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.

http://www.springer.com/978-1-4614-1504-6

Advances in Non-Linear Modeling for Speech Processing
Processing
Holambe, R.S.; Deshpande, M.S.
2012, XIII, 102 p. 32 illus., Softcover
ISBN: 978-1-4614-1504-6