SpringerBriefs in Electrical and Computer Engineering

Speech Technology

Series editor

Amy Neustein, Fort Lee, USA

More information about this series at http://www.springer.com/series/10043

K. Sreenivasa Rao · V. Ramu Reddy
Sudhamay Maity

Language Identification Using Spectral and Prosodic Features

Springer

K. Sreenivasa Rao
Indian Institute of Technology Kharagpur
Kharagpur, West Bengal
India

Sudhamay Maity
Indian Institute of Technology Kharagpur
Kharagpur, West Bengal
India

V. Ramu Reddy
Innovation Lab Kolkata
Kolkata, West Bengal
India

ISSN 2191-8112 ISSN 2191-8120 (electronic)
SpringerBriefs in Electrical and Computer Engineering
ISSN 2191-737X ISSN 2191-7388 (electronic)
Speech Technology
ISBN 978-3-319-17162-3 ISBN 978-3-319-17163-0 (eBook)
DOI 10.1007/978-3-319-17163-0

Library of Congress Control Number: 2015935735

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

The basic goal of a language identification (LID) system is to accurately identify the language from the given speech sample. In literature, LID studies were carried out extensively on western and eastern languages by exploring various language-specific features and models. But, in India, language discrimination tasks have been carried out using very few (less than six) Indian languages (ILs), and no systematic study was observed. Discrimination of ILs is a tough task due to high similarity of languages. This is mainly due to common origin (i.e., Sanskrit) for all ILs. In this work, we have analyzed the LID performance on 27 Indian languages with spectral and prosodic features. In our study, spectral features extracted from block processing (BP), pitch synchronous analysis (PSA), and glottal closure regions (GCR) are explored for identification of languages. It was observed that the performance of LID system is better by using the spectral features derived from PSA and GCR compared to BP. Prosodic features represented by intonation, rhythm, and stress (IRS) are proposed at syllable and word levels for discriminating the languages. For representing language-specific information at the global level, dynamics of fundamental frequency (F_0), duration, and energy patterns are proposed in this study. Complementary information present in different features is exploited by combining the systems developed by using individual features. It is observed that the performance of LID system has been improved by combining various components of prosodic features and spectral features. The proposed features are also verified using OGI database, and observed the similar trend. The major contributions of this book are given below:

- IITKGP-MLILSC (Indian Institute of Technology Kharagpur Multilingual Indian Language Speech Corpus) is developed for promoting LID research on ILs.
- Spectral features from pitch synchronous analysis and glottal closure regions are proposed for discriminating the ILs.
- Intonation, rhythm, and stress features at individual syllable level and in the sequence of syllables within a word are proposed for identifying the ILs.

- Prosodic patterns at global level are proposed in terms of variation of F_0, intensity, and duration patterns for recognizing the ILs.
- Language-specific information from the spectral and prosodic features is combined for improving the accuracy of LID system.

This book is mainly intended for researchers working on language identification area. The book is also useful for young researchers, who want to pursue research in speech processing with an emphasis on spectral and prosodic features. Hence, this may be recommended as the text or reference book for the postgraduate level advanced speech processing course.

The book has been organized as follows: Chap. 1 contains a brief introduction about language identification and its applications. General issues in language identification, language-specific cues in speech signal and specific issues in identification of Indian languages are discussed. Chapter 2 briefly reviews explicit and implicit LID systems present in the literature. Chapter 3 introduces multi-lingual Indian language speech corpus and discusses spectral features extracted from conventional block processing, pitch synchronous analysis, and glottal closure regions for discriminating the languages. Chapter 4 discusses language identification using intonation, rhythm, and stress features derived from syllable and word levels. Chapter 5 provides a brief summary and conclusion of the book with a glimpse towards the scope for possible future work.

We would especially like to thank all professors of School of Information and 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, 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 wellwishers.

K. Sreenivasa Rao
V. Ramu Reddy
Sudhamay Maity

Contents

Acronyms

AANN	Auto-Associative Neural Network
AIR	All India Radio
ASR	Automatic Speech Recognition
BP	Block Processing
CV	Consonant-Vowel
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
EER	Equal Error Rate
FFNN	Feed-Forward Neural Network
FOD	First-Order Difference
FOGD	First-Order Gaussian Difference
GC	Glottal Closure
GCI	Glottal Closure Instants
GCR	Glottal Closure Region
GMM	Gaussian Mixture Models
IDFT	Inverse Discrete Fourier Transform
IL	Indian Language
IRS	Intonation Rhythm Stress
ISE	Instant of Significant Excitation
LID	Language Identification
LP	Linear Prediction
LPCC	Linear Prediction Cepstral Coefficients
MFCC	Mel-Frequency Cepstral Coefficients
ms	Milli Seconds
PLP	Perceptual Linear Prediction
PSA	Pitch Synchronous Analysis
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machines
VOP	Vowel Onset Point
VT	Vocal Tract
ZFF	Zero Frequency Filter