SpringerBriefs in Electrical and Computer Engineering

Speech Technology

Series Editor

Amy Neustein

For further volumes:
http://www.springer.com/series/10043

Editor's Note

The authors of this series have been hand selected. They comprise some of the most outstanding scientists—drawn from academia and private industry—whose research is marked by its novelty, applicability, and practicality in providing broad-based speech solutions. The Springer Briefs in Speech Technology series provides the latest findings in speech technology gleaned from comprehensive literature reviews and *empirical investigations* that are performed in both laboratory and *real life* settings. Some of the topics covered in this series include the presentation of real life commercial deployment of spoken dialog systems, contemporary methods of speech parameterization, developments in information security for automated speech, forensic speaker recognition, use of sophisticated speech analytics in call centers, and an exploration of new methods of soft computing for improving human–computer interaction. Those in academia, the private sector, the self service industry, law enforcement, and government intelligence are among the principal audience for this series, which is designed to serve as an important and essential reference guide for speech developers, system designers, speech engineers, linguists, and others. In particular, a major audience of readers will consist of researchers and technical experts in the automated call center industry where speech processing is a key component to the functioning of customer care contact centers.

Amy Neustein, Ph.D., serves as editor in chief of the *International Journal of Speech Technology* (Springer). She edited the recently published book *Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics* (Springer 2010), and serves as quest columnist on speech processing for Womensenews. Dr. Neustein is the founder and CEO of Linguistic Technology Systems, a NJ-based think tank for intelligent design of advanced natural language-based emotion detection software to improve human response in monitoring recorded conversations of terror suspects and helpline calls.

Dr. Neustein's work appears in the peer review literature and in industry and mass media publications. Her academic books, which cover a range of political, social, and legal topics, have been cited in the Chronicles of Higher Education and have won her a pro Humanitate Literary Award. She serves on the visiting faculty of the National Judicial College and as a plenary speaker at conferences in artificial intelligence and computing. Dr. Neustein is a member of MIR (machine intelligence research) Labs, which does advanced work in computer technology to assist underdeveloped countries in improving their ability to cope with famine, disease/illness, and political and social affliction. She is a founding member of the New York City Speech Processing Consortium, a newly formed group of NY-based companies, publishing houses, and researchers dedicated to advancing speech technology research and development.

Raghunath S. Holambe
Mangesh S. Deshpande

Advances in Non-Linear Modeling for Speech Processing

Springer

Raghunath S. Holambe
Department of Instrumentation
SGGS Institute of Engineering
and Technology
Vishnupuri
Nanded 431606
India

Mangesh S. Deshpande
Department of E & TC Engineering
SRES College of Engineering
Kopargaon 423603
India

ISSN 2191-8112
ISBN 978-1-4614-1504-6
DOI 10.1007/978-1-4614-1505-3
Springer New York Heidelberg Dordrecht London

e-ISSN 2191-8120
e-ISBN 978-1-4614-1505-3

Library of Congress Control Number: 2012931407

Printed on acid-free paper

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

To our families
For their love, encouragement and support

Preface

Speech production and perception, man's most widely used means of communication, has been the subject of research and intense study for more than 10 decades. Conventional theories of speech production are based on linearization of pressure and volume velocity relations and the speech production system is modeled as a linear source-filter model. This source-filter model is the foundation of many speech processing applications such as speech coding, speech synthesis, speech recognition and speaker recognition technology. However, this modeling technique neglects some nonlinear aspects of speech production. The main purpose of this book is to investigate advanced topics in nonlinear estimation and modeling techniques and their applications to speaker recognition.

The text consists of six chapters that are outlined in detail in Chap. 1. Chapter 2 reviews the fundamentals of speech production and speech perception mechanisms. Some important aspects of physical modeling of speech production system like vocal fold oscillations, the turbulent sound source, aerodynamics observations regarding nonlinear interactions between the air flow and the acoustic field etc. are discussed in this chapter. In Chap. 3, the linear as well as nonlinear modeling techniques of the speech production system are discussed. The widely used source-filter model, its limitations and introduction to dynamic system model are covered in this chapter. Finally, different parametric as well as nonparametric approaches for approximations of nonlinear model are presented.

Advanced topics in nonlinear estimation and modeling are investigated in Chap. 4. Introduction to Teager energy operator (TEO), energy separation algorithms and noise suppression capability of TEO is discussed in this chapter. In Chap. 5, the speech production process is modeled as an AM-FM model which overcomes the limitations of linear source-filter model of speech production and features derived from it like linear prediction cepstral coefficients (LPCC) and mel frequency cepstral coefficients (MFCC). Demodulation techniques like energy separation algorithm using TEO and Hilbert transform demodulation are discussed in this chapter. Based on the foundational Chaps. 2–5, in Chap. 6, an application of the nonlinear modeling techniques is discussed. This chapter covers the performance evaluation of different features based on nonlinear modeling techniques

applicable to a speaker identification system. Session variability is one of the challenging tasks in speaker identification. This variability in terms of mismatched environments seriously degrades the identification performance. In order to address the problem of environment mismatch due to noise, different types of robust features are discussed in this chapter. These features make use of nonlinear aspects of speech production model and outperform the most widely accepted MFCC features. The proposed features like Teager energy operator based cepstral coefficients (TEOCC) and amplitude-frequency modulation (AM-FM) based 'Q' features show significant improvement in speaker identification rate in mismatched environments. The performance of these features is evaluated for different types of noise signals in the NOISEX-92 database with clean training and noisy testing environments.

More recently, speech and signal processing researchers have espoused the importance of nonlinear techniques. As this book covers the basics as well as some applications related to speaker recognition technology, this book may be very useful for the researchers working in the speaker recognition area. As compared to the state-of-the-art features which are based on speech production or speech perception mechanism, a new idea is explored to combine the speech production and speech perception systems to derive robust features.

Acknowledgments

We are grateful to many teachers, colleagues and researchers, who directly or indirectly helped us in preparing this book. Very special cordial thanks go to Dr. Hemant A. Patil, Professor, Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gujarat, for his thorough comments and astute suggestions which immensely enhanced the quality of the work. We are thankful to Dr. S. R. Kajale, Director, Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded and Dr. D. N. Kyatanavar, Principal, Prof. R. P. Sood, Director, SRES's College of Engineering, Kopargaon, for their motivation and constant support while preparing the manuscript. We are also thankful to all faculty members of the Department of Instrumentation Engineering, Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded as well as Department of Electronics and Telecommunication Engineering, SRES's College of Engineering, Kopargaon, for their direct and indirect help during completion of this work. We would like to acknowledge our colleagues who have involved indirectly with this work, Dr. J. V. Kulkarni, Head, Department of Instrumentation Engineering, VIT, Pune and Dr. D. V. Jadhav, Principal, Bhiwarabai Sawant College of Engineering, Pune. Finally, we wish to acknowledge Dr. Amy Neustein, Editor, Series in Speech Processing (Springer Verlag) for her unusually great help and efforts during the period of preparing the manuscript and producing the book.

Raghunath Holambe
Mangesh Deshpande

Contents