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

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AP	Anterior Posterior
APQ	Amplitude Perturbation Quotient
AUC	Overall Accuracy
BBA	Best Basis Algorithm
BW	Bandwidth
CWT	Continuous Wavelet Transform
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
ENT	Ear, Nose and Throat
EGG	Electroglottograph
FFT	Fast Fourier Transform
FN	False Negative
Fo	Fundamental Frequency
FP	False Positive
FT	Fourier Transform
GA	Genetic Algorithm
GMMS	Gaussian Mixture Models
GNE	Glottal to Noise Ratio
GRNN	General Regression Neural Network
GUI	Graphical User Interface
HCF	Higher Cut off Frequency
HMM	Hidden Markov Model

HNR	Harmonics to Noise Ratio
k-NN	k-Nearest Neighbor
LCF	Lower Cut off Frequency
LD	Linear Discriminants
LDA	Linear Discriminant Analysis
LDB	Local Discriminant Bases
LPC	Linear Prediction Coding
LVQ	Learning Vector Quantization
MEEI	Massachusetts Eye and Ear Infirmary
MFCCs	Mel Frequency Cepstral Coefficients
MLP	Multilayer Perceptron
NNE	Normalized Noise Energy
PC	Personal Computer
PDF	Probability Density Function
PFR	Phonatory Frequency Range
PNN	Probabilistic Neural Network
PP	Positive Predictivity
PPQ	Pitch Perturbation Quotient
SE	Sensitivity
SF	Spread Factor
SNR	Signal to Noise Ratio
SP	Specificity
SPI	Soft Phonation Index

SVD	Singular Value Decomposition
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
WP	Wavelet Packet
WPT	Wavelet Packet Transform
WT	Wavelet Transform

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, pemeriksaan 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, pebagai 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 dari 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 mengasingkan suara patologikal daripada suara biasa. Dalam penyelidikan ini, tiga pangkalan data seperti "Massachusett 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 kelainan 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 disiasat. Dalam eksperimen kedua, pengesanan jenis masalah suara yang specifik 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 kaedah yang dicadangkan memberikan ketepatan klasifikasi yang memberangsangkan untuk klasifikasi suara biasa dan patologikal di bawah keadaan hingar dan senyap. Dalam kes pengesanan masalah tertentu, ciri- ciri paket "wavelet" dan entropi memberikan kesan yang lebih baik berbanding dengan ciri- ciri berdasarkan perbezaan tenaga domain masa dan 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 pangkalan data masalah suara MEEI, kadar kejayaan pengasing tersebut adalah melebihi 99% untuk pengklasifikasian suara biasa dan patalogikal dan untuk pengesanan masalah tertentu, kadar kejayaan terbaik adalah 100% telah diperolehi. Eksperimen ini juga telah diulangi untuk "MAPACI speech pathology database" dan "dataset- III" di bawah keadaan hingar dan tidak hingar. Keputusan tersebut menunjukkan bahawa ciri- ciri berdasarkan paket wavelet dan entropi menghasilkan ketepatan klasifikasi yang lebih baik berbanding dengan ciri- ciri berdasarkan tenaga domain masa dan ciri- ciri berdasarkan MFCC dan SVD untuk dua lagi pangkalan data. Kesimpulannya, algoritma pengekstrakan ciri- ciri dan pengklasifikasian yang telah dicadangkan boleh diterapkan untuk membantu pakar perubatan dalam siasatan awal bagi masalah suara mengikut aliran perubatan.

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

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 variations 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