Automatic Speaker Classification Based on Voice Characteristics

A thesis submitted for the degree
of Master of Information Sciences (Research) of
the University of Canberra

Phuoc Thanh Nguyen

December 2010

Summary of Thesis

Gender, age, accent and emotion are some of speaker characteristics being investigated in voice-based speaker classification systems. Classifying speaker characteristics is an important task in the fields of Dialog, Speech Synthesis, Forensics, Language Learning, Assessment, and Speaker Recognition.

It is well known that reducing classification error rate has been a challenge in those research fields. This research thesis investigates new methods for speech feature extraction and classification to meet this challenge. Extracted speech features range from traditional features in speech recognition such as mel-frequency cepstral coefficients (MFCCs) to recently developed prosodic and voice quality features in speaker classification such as pitch, shimmer and jitter. Feature selection was then performed to find a more suitable feature set for building speaker models. For classification methods, feature weighting vector quantisation, Gaussian mixture models (GMMs), Support Vector Machine (SVM) and Fuzzy Support Vector Machine (FSVM) are investigated. Those new feature extraction and classification methods are then applied to gender, age, accent and emotion classification. Four well-known data sets including Australian National Database of Spoken Language (ANDOSL), aGender, EBO-DB, and FAU AIBO are used to evaluate those methods.

The contributions of this thesis to classification of speaker characteristics include:

1. The use of different speech features. Up to 1582 features and transliteration have been investigated.

2. Application of new feature selection method. Correlation based feature subset selection with SFFS was employed to eliminate redundant features because of large databases.

3. The use of fuzzy SVM (FSVM) as a new speaker classification method. FSVM assigns

a fuzzy membership value as a weight to each training data point to allow the decision boundary to move to overlapping regions to reduce empirical errors.

4. A detailed comparison of speaker classification performance for GMMs, SVM and FSVM.

5. A depth investigation on the relevance of feature type for classification of age and gender. Extensive experiments are performed to determine which features in the speech signal are suited to representation of age and gender in human speech.

6. Classification of age, gender, accent, and emotion characteristics is performed on four well-known data sets including ANDOSL, aGender, EBO-DB and FAU AIBO.

Certificate of Authorship of Thesis

Except where clearly acknowledged in footnotes, quotations and the bibliography, I certify that I am the sole author of the thesis submitted today entitled —

Automatic Speaker Classification Based on Voice Characteristics

I further certify that to the best of my knowledge the thesis contains no material previously published or written by another person except where due reference is made in the text of the thesis.

The material in the thesis has not been the basis of an award of any other degree or diploma except where due reference is made in the text of the thesis.

The thesis complies with University requirements for a thesis as set out in *Gold Book Part 7: Examination of Higher Degree by Research Theses Policy, Schedule Two (S2).*
Refer to http://www.canberra.edu.au/research-students/goldbook

Signature of Candidate

Signature of chair of the supervisory panel

Date

Acknowledgements

First and foremost, I would like to thank my supervisor, A/Prof. Dat Tran, for his enormous support during my study at the University of Canberra. I am also thankful for his valuable guidance both in research and life, his encouragement and attention to important research milestones and events, his very quick response to my questions, and his patience to help me enhance the thesis.

I would also like to thank my co-supervisor, A/Prof. Xu Huang, for his encouragement, advice, support and suggestions on research plans. I am also thankful for his patience to revised my thesis and careful feedbacks.

I would also like to thank the Faculty of Information Sciences and Engineering for supporting conference travels and maintaining the excellent computing facilities which were crucial for carrying out my research. Thanks to staff members as well as research students for discussions and seminars. A grateful thanks to Prof. John Campbell for his Research Proposal and Research Methodologies courses. A warm thanks to Mr. Hanh Huynh for his interesting discussions about life and encouragement. A special thanks to Trung Le for his valuable discussions.

More importantly, I would like to thank the HCMC University of Pedagogy, Viet Nam for providing me the scholarship which enabled me to undertake this research at the University of Canberra. I would like to express my gratitude to all my lecturers and colleagues at the Faculty of Mathematics and Informatics, HCMC University of Pedagogy. I wish to express my warm and sincere thanks to Dr. Nguyen Thai Son and Msc. Ly Anh Tuan, Faculty of Mathematics and Informatics, HCMC University of Pedagogy for their important guidance, support and encouragement during my first steps in the Faculty.

I devote my deepest gratitude to my parents for their unlimited love and support. They have encouraged me throughout the years of my study. The most special thanks belong to my wife Huyen, for her understanding about my leaving during all these years of my absence, her selfless love and support all along and encouragement.

Contents

List of Figures

List of Tables

Abbreviation

GMMs	Gaussian Mixture Models
SVM	Support Vector Machine
FSVM	Fuzzy Support Vector Machine
HTK	Hidden Markov Model Toolkit
HMM	Hidden Markov Model
SFFS	Sequential Forward Floating Search
MFCCs	Mel-Frequency Cepstral Coefficients
LPC	Linear Prediction Coding