Preface

Robust speech systems in mobile environment have gained a special interest in recent years in order to enable access to remote voice-activated services. In this context, three major challenges that need to be considered are: varying background conditions, speech coding, and transmission channel errors. In this book, we focus on improving the recognition performance of speech systems in the presence of speech coding and background noise conditions by using vowel onset points (VOPs) as anchor points. VOP is an important event in speech production, and it is defined as the instant at which the onset of vowel takes place. Speech coders considered in this work are GSM full rate (ETSI 06.10), GSM enhanced full rate (ETSI 06.60), CELP (FS-1016), and MELP (TI 2.4 kbps).

The major works presented in this book are:

- Methods are proposed for the detection of VOPs in the presence of speech coding and background noise conditions.
- A two-stage hybrid approach based on hidden Markov models (HMMs) and support vector machines (SVMs) is proposed for improving the performance of consonant-vowel (CV) recognition system.
- Two-stage VOP detection method is proposed for spotting CV units from continuous speech.
- Combined temporal and spectral preprocessing methods are explored to improve the performance of CV recognition system under background noise.
- A method based on VOPs is proposed to improve the performance of speaker identification (SI) system in the presence of coding.
- A method is proposed for nonuniform time scale modification using VOPs and instants of significant excitation.

Some important conclusions drawn out of this work are: (i) Performance of the proposed VOP detection method based on spectral energy in the glottal closure region is found to be better compared to existing methods under clean, coded and noisy conditions. (ii) Performance of the proposed two-stage hybrid CV recognition approach has shown significant improvement compared to other approaches, under clean, coded, and noisy conditions. (iii) Performance of CV recognition system

under background noise is improved by using combined temporal and spectral processing-based preprocessing method. (iv) Proposed two-stage VOP detection method used for spotting CV segments from continuous speech has found to be efficient in minimizing the missing and spurious VOPs. (v) In the presence of coding, performance of SI system is improved by using features extracted from steady vowel speech segments. Improvement in SI system performance is mainly due to the presence of crucial speaker-specific information in the steady vowel segments of speech, even after coding. (vi) Performance of the proposed time scale modification method is superior compared to existing methods. The superior performance of the proposed method is due to the nonuniform modification of different speech segments and accurate detection of various speech segments with the help of instants of significant excitation and VOPs.

This book is mainly intended for researchers working on building robust speech systems in mobile environment. This book is also useful for the young researchers, who want to pursue the research in speech processing. Hence, this may be recommended as the text or reference book for the postgraduate level advanced speech processing course.

Many people have helped us during the course of preparation of this book. We would especially like to thank all professors of G.S. Sanyal School of Telecommunication and School of Information Technology, IIT Kharagpur for their moral encouragement and technical discussions during the course of editing and organization of the book. Special thanks to our colleagues at IIT Kharagpur for their cooperation and coordination to carry out the work. Finally, we thank all our friends and well-wishers.

West Bengal, India   K. Sreenivasa Rao
Hyderabad, India   Anil Kumar Vuppala