

# Preface

The goal of developing a phone recognition system (PRS) is to derive the sequence of basic sound units from the speech signal. Most of the state-of-the-art PRSs are developed using spectral features such as Mel frequency cepstral coefficients. Spectral features mainly represent the gross shape of the vocal tract, but not the information related to the excitation source or the positioning and movements of various articulators. But, the production of each sound unit is characterized by articulatory and excitation source features in addition to vocal tract features. It is impossible to produce a sound unit without having an appropriate source of excitation. The rate of vibration of vocal folds varies from one phone to another phone based on their inherent characteristics as well as the influence of coarticulation characteristics due to the presence of adjacent phones. The positioning and movement of various articulators during the production of a sound unit change from one sound unit to another. A unique combination of articulators in the vocal tract and specific source of excitation results in production of a particular sound unit.

In this work, the articulatory and excitation source features are explored for improving the performance of PRSs. The articulatory features (AFs) are derived from the spectral features using feedforward neural networks (FFNNs). Five AF groups, namely manner, place, roundness, frontness, and height, are considered. Five different AF-based tandem PRSs are developed using the combination of spectral features and AFs derived from FFNNs of each AF group. The systematic analysis of phone-level accuracies contributed by each AF group is carried out. Hybrid PRSs are developed by combining the evidences from AF-based tandem PRSs using weighted combination approach. It is observed that the use of AFs in addition to spectral features has lead to improvement in the performance of PRSs.

The excitation source information is derived by processing linear prediction (LP) residual of the speech signal. The use of excitation source information has shown improvement in the performance of PRSs. The robustness of proposed excitation source features is demonstrated using white and babble noisy speech samples. The PRSs developed using the combination of vocal tract and excitation source features are more robust to noise than the PRSs developed using vocal tract features alone. The performance of tandem PRSs is improved using excitation

source features in addition to spectral features. The performance of PRSs developed using articulatory and excitation source features across read, extempore, and conversation modes of speech is analyzed, and results are compared. The use of articulatory and excitation source features has shown improvement in all the three modes of speech.

This book is mainly intended for researchers working on speech recognition area. This book is also useful for the young researchers, who want to pursue research in speech processing with an emphasis on articulatory and excitation source features. Hence, this may be recommended as the text or reference book for the postgraduate level advanced speech processing course. This book has been organized as follows:

Chapter 1 introduces basic concepts of speech recognition and its applications. The articulatory and excitation source features are described briefly. Chapter 2 provides compendious reviews about the use of articulatory and excitation source features to develop speech recognition systems. Chapter 3 discusses the proposed approaches to derive and use AFs for phone recognition task. The development of tandem and hybrid PRSs using AFs is proposed. Chapter 4 describes the proposed methods to parameterize and use the excitation source information to perform phone recognition. The use of excitation source features to improve robustness of PRSs is also discussed. Chapter 5 investigates the use of articulatory and excitation source features to improve performance of PRSs across read, extempore, and conversation modes of speech. Chapter 6 provides a brief summary and conclusion of this book with a glimpse toward the scope for possible future work.

We would especially like to thank all professors of computer science and engineering, IIT Kharagpur for their moral encouragement and technical discussions during course of editing and organization of this book. Special thanks to our colleagues at Indian Institute of Technology Kharagpur, India, for their cooperation to carry out the work. We are grateful to our parents and family members for their constant support and encouragement. Finally, we thank all our friends and well-wishers.

Kharagpur, India  
Bangalore, India

K. Sreenivasa Rao  
Manjunath K.E.

<http://www.springer.com/978-3-319-49219-3>

Speech Recognition Using Articulatory and Excitation  
Source Features

Rao, K.S.; K.E., M.

2017, XI, 92 p. 23 illus., 4 illus. in color., Softcover

ISBN: 978-3-319-49219-3