIPCO-2009)*, 2009.

Bibliography

- [1] H. Kiyokawa, M. Yonemaru, S. Horie, I. Kasuga, Y. Ichinose, and Y. Toyama. Detection of nocturnal wheezing in bronchial asthma using sleep intermittent tracheal sound recording. *Respirology*, 4:37–45, 1999.
- [2] M. L. Hsueh, J. C. Chien, F. C. Chang, H. D. Wu, and F. C. Chong. Respiratory wheeze detection system. *Proc. 27th IEEE Eng. Med. Biol. Soc. (EMBS)*, pages 7553–7559, 2005.
- [3] M. L. Levy, S. Godfrey, C. S. Irving, A. Sheikh, W. Hanekom, Ambulatory Care Nurses, A. Bush, and P. Lachman. Wheeze detection: Recordings vs. assessment of physician and parent. *Journal of Asthma*, 41:845–853, 2004.
- [4] L. J. Hadjileontiadis. Lung sounds: An advanced signal processing perspective. In J. D. Enderle, editor, *Synthesis Lectures on Biomedical Engineering*. Morgan & Claypool Publishers, 2009.
- [5] L. J. Hadjileontiadis and S. M. Panas. Adaptive reduction of heart sounds from lung sounds using fourth-order statistics. *IEEE Trans. Biomed. Eng.*, 44(7):642–648, 1997.
- [6] S. Cortés, R. Jané, A. Torres, J. A. Fiz, and J. Morera. Detection and adaptive cancellation of heart sound interference in tracheal sounds. *Proc. 28th IEEE Eng. Med. Biol. Soc. (EMBS)*, pages 2860–2863, 2006.
- [7] D. Flores-Tapia, Z. Moussavi, and G. Thomas. Heart sound cancellation based on multiscale products and linear prediction. *IEEE Trans. Biomed. Eng.*, 54(2):234–243, 2007.
- [8] I. Yildirim, R. Ansari, and Z. Moussavi. Automated respiratory phase and onset detection using only chest sound signal. *Proc. 30th IEEE Eng. Med. Biol. Soc. (EMBS)*, pages 2578–2581, 2008.

-
- [9] P. Hult, T. Fjällbrant, B. Wranne, O. Engdahl, and P. Ask. An improved bioacoustic method for monitoring of respiration. *Tech. Health Care*, pages 323–332, 2004.
 - [10] A. Yadollahi and Z. Moussavi. A robust method for estimating respiratory flow using tracheal sounds entropy. *IEEE Trans. Biomed. Eng.*, 53(4):662–668, 2006.
 - [11] E. Conte, A. Vena, A. Federici, R. Giuliani, and J. P. Zbilut. A brief note on possible detection of physiological singularities in respiratory dynamics by recurrence quantification analysis of lung sounds. *J. Chaos Solitons Fractals*, 21:869–877, 2004.
 - [12] J. Gnitecki and Z. Moussavi. The fractality of lung sounds: A comparison of three waveform fractal dimension algorithms. *J. Chaos Solitons Fractals*, 26(4):1065–1072, 2005.
 - [13] C. Ahlstrom, A. Johansson, P. Hult, and P. Ask. Chaotic dynamics of respiratory sounds. *J. Chaos, Solitons Fractals*, 29:1054–1062, 2006.
 - [14] N. Gavriely and D. W. Cugell. *Breath Sounds Methodology*. CRC Press, 1995.
 - [15] L. Pesu, P. Helistö, E. Ademovic, J. C. Pesquet, A. Saarinen, and A. R. Sovijärvi. Classification of respiratory sounds based on wavelet packet decomposition and learning vector quantization. *Tech. Health Care*, 6(1):65–74, 1998.
 - [16] L. J. Hadjileontiadis. Discrimination analysis of discontinuous breath sounds using higher-order crossings. *Med. Biol. Eng. Comput.*, 41(4):445–455, 2003.
 - [17] L. J. Hadjileontiadis and I. T. Rekanos. Neural classification of lung sounds using wavelet coefficients. *Comput. Biol. Med.*, 34:523–537, 2004.
 - [18] M. Rossi, A. R. A. Sovijärvi, P. Piirilä, L. Vannuccini, F. Dalmasso, and J. Vanderschoot. Environmental and subject conditions and breathing manoeuvres for respiratory sound recordings. *Eur Respir Rev*, 10(77):611–615, 2000.
 - [19] F. Dalmay, M. T. Antonini, P. Marquet, and R. Menier. Acoustic properties of the normal chest. *Eur Respir J*, 8:1761–1769, 1995.
 - [20] V. A. McKusick, J. T. Jenkins, and G. N. Webb. The acoustic basis of the chest examination: Studies by means of sound spectrography. *Am Rev Tuberc*, 21:49–54, 1955.
 - [21] R. Loudon and R. L. H. Murphy. Lung sounds. *Am Rev Respir Dis*, 130:663–673, 1984.

-
- [22] M. J. Mussel. The need for standards in recording respiratory sounds. *Med. Biol. Eng. Comput.*, 30:129–139, 1992.
 - [23] S. S. Kraman. Lung sounds: Relative site of origin and comparative amplitudes in normal subjects. *Lung*, 161:57–64, 1983.
 - [24] S. S. Kraman. Does the vesicular lung sound come only from the lungs? *Am Rev Respir Dis*, 128:622–626, 1983.
 - [25] N. Gavriely, Y. Palti, G. Alroy, and J. B. Grotberg. Measurement and theory of wheezing breath sounds. *J Appl Physiol*, 57:481–492, 1984.
 - [26] W. W. Waring, R. C. Beckerman, and R. L. Hopkins. Continuous adventitious lung sounds: Site and method of production and significance. *Semin Respir Med*, 6:201–209, 1985.
 - [27] P. E. Krumpe, J. Hadley, and R. A. Marcum. Evaluation of bronchial air leaks by auscultation and phonopneumography. *Chest*, 85:777–781, 1984.
 - [28] M. J. Mussel, Y. Nakazono, Y. Miyamoto, S. Okabe, and T. Takishima. Distinguishing normal and abnormal tracheal breathing sounds by principal component analysis. *Jpn J Physiol*, 40:713–721, 1990.
 - [29] A. Cohen and D. Landsberg. Analysis and automatic classification of breath sounds. *IEEE Trans. Biomed. Eng.*, 31(9):585–590, 1984.
 - [30] http://images.main.uab.edu/healthsys/ei_0094.jpg.
 - [31] A. G. Tilkian and M. B. Conover. *Understanding Heart Sounds and Murmurs: with an Introduction to Lung Sounds*. Philadelphia, Audio CD., 2001.
 - [32] K. Ishizaka, J. C. French, and J. L. Flanagan. Direct determination of vocal tract wall impedance. *IEEE Trans. Acoust. Speech Signal Process.*, ASSP-23:370–373, 1975.
 - [33] J. J. Fredberg and A. Hoenig. Mechanical response of the lungs at high frequencies. *J. Biomech. Eng.*, 100:57–66, 1978.
 - [34] G. R. Wodicka, K. N. Stevens, H. L. Golub, E. G. Cravalho, and D. C. Shannon. A model of acoustic transmission in the respiratory system. *IEEE Trans. Biomed. Eng.*, 36(9):925–934, 1989.
 - [35] S. Lu, P. C. Doerschuk, and G. R. Wodicka. Parametric phase-delay estimation of sound transmitted through intact human lung. *Med. Biol. Eng. Comput.*, 33(3):293–298, 1995.

-
- [36] A. R. A. Sovijärvi, L. P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, and J. E. Earis. Characteristics of breath sounds and adventitious respiratory sounds. *Eur Respir Rev*, 10(77):591–596, 2000.
 - [37] S. Lehrer. *Understanding Lung Sounds*. Philadelphia, Audio CD., 2002.
 - [38] H. Pasterkamp, R. Fenton, A. Tal, and V. Chernick. Tracheal vs lung sounds in acute asthma. *Am Rev Respir Dis*, 129:256A, 1984.
 - [39] Z. Moussavi. Fundamentals of respiratory sounds and analysis. In J. D. Enderle, editor, *Synthesis Lectures on Biomedical Engineering*. Morgan & Claypool Publishers, 2007.
 - [40] A. Yadollahi and Z. Moussavi. A robust method for heart sounds localization using lung sounds entropy. *IEEE Trans. Biomed. Eng.*, 53(3):497–502, 2006.
 - [41] G. Sierra, V. Telfort, B. Popov, L. G. Durand, R. Agarwal, and V. Lanzo. Monitoring respiratory rate based on tracheal sounds. First experiences. *Proc. 26th IEEE Eng. Med. Biol. Soc. (EMBS)*, 1:317–320, 2004.
 - [42] J. Chuah and Z. Moussavi. Automated respiratory phase detection by acoustical means. *University of Manitoba, Dept. of Electrical and Computer Engineering*, 2004.
 - [43] Y. L. Yap and Z. Moussavi. Respiratory onset detection using variance fractal dimension. *Proc. 23th IEEE Eng. Med. Biol. Soc. (EMBS)*, 2:1554–1556, 2001.
 - [44] S. L. Cam, Ch. Collet, and F. Salzenstein. Acoustical respiratory signal analysis and phase detection. *Proc. IEEE Acoust. Speech Signal Process. (ICASSP'08)*, pages 3629–3632, 2008.
 - [45] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.
 - [46] W. Pieczynski. Modeling non stationary hidden semi-markov chains with triplet markov chains and theory of evidence. *Proc. IEEE Workshop Statistical Signal Process.*, pages 727–732, 2005.
 - [47] J. Gnitecki, Z. Moussavi, and H. Pasterkamp. Classification of lung sounds during bronchial provocation using waveform fractal dimensions. *Proc. 26th IEEE Eng. Med. Biol. Soc. (EMBS)*, 2:3844–3847, 2004.
 - [48] J. C. Chien, H. D. Wu, F. C. Chong, and C. I. Li. Wheeze detection using cepstral analysis in gaussian mixture models. *Proc. 29th IEEE Eng. Med. Biol. Soc. (EMBS)*, pages 3168–3171, 2007.

-
- [49] K. S. Baydar, A. Ertuzun, and Y. P. Kahya. Analysis and classification of respiratory sounds by signal coherence method. *Proc. 25th IEEE Eng. Med. Biol. Soc. (EMBS)*, 3:2950–2953, 2003.
 - [50] S. Alsmadi and Y. P. Kahya. Design of a DSP-based instrument for real-time classification of pulmonary sounds. *Comput. Bio. Med.*, 38:53–61, 2008.
 - [51] H. Pasterkamp, S. S. Kraman, and G. R. Wodicka. Respiratory sounds: Advances beyond the stethoscope. *Am. J. Respir. Crit. Care Med.*, 156:974–987, 1997.
 - [52] L. Vannuccini, J. E. Earis, P. Helistö, B. M. G. Cheetham, M. Rossi, A. R. A Sovijärvi, and J. Vanderschoot. Capturing and preprocessing of respiratory sounds. *Eur Respir Rev*, 10(77):616–620, 2000.
 - [53] P. H. Quanjer. Standardization of lung function tests – 1993 update. Report of a working party for the european community for steel and coal. *Eur Respir J*, 6(Suppl 16):5–40, 1993.
 - [54] F. Jin and F. Sattar. Enhancement of recorded respiratory sound using signal processing techniques. In A. Cartelli and M. Palma, editors, *Encyclopedia of Information Communication Technology*. Idea Group Inc., USA, 2008.
 - [55] B. widrow, J. R. Jr. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, Jr. Eugene Dong, and R. C. Goodlin. Adaptive noise cancelling: Principles and applications. *Proc. IEEE*, 63(12):1692–1716, 1975.
 - [56] H. Pasterkamp, R. Fenton, A. Tal, and V. Chernick. Interference of cardiovascular sounds with phonopneumography in children. *Am Rev Respir Dis*, 131(1):61–64, 1985.
 - [57] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Information Theory*, 41(3):613–627, 1995.
 - [58] K. Iyer, P. A. Pamamoorthy, H. Fan, and Y. Ploysongsang. Reduction of heart sounds from lung sounds by adaptive filtering. *IEEE Trans. Biomed. Eng.*, 33:1141–1148, 1986.
 - [59] M. Kompis and E. Russi. Adaptive heart-noise reduction of lung sounds recorded by a single microphone. *Proc. 14th IEEE Eng. Med. Biol. Soc. (EMBS)*, 2:691–692, 1992.
 - [60] M. T. Pourazad, Z. Moussavi, and G. Thomas. Heart sound cancellation from lung sound recordings using time-frequency filtering. *Med. Biol. Eng. Comput.*, 44(3):216–225, 2006.

-
- [61] C. Ahlstrom, O. Liljefeldt, P. Hult, and P. Ask. Heart sound cancellation from lung sound recordings using recurrence time statistics and nonlinear prediction. *IEEE Signal Process. Letters*, 12(12):812–815, 2005.
 - [62] J. Gnitecki and Z. Moussavi. Variance fractal dimension trajectory as a tool for heart sound localization in lung sounds recordings. *Proc. 25th IEEE Eng. Med. Biol. Soc. (EMBS)*, 3(3):2420–2423, 2003.
 - [63] J. Gnitecki and Z. Moussavi. Separating heart sounds from lung sounds. *IEEE Eng. Med. Bio. Mag.*, 26(1):20–29, 2007.
 - [64] R. L. Wilkins, J. E. Hodgkin, and B. Lopez. *Fundamentals of Lung and Heart Sounds*. Mosby, Audio CD., 2004.
 - [65] M. Kompis, H. Pasterkamp, Y. Motai, and G. R. Wodicka. Spatial representation of thoracic sounds. *Proc. 20th IEEE Eng. Med. Biol. Soc. (EMBS)*, 20(3):1661–1664, 1998.
 - [66] C. L. Nikias and P. A. Petropulu. *Higher-Order Spectral Analysis: A Nonlinear Signal Processing Framework*. Prentice-Hall, 1993.
 - [67] H. Liang, S. Lukkarinen, and I. Hartimo. Heart sound segmentation algorithm based on heart sound envelopogram. *Computers in Cardiology*, pages 105–108, 1997.
 - [68] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Process. Mag.*, 8(4):14–38, 1991.
 - [69] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, 1991.
 - [70] A. J. Van Der Veen, E. F. Deprettere, and A. L. Swindlehurst. Sub-space based signal analysis using singular value decomposition. *Proc. IEEE*, 81(9):1277–1307, 1993.
 - [71] S. Charleston, M. R. Azimi-Sadjadi, and R. Gonzalez-Camarena. Interference cancellation in respiratory sounds via a multiresolution joint time-delay and signal-estimation scheme. *IEEE Trans. Biomed. Eng.*, 44(10):1006–1019, 1997.
 - [72] L. J. Hadjileontiadis and S. M. Panas. A wavelet-based reduction of heart sound noise from lung sounds. *Int J Med Inf*, 52(1-3):183–190, 1998.
 - [73] I. Hossain and Z. Moussavi. An overview of heart-noise reduction of lung sound using wavelet transform based filter. *Proc. 25th IEEE Eng. Med. Biol. Soc. (EMBS)*, 1:458–461, 2003.

-
- [74] S. Charleston and M. R. Azimi-Sadjadi. Reduced order kalman filtering for the enhancement of respiratory sounds. *IEEE Trans. Biomed. Eng.*, 43(4):421–424, 1996.
 - [75] L. Yip and Y. T. Zhang. Reduction of heart sounds from lung sound recordings by automated gain control and adaptive filtering techniques. *Proc. 23th IEEE Eng. Med. Biol. Soc. (EMBS)*, 3(3):2154–2156, 2001.
 - [76] W.G. Gardner. *3-D Audio Using Loudspeakers*. Kluwer Academic Publishers, 1998.
 - [77] H. S. Malvar. *Signal Processing with Lapped Transforms*. Artech House, Inc. Norwood, MA, USA, 1992.
 - [78] P. P. Vaidyanathan. *Multirate Systems and Filter Banks*. NJ: Prentice Hall, 1993.
 - [79] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, 1991.
 - [80] J. Gnitecki, Z. Moussavi, and H. Pasterkamp. Recursive least square adaptive noise cancellation filtering for heart sound in lung sounds recording. *Proc. 25th IEEE Eng. Med. Biol. Soc. (EMBS)*, 3(3):2416–2419, 2003.
 - [81] M. T. Pourazad, Z. Moussavi, and G. Thomas. Heart sound cancellation from lung sound recordings using adaptive threshold and 2D interpolation in time-frequency domain. *Proc. 25th IEEE Eng. Med. Biol. Soc. (EMBS)*, 3:2586–2589, 2003.
 - [82] A. L. Klein, L. K. Hatle, D. J. Burstow, C. P. Taliercio, J. B. Seward, R. A. Kyle, K. R. Bailey, M. A. Gertz, and A. J. Tajik. Comprehensive Doppler assessment of right ventricular diastolic function in cardiac amyloidosis. *J Am Coll Cardiol*, 15:99–108, 1990.
 - [83] S. Henneberg, B. Hök, L. Wiklund, and G. Sjödin. Remote auscultatory patient monitoring during magnetic resonance imaging. *J Clin Monit*, 8(1):37–43, 1992.
 - [84] N. Meslier, G. Charbonneau, and J. L. Racineux. Wheezes. *Eur Respir J*, 8:1942–1948, 1995.
 - [85] U. R. Abeyratne, A. S. Karunajeewa, and C. Hukins. Mixed-phase modeling in snore sound analysis. *Med. Biol. Eng. Comput.*, 45(8):791–806, 2007.
 - [86] A. Kulkas, E. Huupponen, J. Virkkala, M. Tenhunen, A. Saastamoinen, E. Rauhala, and S. L. Himanen. New tracheal sound feature for apnoea analysis. *Med. Biol. Eng. Comput.*, 47(4):405–412, 2009.

-
- [87] S. A. Taplidou and L. J. Hadjileontiadis. Nonlinear analysis of wheezes using wavelet bicoherence. *Comput. Biol. Med.*, 37:563–570, 2007.
 - [88] S. Cortés, R. Jané, J. A. Fiz, and J. Morera. Monitoring of wheeze duration during spontaneous respiration in asthmatic patients. *Proc. 27th IEEE Eng. Med. Biol. Soc. (EMBS)*, 2005.
 - [89] J. Ashkanazi, P. Silverberg, R. Foster, A. Hyman, J. Milic-Emili, and J. Kinney. Effects of respiratory apparatus on breathing pattern. *J Appl Physiol*, 48:577–580, 1980.
 - [90] G. Souffet, G. Charbonneau, M. Polit, P. Attal, A. Denjean, P. Escourrou, and C. Gaultier. Interaction between tracheal sound and flow rare: A comparison of some different flow evaluations from lung sounds. *IEEE Trans. Biomed. Eng.*, (4):384–391, 1990.
 - [91] M. Golabbakhsh. Tracheal breath sound relationship with respiratory flow: Modeling, the effect of age and airflow estimation. *M. Sc. Thesis, University of Manitoba*, 2004.
 - [92] R. C. Nongpiur. Impulse noise removal in speech using wavelets. *Proc. 20th IEEE Eng. Med. Biol. Soc. (EMBS)*, pages 1593–1596, 2008.
 - [93] S. Mukhopadhyay and G. C. Ray. A new interpretation of nonlinear energy operator and its efficacy in spike detection. *IEEE Trans. Biomed. Eng.*, 45(2):180–187, 1998.
 - [94] A. V. Oppenheim and R.W. Schaffer. *Discrete-Time Signal Processing*. Prentice-Hall, 1999.
 - [95] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.
 - [96] M. Schatzman. *Numerical Analysis: A Mathematical Introduction*. Clarendon Press, Oxford, 2002.
 - [97] A. R. A. Sovijärvi, J. Vanderschoot, and J. R. Eavis. Standardization of computerized respiratory sound analysis. *Eur Respir Rev*, 10(77):585–649, 2000.
 - [98] Y. Shao and C. H. Chang. A generalized time–frequency subtraction method for robust speech enhancement based on wavelet filter banks modeling of human auditory system. *IEEE Trans. Sys. Man Cyber. Part B*, 37(4):877–889, 2007.
 - [99] Y. Huang, J. Benesty, and J. D. Chen. *Acoustic MIMO Signal Processing*. Springer, 2006.

-
- [100] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol*, 278(6):H2039–2049, 2000.
- [101] D. A. Coley. *An Introduction to Genetic Algorithms for Scientists and Engineers*. World Scientific, 2001.
- [102] K. S. Tang, K. F. Man, S. Kwong, and Q. He. Genetic algorithms and their applications. *IEEE Signal Process. Mag.*, 13(6):22–37, 1996.
- [103] Q. Huang and B. Dom. Quantitative methods of evaluating image segmentation. *Proc. IEEE Int. Conf. Image Process.*, 3:53–56, 1995.
- [104] A. Chipperfield, P. Fleming, H. Pohlheim, and C. Fonseca. Genetic algorithm toolbox. *Department of Automatic Control and Systems Engineering, University of Sheffield*, 1995.
- [105] E. R. Larsen. *Audio Bandwidth Extension : Application of Psychoacoustics, Signal Processing and Loudspeaker Design*. John Wiley & Sons, 2004.
- [106] S. Marshall and G. L. Sicuranza. *Advances in Nonlinear Signal and Image Processing*. Hindawi Publishing Corporation, 2006.
- [107] M. Berouti, R. Schwartz, and J. Makhoul. Enhancement of speech corrupted by acoustic noise. *Proc. IEEE Acoust. Speech Signal Process. (ICASSP’79)*, 4(208–211), 1979.
- [108] S. M. Pincus. Approximate entropy (ApEn) as a complexity measure. *Chaos*, 5:110–117, 1995.
- [109] X. N. Chen, I. C. Solomon, and K. H. Chon. Comparison of the use of approximate entropy and sample entropy: Applications to neural respiratory signal. *Proc. 27th IEEE Eng. Med. Biol. Soc. (EMBS)*, pages 4212–4215, 2005.
- [110] D. E. Goldberg. *Genetic Algorithm in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
- [111] J. G. Qi, G. R. Burns, and D. K. Harrison. The application of parallel multipopulation genetic algorithms to dynamic job-shop scheduling. *Int. J. Adv. Manufacturing Tech.*, 16(8):609–615, 2000.
- [112] J. S. Erkelens and R. Heusdens. Tracking of nonstationary noise based on data-driven recursive noise power estimation. *IEEE Trans. Audio Speech Language Process.*, 16(6):1112–1123, 2008.

-
- [113] S. Sello, S. K. Strambi, G. D. Michele, and N. Ambrosino. Respiratory sound analysis in healthy and pathological subjects: A wavelet approach. *Biomed. Signal Process. Control*, 3(3).
 - [114] S. V. Vaseghi. *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons Ltd, 2000.
 - [115] S. S. Kraman. Lung sounds: An introduction to the interpretation of the auscultatory finding. *American College of Chest Physicians, Northbrook, IL*, pages 19–21, 1993.
 - [116] R. L. Murphy Jr, E. A. Del Bono, and F. Davidson. Validation of an automatic crackle (rale) counter. *Am Rev Resir Dis*, 140(4):1017–1020, 1989.
 - [117] M. Munakata, H. Ukita, I. Doi, Y. Ohtsuka, Y. Masaki, Y. Homma, and Y. Kawakami. Spectral and waveform characteristics of fine and coarse crackles. *British Med J*, 46(9):651–657, 1991.
 - [118] Y. Shabtai-Musih, J. B. Grotberg, and N. Gavriely. Spectral Content of Forced Expiratory Wheezes During Air, He, and SF6 Breathing in Normal Humans. *Journal of Applied Physiology*, 72(2):629–635, 1992.
 - [119] Y. Qiu, A. R. Whittaker, M Lucas, and K. Anderson. Automatic wheeze detection based on auditory modelling. *Proc. I MechE Part H J. Eng. Med.*, 51:219–227, 2005.
 - [120] Wheeze detection based on time-frequency analysis of breath sounds.
 - [121] C. Ahlstrom, P. Hult, and P. Ask. Wheeze analysis and detection with nonlinear phase space embedding. *Proc. 13th Nordic Baltic Biomed. Eng. Med. Phy. (NBC05)*, pages 305–306, 2005.
 - [122] S. S. Kraman, H. Pasterkamp, M. Kompis, M. Takase, and G. R. Wodicka. Effects of breathing pathways on tracheal sound spectral features. *Respir Physiol*, 111:295–300, 1998.
 - [123] L. J. Hadjileontiadis and I. T. Rekanos. Detection of explosive lung and bowel sounds by means of fractal dimension. *IEEE Signal Process. Letters*, 10:311–314, 2003.
 - [124] S. M. Yamashiro. Non-linear dynamics of human periodic breathing and implications for sleep apnea therapy. *Med. Biol. Eng. Comput.*, 45(4):345–356, 2007.
 - [125] A. Swami and J. M. Mendel. Cumulant-based approach to the harmonic retrieval and related problems. *IEEE Trans. Signal Process.*, 39(5):1099–1109, 1991.

-
- [126] D. M. Tsai and R. H. Yang. An eigenvalue-based similarity measure and its application in defect detection. *Image and Vision Computing*, 23:1094–1101, 2005.
 - [127] E. Micheli-Tzanakou. *Supervised and Unsupervised Pattern Recognition: Feature Extraction and Computational Intelligence*. CRC Press LLC, 2000.
 - [128] R. Folland, E. Hines, R. Dutta, P. Boilot, and D. Morgan. Comparison of neural network predictors in the classification of tracheal-bronchial breath sounds by respiratory auscultation. *Artificial Intelligence Med.*, 31(3):211–220, 2004.
 - [129] Y. P. Kahya, E. Bayatli, M. Yeginer, K. Ciftci, and G. Kilinc. Comparison of different feature sets for respiratory sounds classifiers. *Proc. 25th IEEE Eng. Med. Biol. Soc. (EMBS)*, pages 2853–2856, 2003.
 - [130] M. Bahoura and C. Pelletier. New parameters for respiratory sound classification. *Proc. IEEE Conf. Electrical Compt. Eng.*, 3:1457–1460, 2003.
 - [131] F. Jin, F. Sattar, and D.Y.T. Goh. Automatic wheeze detection using histograms of sample entropy. pages 1890–1893, 2008.
 - [132] S. Qian and D. Chen. *Joint Time-Frequency Analysis - Methods and Applications*. Prentice-Hall, 1996.
 - [133] D. J. Nelson and D. C. Smith. A linear model for the distribution of signals in time and frequency with a historical perspective of conventional time frequency distribution. *Digital Signal Processing*, 18:81–102, 2000.
 - [134] J. He, A. H. Tan, C. L. Tan, and S. Y. Sung. On quantitative evaluation of clustering systems. In W. Wu, H. Xiong, and S. Shekhar, editors, *Clustering and Information Retrieval*, pages 105–133. Kluwer Academic Publishers, 2003.
 - [135] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.
 - [136] M. Halkidi, M. Vazirgiannis, and I. Batistakis. Quality scheme assessment in the clustering process. *Proc. 4th European Conf. Principles of Data Mining and Knowledge Discovery (PKDD)*, pages 265–276, 2000.
 - [137] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.

- [138] M. J. Islam, Q. M. J. Wu, M. Ahmadi, and M. A. Sid-Ahmed. Investigating the performance of naive-bayes classifiers and k-nearest neighbor classifiers. *Proc. Int. Conf. Convergence Inf. Tech.*, pages 1541–1546, 2007.
- [139] G. W. Milligan and M. C. Cooper. A study of standardization of variables in cluster analysis. *Journal of Classification*, 5(2):181–204, 1988.
- [140] T. S. Lim, W. Y. Loh, and Y. S. Shih. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning*, 40(3):203–228, 2000.
- [141] *Neural Network Toolbox (6) Users' Guide*. The Math Works, Natick, MA, 2009.
- [142] J. Laroche and M. Dolson. Improve phase vocoder time-scale modification of audio. *IEEE Trans. Speech Audio Process.*, 7(3):323–332, 1999.
- [143] D. Arfib, F. Keiler, and U. Zöler. *DAFX - Digital Audio Effects*. John Wiley Publisher, 2002.
- [144] S. Qian, Y. Rao, and D. Chen. A fast gabor spectrogram. *Proc. IEEE Acoustics Speech Signal Process. (ICASSP'00)*, 2:653–656, 2000.