NON-INVASIVE PATHOLOGICAL VOICE CLASSIFICATIONS USING LINEAR AND NON-LINEAR CLASSIFIERS

HARIHARAN MUTHUSAMY

UNIVERSITI MALAYSIA PERLIS
2010
NON-INVASIVE PATHOLOGICAL VOICE CLASSIFICATIONS USING LINEAR AND NON-LINEAR CLASSIFIERS

by

Hariharan Muthusamy
(0640610093)

A thesis submitted
In fulfillment of the requirements for the degree of Doctor of Philosophy (Mechatronic Engineering)

School of Mechatronic Engineering
UNIVERSITI MALAYSIA PERLIS

2010
ACKNOWLEDGEMENT

I would like to thank the Acoustic Applications Research Cluster, School of Mechatronic Engineering and University Malaysia Perlis (UniMAP) for providing a highly supportive research environment.

I would like to sincerely thank my supervisor Prof. Dr. Sazali Yaacob, Deputy Vice-chancellor (Academic and International), Universiti Malaysia Perlis for his guidance, encouragement, patience, inspiration and constructive feedback throughout the research and preparation of this thesis. His support towards my participation in international conferences, and commitment to an environment with adequate equipment and facilities are deeply appreciated. This thesis would not have been possible at all without his encouragement and support.

I would like to extend my sincere appreciation to my co-supervisor Assoc. Prof. Dr. Paulraj M P for his assistance and discussion throughout this research work. My special thanks to UniMAP for providing me a financial support through Graduate Assistantship (GA).

I would like to thank Dr. Indrani Christina Das, Dr. Mohd. Zambri B. Ibrahim, and Cik Fairus bt Mukthar from Hospital Tuanku Fauziah, Kangar, Perlis for allowing me to collect data from real patients.
I would also like to express my sincere appreciation to all members of Acoustic Applications Research Cluster as well as those who have contributed indirectly towards the completion of this research.

I sincerely acknowledge the FRGS fund (No: 9003-00043) received from the Ministry of Science, Technology and Innovation, Malaysia, through UniMAP.

Last but not the least, I must thank my parents, my uncle and my brother. Their love, encouragement, and constant support enabled me to pursue and achieve my goals.
## TABLE OF CONTENTS

ACKNOWLEDGEMENT .............................................. ii
TABLE OF CONTENTS ................................................ iv
LIST OF TABLES .................................................... x
LIST OF FIGURES .................................................... xv
LIST OF ABBREVIATIONS ........................................ xviii
ABSTRAK .............................................................. xxi
ABSTRACT .............................................................. xxii

1 INTRODUCTION
  1.1 Preamble ..................................................... 1
  1.2 Problem Statement and Significance of the Study ............ 4
  1.3 Research Objectives ......................................... 6
  1.4 Thesis Organization ......................................... 11

2 LITERATURE REVIEW
  2.1 Voice Disorders .............................................. 14
  2.2 Types of Voice Disorders .................................... 15
  2.3 Prevalence of Voice Disorders ................................ 16
  2.4 Review of Previous Works ................................... 18
      2.4.1 Long-time Acoustical Parameters ...................... 18
      2.4.2 Short-time Acoustical Parameters .................... 25
      2.4.3 Time-Frequency Analysis .............................. 27
      2.4.4 Application of Non-Linear Analysis ................. 28
2.4.5 Detection of Specific Disorders 31
2.5 Observations from Previous Works 33
2.6 Research Contributions 35
2.7 Voice Disorders Database used in this Research 38
  2.7.1 MEEI Voice Disorders Database 38
  2.7.2 MAPACI Speech Pathology Database 39
  2.7.3 Dataset –III (collected at Hospital Tuanku Fauziah, Kangar, Perlis) 41
2.8 Summary and Conclusion 42

3 ACOUSTIC FEATURE EXTRACTION METHODS FOR VOICE DISORDERS

3.1 Pre-processing of the Speech Signals 43
3.2 Time-Domain Energy based Features 44
3.3 Feature Extraction based on MFCC and SVD 49
  3.3.1 Mel Frequency Cepstral Coefficients 49
  3.3.2 Singular Value Decomposition 51
  3.3.3 Parameterization using Mel-Frequency Cepstral Coefficients and SVD 52
  3.3.4 Summarization of MFCC features using SVD 53
3.4 Wavelet Packet Transform and Entropy based Features 54
  3.4.1 Wavelet 55
3.4.2 Wavelet Transform 56
3.4.3 Continuous Wavelet Transform 56
3.4.4 Discrete Wavelet Transform 57
3.4.5 Wavelet Packets 58
3.4.6 Parameterization using Wavelet Packet Transform and Entropy Measures 59

3.5 Effectiveness of the Proposed Features 62
3.5.1 Time-Domain Energy Based Features 62
3.5.2 MFCC and SVD Based Features 65
3.5.3 Wavelet Packet and Entropy Based Features 69

3.6 Summary and Conclusion 74

4 DEVELOPMENT OF CLASSIFIERS FOR THE CLASSIFICATION OF VOICE DISORDERS

4.1 Linear Classifier 76
4.1.1 Linear Discriminant Analysis based Classifier 77

4.2 Non-Linear Classifiers 79
4.2.1 k-Nearest Neighbor Classifier 80
4.2.2 Multilayer Perceptron Classifier 81
4.2.3 Probabilistic and General Regression Neural Networks 84
4.2.4 Probabilistic Neural Network Classifier 84
4.2.5 General Regression Neural Network Classifier 86
5 RESULTS AND DISCUSSIONS OF NORMAL AND PATHOLOGICAL VOICE CLASSIFICATION

5.1 Results for the Speech Signals in MEEI Voice Disorders Database under Controlled Environment
   5.1.1 Results of k-NN Classifier
   5.1.2 Results of LDA based Classifier
   5.1.3 Results of MLP Classifier
   5.1.4 Results of PNN Classifier
   5.1.5 Results of GRNN Classifier

5.2 Results for the Speech Signals in MEEI Voice Disorders Database under Noisy Environment (SNR=30dB)
   5.1.1 Results of k-NN Classifier (SNR=30dB)
   5.1.2 Results of LDA based Classifier (SNR=30dB)
   5.1.3 Results of MLP Classifier (SNR=30dB)
   5.1.4 Results of PNN Classifier (SNR=30dB)
   5.1.5 Results of GRNN Classifier (SNR=30dB)

5.3 Development of Graphical User Interface for An Intelligent Voice Disorders Diagnosing System
   5.3.1 Software Description

5.4 Comparison of Results with Previous Works (Two Class Problem – Normal or Pathological)

5.5 Summary and Conclusion
6 DETECTION OF SPECIFIC DISORDERS

6.1 Detection of Three Specific Disorders under Controlled Environment

6.1.1 k-NN Classifier 123
6.1.2 LDA based Classifier 126
6.1.3 MLP Classifier 127
6.1.4 PNN Classifier 130
6.1.5 GRNN Classifier 132

6.2 Detection of Three Specific Disorders under Noisy Environment (SNR=30dB)

6.2.1 k-NN Classifier (SNR=30dB) 135
6.2.2 LDA based Classifier (SNR=30dB) 137
6.2.3 MLP Classifier (SNR=30dB) 138
6.2.4 PNN Classifier (SNR=30dB) 140
6.2.5 GRNN Classifier (SNR=30dB) 142

6.3 Comparison of Results with Previous Works (Detection of Specific Disorders)

6.4 Summary and Conclusion 151

7 CONCLUSION

7.1 Thesis summary 153
7.2 Contributions 158
7.3 Suggestions for Future Research 159
REFERENCES

APPENDICES

Appendix A  Sensitivity and Specificity Results for MEEI Voice Disorders Database (Two Class Problem-Normal or Pathological) 174

Appendix B  Sensitivity and Specificity Results for MEEI Voice Disorders Database (Detection of Specific Disorders) 182

Appendix C  Sensitivity and Specificity Results for MAPACI Speech Pathology Database 206

Appendix D  Sensitivity and Specificity Results for Dataset-III (collected at Hospital Tuanku Fauziah, Kangar, Perlis) 214

LIST OF PUBLICATIONS 222

ACHIEVEMENTS 224
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Summary of Significant Research Works using Long-time Acoustical Parameters</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Summary of Significant Research Works using Short-time Acoustical Parameters</td>
<td>26</td>
</tr>
<tr>
<td>2.3</td>
<td>Summary of Significant Research Works using Time-Frequency Analysis</td>
<td>28</td>
</tr>
<tr>
<td>2.4</td>
<td>Summary of Significant Research Works using Non-Linear Analysis</td>
<td>30</td>
</tr>
<tr>
<td>2.5</td>
<td>Summary of Significant Research Works Done for the Detection of Specific Disorders</td>
<td>31</td>
</tr>
<tr>
<td>2.6</td>
<td>Details of the Investigations Made on the Specific Disorders</td>
<td>36</td>
</tr>
<tr>
<td>2.7</td>
<td>Pathological Female Patients’ Information Sheet</td>
<td>39</td>
</tr>
<tr>
<td>2.8</td>
<td>Pathological Male Patients’ Information Sheet</td>
<td>40</td>
</tr>
<tr>
<td>2.9</td>
<td>Patient’s Information Sheet</td>
<td>41</td>
</tr>
<tr>
<td>3.1</td>
<td>Filters and Their Corresponding Frequency Bands Considered in MFCC Feature Extraction</td>
<td>53</td>
</tr>
<tr>
<td>3.2</td>
<td>Filters and Their Corresponding Frequency Bands Achieved by Wavelet Packet Decomposition</td>
<td>60</td>
</tr>
<tr>
<td>3.3</td>
<td>Difference of Mean of the Time Domain Energy Variation based Features of Normal and Pathological Speech Signals in MEEI Database</td>
<td>63</td>
</tr>
<tr>
<td>3.4</td>
<td>Difference of Mean of the Time Domain Energy Variation based Features of Normal and Pathological (AP Squeezing, Vocal Fold Edema, and Vocal Fold Paralysis) Speech Signals in MEEI Database for 50% Overlap</td>
<td>64</td>
</tr>
<tr>
<td>3.5</td>
<td>Difference of Mean of the Time Domain Energy Variation based Features of Normal and Pathological (AP Squeezing, Vocal Fold Edema, and Vocal Fold Paralysis) Speech Signals in MEEI Database for 75% Overlap</td>
<td>65</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.6</td>
<td>Difference of Mean of the MFCC and SVD based Features of Normal and Pathological Speech Signals in MEEI Database</td>
<td>66</td>
</tr>
<tr>
<td>3.7</td>
<td>Difference of Mean of the MFCC and SVD based Features of Normal and Pathological (AP Squeezing, Vocal Fold Edema, and Vocal Fold Paralysis) Speech Signals in MEEI Database for 50% Overlap</td>
<td>67</td>
</tr>
<tr>
<td>3.8</td>
<td>Difference of Mean of the MFCC and SVD based Features of Normal and Pathological (AP Squeezing, Vocal Fold Edema, and Vocal Fold Paralysis) Speech Signals in MEEI Database for 75% Overlap</td>
<td>68</td>
</tr>
<tr>
<td>3.9</td>
<td>Difference of Mean of the Wavelet Packet and Entropy based Features of Normal and Pathological Speech signals in MEEI Database</td>
<td>70</td>
</tr>
<tr>
<td>3.10</td>
<td>Difference of Mean of the Wavelet Packet and Entropy based Features of Normal and Pathological (AP Squeezing, Vocal Fold Edema, and Vocal Fold Paralysis) Speech Signals in MEEI Database for db4</td>
<td>71</td>
</tr>
<tr>
<td>3.11</td>
<td>Difference of Mean of the Wavelet Packet and Entropy based Features of Normal and Pathological (AP Squeezing, Vocal Fold Edema, and Vocal Fold Paralysis) Speech Signals in MEEI Database for db10</td>
<td>72</td>
</tr>
<tr>
<td>3.12</td>
<td>Difference of Mean of the Wavelet Packet and Entropy based Features of Normal and Pathological (AP Squeezing, Vocal Fold Edema, and Vocal Fold Paralysis) Speech Signals in MEEI Database for db20</td>
<td>73</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.1</td>
<td>Confusion Matrix</td>
<td>90</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.1</td>
<td>Results of the KNN Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database</td>
<td>94</td>
</tr>
<tr>
<td>5.2</td>
<td>Results of the LDA Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database</td>
<td>95</td>
</tr>
<tr>
<td>5.3(a)</td>
<td>Results of the MLP Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database (Overall Accuracy)</td>
<td>97</td>
</tr>
</tbody>
</table>
5.3(b) Results of the MLP Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database (Number of Epoch)

97

5.3(c) Results of the MLP Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database (Average Training Time)

97

5.4 Results of the PNN Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database

99

5.5 Results of the GRNN Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database

101

5.6 Results of the KNN Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database with 30dB Noise

103

5.7 Results of the LDA Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database with 30dB noise

104

5.8(a) Results of the MLP Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database with 30dB noise (Overall Accuracy)

105

5.8(b) Results of the MLP Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database with 30dB noise (Number of Epoch)

105

5.8(c) Results of the MLP Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database with 30dB Noise (Training Time)

105

5.9 Results of the PNN Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database with 30dB Noise

107

5.10 Results of the GRNN Classifier for the Classification of Normal and Pathological Voices in MEEI Voice Disorder Database with 30dB Noise

108

5.11 Comparison of Results with Previous Research Works (Two Class Problem – Normal or Pathological)

114
6.1 Results of the KNN Classifier for the Classification of Normal and Pathological Voices (AP Squeezing) 124
6.2 Results of the KNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Edema) 125
6.3 Results of the KNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Paralysis) 125
6.4 Results of the LDA Classifier for the Classification of Normal and Pathological Voices (AP Squeezing, Vocal Fold Edema, Vocal Fold Paralysis) 126
6.5 Results of the MLP Classifier for the Classification of Normal and Pathological Voices (AP Squeezing) 128
6.6 Results of the MLP Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Edema) 129
6.7 Results of the MLP Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Paralysis) 129
6.8 Results of the PNN Classifier for the Classification of Normal and Pathological Voices (AP Squeezing) 131
6.9 Results of the PNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Edema) 131
6.10 Results of the PNN Classifier for the Classification of Normal and Pathological voices (Vocal Fold Paralysis) 132
6.11 Results of the GRNN Classifier for the Classification of Normal and Pathological Voices (AP Squeezing) 134
6.12 Results of the GRNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Edema) 134
6.13 Results of the GRNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Paralysis) 135
6.14 Results of the KNN Classifier for the Classification of Normal and Pathological Voices (AP Squeezing) with 30dB Noise 136
6.15 Results of the KNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Edema) with 30dB Noise 136
6.16 Results of the KNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Paralysis) with 30dB Noise 137
6.17 Results of the LDA Classifier for the Classification of Normal and Pathological Voices (AP Squeezing, Vocal Fold Edema, Vocal Fold Paralysis) with 30dB Noise 138
6.18 Results of the MLP Classifier for the Classification of Normal and Pathological Voices (AP Squeezing) with 30dB Noise 139
6.19 Results of the MLP Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Edema) with 30dB Noise 139
6.20 Results of the MLP Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Paralysis) with 30dB Noise 139
6.21 Results of the PNN Classifier for the Classification of Normal and Pathological Voices (AP Squeezing) with 30dB Noise 140
6.22 Results of the PNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Edema) with 30dB Noise 141
6.23 Results of the PNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Paralysis) with 30dB Noise 141
6.24 Results of the GRNN Classifier for the Classification of Normal and Pathological Voices (AP Squeezing) with 30dB Noise 142
6.25 Results of the GRNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Edema) with 30dB Noise 143
6.26 Results of the GRNN Classifier for the Classification of Normal and Pathological Voices (Vocal Fold Paralysis) with 30dB Noise 143
6.27 Comparison of Results with the Previous Works (Detection of Specific Disorders) 144
LIST OF FIGURES

Figure
2.1(a)  Cross Section of Mouth and Throat  15
2.1(b)  Vocal Cords Diagram  15
2.2  Classification of Voice Disorders  16
2.3  Overall Methodology  37
3.1(a)  Time-Domain Energy Plot of a Normal Speech Signal  45
3.1(b)  Time-Domain Energy Plot of a Pathological Speech Signal  46
3.2(a)  Illustration of the Energy Peaks  47
3.2(b)  Area Enclosed by the Two Energy Peaks of a Speech Signal in One Short-Time Window  47
3.2(c)  Basic Block Diagram of MFCC Feature Extraction  50
3.3  An Example of Mel-Spaced Filterbank  50
3.4  Feature Extraction using MFCC and SVD  52
3.5  Basic and Multilevel Wavelet Decomposition  58
3.6  Feature Extraction using Wavelet Packet Transform and Shannon Entropy Measures  62
4.1  Multilayer Neural Network with One Hidden Layer  82
4.2  Architecture of General Regression Neural Network  88
5.1  GUI Layout  111
5.2  Open Dialog Box to Load a Speech File  111
5.3  Software Results after Loading a Pathological Speech File  112
5.4  Software Results after Loading a Normal Speech File  112
5.5  Comparison of Positive Predictivity Results of the Five Classifiers under Clean and Noisy Environment (MEEI Database)  115
5.6 Comparison of Overall Accuracy Results of the Five Classifiers under Clean and Noisy Environment (MEEI Database) 115
5.7 Comparison of Training Time(s) Results of the Five Classifiers under Clean and Noisy Environment (MEEI Database) 116
5.8 Comparison of Overall Accuracy Results of the Five Classifiers under Clean and Noisy Environment (MAPACI Database) 116
5.9 Comparison of Positive Predictivity Results of the Five Classifiers under Clean and Noisy Environment (MAPACI Database) 117
5.10 Comparison of Training Time(s) Results of the Five Classifiers under Clean and Noisy Environment (MAPACI Database) 117
5.11 Comparison of Positive Predictivity Results of the Five Classifiers under Clean and Noisy Environment (Dataset III) 118
5.12 Comparison of Overall Accuracy Results of the Five Classifiers under Clean and Noisy Environment (Dataset III) 118
5.13 Comparison of Training Time(s) Results of the Five Classifiers under Clean and Noisy Environment (Dataset III) 119
6.1 Comparison of Overall Accuracy Results of the Five Classifiers under Clean and Noisy Environment (AP Squeezing) 146
6.2 Comparison of Overall Accuracy Results of the Five Classifiers under Clean and Noisy Environment (Vocal Fold Edema) 147
6.3 Comparison of Overall Accuracy Results of the Five Classifiers under Clean and Noisy Environment (Vocal Fold Paralysis) 147
6.4 Comparison of Positive Predictivity Results of the Five Classifiers under Clean and Noisy Environment (AP Squeezing) 148
6.5 Comparison of Positive Predictivity Results of the Five Classifiers under Clean and Noisy Environment (Vocal Fold Edema) 148
6.6 Comparison of Positive Predictivity Results of the Five Classifiers under Clean and Noisy Environment (Vocal Fold Paralysis) 149
6.7 Comparison of Training Times(s) of the Five Classifiers under Clean and Noisy Environment (AP Squeezing)  
6.8 Comparison of Training Times(s) of the Five Classifiers under Clean and Noisy Environment (Vocal Fold Edema)  
6.9 Comparison of Training Times(s) of the Five Classifiers under Clean and Noisy Environment (Vocal Fold Paralysis)
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>Anterior Posterior</td>
</tr>
<tr>
<td>APQ</td>
<td>Amplitude Perturbation Quotient</td>
</tr>
<tr>
<td>AUC</td>
<td>Overall Accuracy</td>
</tr>
<tr>
<td>BBA</td>
<td>Best Basis Algorithm</td>
</tr>
<tr>
<td>BW</td>
<td>Bandwidth</td>
</tr>
<tr>
<td>CWT</td>
<td>Continuous Wavelet Transform</td>
</tr>
<tr>
<td>DCT</td>
<td>Discrete Cosine Transform</td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete Wavelet Transform</td>
</tr>
<tr>
<td>ENT</td>
<td>Ear, Nose and Throat</td>
</tr>
<tr>
<td>EGG</td>
<td>Electroglottograph</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>Fo</td>
<td>Fundamental Frequency</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier Transform</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td>GMMS</td>
<td>Gaussian Mixture Models</td>
</tr>
<tr>
<td>GNE</td>
<td>Glottal to Noise Ratio</td>
</tr>
<tr>
<td>GRNN</td>
<td>General Regression Neural Network</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HCF</td>
<td>Higher Cut off Frequency</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HNR</td>
<td>Harmonics to Noise Ratio</td>
</tr>
<tr>
<td>k-NN</td>
<td>k-Nearest Neighbor</td>
</tr>
<tr>
<td>LCF</td>
<td>Lower Cut off Frequency</td>
</tr>
<tr>
<td>LD</td>
<td>Linear Discriminants</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
<tr>
<td>LDB</td>
<td>Local Discriminant Bases</td>
</tr>
<tr>
<td>LPC</td>
<td>Linear Prediction Coding</td>
</tr>
<tr>
<td>LVQ</td>
<td>Learning Vector Quantization</td>
</tr>
<tr>
<td>MEEI</td>
<td>Massachusetts Eye and Ear Infirmary</td>
</tr>
<tr>
<td>MFCCs</td>
<td>Mel Frequency Cepstral Coefficients</td>
</tr>
<tr>
<td>MLP</td>
<td>Multilayer Perceptron</td>
</tr>
<tr>
<td>NNE</td>
<td>Normalized Noise Energy</td>
</tr>
<tr>
<td>PC</td>
<td>Personal Computer</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PFR</td>
<td>Phonatory Frequency Range</td>
</tr>
<tr>
<td>PNN</td>
<td>Probabilistic Neural Network</td>
</tr>
<tr>
<td>PP</td>
<td>Positive Predictivity</td>
</tr>
<tr>
<td>PPQ</td>
<td>Pitch Perturbation Quotient</td>
</tr>
<tr>
<td>SE</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>SF</td>
<td>Spread Factor</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>SP</td>
<td>Specificity</td>
</tr>
<tr>
<td>SPI</td>
<td>Soft Phonation Index</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>WP</td>
<td>Wavelet Packet</td>
</tr>
<tr>
<td>WPT</td>
<td>Wavelet Packet Transform</td>
</tr>
<tr>
<td>WT</td>
<td>Wavelet Transform</td>
</tr>
</tbody>
</table>
ANALISIS AKUSTIK DAN KLASIFIKASI BAGI SUARA PATOLOGIKAL
DENGAN MENGGUNAKAN PENGELASAN LINEAR DAN TIDAK LINEAR

ABSTRAK

Penyakit vokal dan suara telah meningkat secara mendadak disebabkan keadaan pekerjaan, tabiat sosial
yang tidak sihat dan penyalahgunaan suara. Penyakit vokal memberi kesan kepada bentuk getaran biasa
dalam peti suara dan menyebabkan perubahan dalam gelombang suara akustik. Pakar perubatan
profesional menggunakan teknik yang subjektif untuk memeriksa masalah suara, contohnya, memeriksa
terus kepada pengetar suara dan pemeriksaan kepada pengetar suara menggunakan 'Laryngoscopy'.
Teknik tersebut adalah sangat mahal, berisiko, memerlukan masa yang banyak, menyebabkan
ketidakselesaan kepada pesakit dan memerlukan sumber yang mahal. Analisis akustik bagi gelombang
suara telah terbukti sebagai alat yang terbaik untuk mengesan penyakit vokal kerana ia adalah salah satu
alat yang tidak memberikan kesan sampingan dan memberikan satu pemeriksaan yang objektif. Dalam
penyelidikan ini, satu kaedah tidak-invasif telah dijalankan untuk mengesan penyakit suara melalui analisis
gelombang suara akustik. Dalam tiga puluh tahun ini, beberapa penyelidikan dan pembangunan telah
dijalankan dalam bidang pengesanan penyakit suara automatik dalam bentuk analisis percakapan masa
panjang, analisis percakapan masa pendek, analisis gelombang 'Electroglottographic (EGG)', analisis
masa- frekuensi, pengesanan pergerakan pengetar suara automatik dan teknik pengimejan dan teknik
pemprosesan gelombang tidak sekata. Sebahagian besar parameter jangka panjang dihasilkan dari
frekuensi asas, namun anggaran yang betul bagi frekuensi asas patologi tertentu adalah satu tugas yang
sukar. Walaubagaimanapun, terdapat kaedah penyelesaian alternatif dengan membangunkan algoritma
pengekstrakan sifat yang berkesan. Tiga kaedah pengekstrakan ciri-ciri telah dicadang berdasarkan
kepada perbezaan tenaga domain masa, "Mel Frequency Cepstral Coefficients (MFCC)" digabungkan
dengan "Singular Value Decomposition (SVD)" dan ciri-ciri paket "wavelet" dan entropi tanpa mengira
frekuensi asas. Pengasing linear seperti pengasing berdasarkan "Linear Discriminant Analysis (LDA)" dan
pengasing tak linear seperti pengasing "k-nearest neighbor (k-NN)", "Multilayer Perceptron (MLP)",
"Probabilistic Neural Network (PNN)" dan "General Regression Neural Network (GRNN)" telah dicadangkan
untuk mengasaskan suara patologi dan pada suara biasa. Dalam penyelidikan ini, tiga pangkalan data
seperti "Massachusetts Eye and Ear Infirmary (MEEI) Voice Disorders database", "MAPACI Speech
Pathology database" dan "Dataset- III" (dikumpulkan di Hospital Tengku Fauziah, Kangar, Perlis) telah
digunakan untuk menguji kelajuan algoritma di antara pangkalan-pangkalan data dan di antara
pengekstrakan ciri-ciri yang telah dicadangkan diuji dalam keadaan kehingaran pada 30dB "signal- to- ratio
(SNR)". Dua jenis eksperimen telah dijalankan menggunakan algoritma pengekstrakan ciri-ciri dan
klasifikasi yang telah dicadangkan. Dalam eksperimen pertama, klasifikasi suara normal dan suara
patologikal telah diuji. Dalam eksperimen kedua, pengesanan jenis masalah suara yang spesifik telah
dilakukan melalui masalah klasifikasi bentuk dua kelas. Pelbagai jenis masalah suara telah dipilih seperti
"AP squeezing", "Vocal fold edema" dan "vocal fold paralysis" berdasarkan penyelidikan sebelum ini. Keputusan
eksperimen menjelaskan keadaan yang dicadangkan memberikan ketepatan klasifikasi yang
memberi manfaat untuk pengekstrakan suara biasa dan patologikal dalam kehingaran dan senyap.
Dalam kes pengesanan masalah tertentu, ciri-ciri paket "wavelet" dan entropi memberikan kesan yang lebih
baik berbanding dengan ciri-ciri berdasarkan "MFCC" dan "SVD". Pengukuran prestasi berikut seperti "positive predictivity (PP)", "specificity (SE)", dan "overall accuracy (AUC)" telah dipertimbangkan untuk menjalankan ujian untuk menguji kehandalan dan
keefektifan pengasing linear dan bukan linear. Untuk pengasalan data masalah suara MEEI, kadar kejayaan
pengasing tersebut adalah melebihi 99% untuk pengklasifikasian suara biasa dan patologikal dan untuk
pengesanan masalah tertentu, kadar kejayaan terbaik adalah 100% telah diperolehi. Eksperimen ini juga
terulang untuk "MAPACI speech pathology database" dan "dataset- III" di bawah keadaan hingaran dan
tidak hingaran. Keputusan tersebut menunjukkan bahawa ciri-ciri berdasarkan paket "wavelet" dan entropi
menhasilkan ketepatan klasifikasi yang lebih baik berbanding dengan ciri-ciri berdasarkan tenaga domain
masa dan ciri-ciri berdasarkan "MFCC" dan "SVD" untuk dua lagi pengasalan data. Kesimpulannya, algoritma
pengekstrakan ciri-ciri dan pengklasifikasian yang telah dicadangkan boleh diterapkan untuk membantu
pakar perubatan dalam diasasatan awal bagi masalah suara mengikut aliran perubatan.
ABSTRACT

In this research work, a non-invasive method is conducted to diagnose the voice diseases through acoustic analysis of voice signal. Three feature extraction methods are proposed based on the time-domain energy variations, Mel frequency cepstral coefficients combined with singular value decomposition and wavelet packet and entropy features. Linear classifier namely LDA based classifier and non-linear classifiers such as k-NN classifier, MLP network, PNN, and GRNN are suggested to discriminate pathological voices from normal voices. In this research work, three databases such as MEEI voice disorders database, MAPACI Speech Pathology database, and dataset-III (collected at Hospital Tuanku Fauziah, Kangar, Perlis) are used to test the independence of the algorithms to the databases and the proposed feature extraction algorithms are also tested in noisy condition at 30dB signal-to-noise ratio. Two types of experiments are conducted using the proposed feature extraction and classification algorithms. In the first experiment, classification of normal and pathological voice has been investigated. In the second experiment, the detection of the specific type of voice disorders has been carried out through two-class pattern classification problems. The different kind of voice disorders are selected such as AP squeezing, vocal fold edema and vocal fold paralysis based on the previous research works. The experiment investigations elucidate that the proposed feature extraction algorithms give very promising classification accuracy for the classification of normal and pathological voices under controlled and noisy environment. In the case of detection of specific disorders, wavelet packet and entropy features perform well compared to time-domain energy based features and MFCCs and SVD based features. The following performance measures such as positive predictivity, specificity, sensitivity, and overall accuracy have been considered, in order to test the reliability and effectiveness of the linear and non-linear classifiers. For the MEEI voice disorders database, the success rate of the classifiers is above 98% for the classification of normal and pathological voices and for the detection of specific disorders the best classification accuracy of 100% is achieved. The experiments have also been repeated for the MAPACI speech pathology database and dataset-III under controlled and noisy environment. The results indicate that the wavelet packet and entropy based features provides better classification accuracy compared to time-domain energy based features and MFCCs and SVD based features for the two more databases. It is concluded that proposed feature extraction and classification algorithms can be employed to help the medical professionals for early investigation of voice disorders.
CHAPTER 1

INTRODUCTION

This chapter gives the introduction to the subject of interest, discussion of the existing methods of voice disorders diagnosing methods, its drawbacks and also the advantages of non-invasive methods. This chapter also deals with the objectives of the proposed research and organization of the thesis.

1.1 Preamble

The voice can indicate an individual moods, age or illness. The voice can be used to attract others, to calm others, to irritate, and to frighten others. In this world, people are realizing the importance of voice, only when they got a voice problem. Voice problems affect the normal vibration pattern of the glottis. Voice is very important for certain professionals like singers, teachers, actors, reporters, lawyers, auctioneers, and phone assistants. Vocal fold problems have an impact on people’s professional carriers and their quality of life (Krischke et al., 2005; Rasch, Günther, Hoppe, Eysholdt, & Rosanowski, 2005).

Voice disorders are due to nature of job, unhealthy social habits and due to vocal fatigue after an extensive period of talking. However, the problems may become chronic if the voice is abused or overused when vulnerable. During the upper respiratory infections, the risk of voice damage is increased (Murry & Rosen, 2000). Due to the vibration of the vocal folds, the structure of vocal folds become