Signals and Communication Technology

For further volumes:
http://www.springer.com/series/4748

Qi (Peter) Li

Speaker Authentication

Springer

Dr. Qi (Peter) Li
Li Creative Technologies (LcT), Inc.
Vreeland Road 30 A, Suite 130
Florham Park, NJ 07932
USA
e-mail: li@licreativetech.com

ISSN 1860-4862
ISBN 978-3-642-23730-0 e-ISBN 978-3-642-23731-7
DOI 10.1007/978-3-642-23731-7
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011939406

Cover design: eStudio Calamar, Berlin/Figueres

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my parents YuanLin Shen and YanBin Li

Preface

My research on speaker authentication started in 1995 when I was an intern at Bell Laboratories, Murray Hill, New Jersey, USA, while working on my Ph.D. dissertation. Later, I was hired by Bell Labs as a Member of Technical Staff, which gave me the opportunity to continue my research on speaker authentication with my Bell Labs colleagues. In 2002, I established Li Creative Technologies, Inc. (LcT), located in Florham Park, New Jersey. At LcT, I am continuing my research in speaker authentication with my LcT colleagues. Recently, when I looked at my publications during the last fifteen years, I found that my research has covered all the major research topics in speaker authentication: from front-end to back-end; from endpoint detection to decoding; from feature extraction to discriminative training; from speaker recognition to verbal information verification. This has motivated me to put my research results together into a book in order to share my experience with my colleagues in the field.

This book is organized by research topic. Each chapter focuses on a major topic and can be read independently. Each chapter contains advanced algorithms along with real speech examples and evaluation results to validate the usefulness of the selected topics. Special attention has been given to the topics related to improving overall system robustness and performance, such as robust endpoint detection, fast discriminative training theory and algorithms, detection-based decoding, and sequential authentication. I have also given attention to those novel approaches that may lead to new research directions, such as a recently developed auditory transform (AT) to replace the fast Fourier transform (FFT) and auditory-based feature extraction algorithms.

For real applications, a good speaker authentication system must first have an acceptable authentication accuracy and then be robust to background noise, channel distortion, and speaker variability. A number of speaker authentication systems can be designed based on the methods and techniques presented in this book. A particular system can be designed to meet required specifications by selecting an authentication method or combining several authentication and decision methods introduced in the book.

Speaker authentication is a subject that relies on the research efforts of many different fields, including, but not limited to, physics, acoustics, psychology, physiology, hearing, auditory nerve, brain, auditory perception, parametric and nonparametric statistics, signal processing, pattern recognition, acoustic phonetics, linguistics, natural language processing, linear and nonlinear programming, optimization, communications, etc. This book only covers a subset of these topics. Due to my limited time and experience, this book only focuses on the topics in my published research. I encourage people with the above backgrounds to consider contributing their knowledge to speech recognition and speaker authentication research. I also encourage colleagues in the field of speech recognition and speaker authentication to extend their knowledge to the above fields in order to achieve breakthrough research results.

This book does not include those fundamental topics which have been very well introduced in other textbooks. This author assumes the reader has a basic understanding of linear systems, signal processing, statistics, and pattern recognition.

This book can also be used as a reference book for government and company officers and researchers working in information technology, homeland security, law enforcement, and information security, as well as for researchers and developers in the areas of speaker recognition, speech recognition, pattern recognition, and audio and signal processing. It can also be used as a reference or textbook for senior undergraduate and graduate students in electrical engineering, computer science, biomedical engineering, and information management.

Acknowledgments

The author would like to thank the many people who helped the author in his career and in the fields of speaker and speech recognition. I am particularly indebted to Dr. Donald W. Tufts at the Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, for his role in guiding and training me in pattern recognition and speech signal processing.

Special thanks are due to Dr. S. Parthasathy and Dr. Aaron Rosenberg who served as mentors when I first joined Bell Laboratories. They led me into the field of speaker verification research. I am particularly grateful to Dr. Biing-Hwang (Fred) Juang for his guidance in verbal information verification research. The work extended speaker recognition to speaker authentication, which has broader applications.

Most topics in this book were prepared based on previously published peer-reviewed journal and conference papers where I served as the first author. I would like to thank all the coauthors of those publications, namely Dr. Donald Tufts, Dr. Peter Swaszek, Dr. S. Parthasarathy, Dr. Aaron Rosenberg, Dr. Biing-Hwang Juang, Dr. Frank Soong, Dr. Chin-Hui Lee, Qiru Zhou, Jinsong Zheng, Dr. Augustine Tsai, and Yan Huang. Also, I would like to

thank the many anonymous reviewers and editors for their helpful comments and suggestions.

The author also would like to thank Dr. Bishnu Atal, Dr. Joe Olive, Dr. Wu Chou, Dr. Oliver Siohan, Dr. Mohan Sondhi, Dr. Oded Ghitza, Dr. Jingdong Chen, Dr. Rafid Sukkar, Dr. Larry O'Gorman, Dr. Richard Rose, and Dr. David Roe, all former Bell Laboratories colleagues, for their useful discussions and their kind help and support on my research there. Also, I would like to thank Dr. Ivan Selesnick for our recent collaborations.

Within Li Creative Technologies, the author would like to thank Yan Huang and Yan Yin for our recent collaborations in speaker identification research. I also would like to thank my colleagues Dr. Manli Zhu, Dr. Bozhao Tan, Uday Jain, and Joshua Hajicek for useful discussions on biometrics, acoustic, speech, and hearing systems.

From 2008 to 2010, the author's research on speaker identification was supported by the U.S. AFRL under the contract number FA8750-08-C-0028. I would like to thank program managers Michelle Grieco, John Parker, and Dr. Stanly Wenndt for their help and support. Some of the research results have been included in Chapter 7 and Chapter 8 of this book. Other results will be published later.

I would like to thank my colleague Craig B. Adams and my daughter Joy Y Li for their work in editing this book. I would like to thank Uday Jain, Dr. Manli Zhu and Dr. Bozhao Tan for their proofreading. Also, this book could not have been finished without the support of my wife Vivian for the many weekends which I spent working on it.

The author also would like to thank the IEEE Intellectual Property Rights Office for permissions to use the IEEE copyright materials which I previously published in IEEE publications in the book.

Finally, I would like to thank Dr. Christoph Baumann, Engineering Editor at Springer, for his kind invitation to prepare and publish this book.

Florham Park, NJ *Qi (Peter) Li*
July 2011

Contents

List of Tables

List of Figures