

Preface

Language identification (LID) is a process of determining the language from the uttered speech. Most LID studies have been carried out using vocal tract and prosodic features. However, the characteristics of excitation source have not been explored for LID study. In this book, implicit and explicit features of excitation source have been explored for language discrimination task. Linear prediction (LP) residual signal is used for representing excitation source signal. The implicit relations among the raw LP residual samples, its magnitude, and phase components are explored to capture the language-specific excitation source information. The proposed implicit features consist of raw LP residual samples, its magnitude and phase components at three different levels: (i) sub-segmental level (within a glottal cycle or pitch cycle), (ii) segmental level (within 2–3 successive glottal cycles), and (iii) supra-segmental level (across 50 glottal cycles). These features capture the implicit language-specific phonotactic constraints embedded in excitation source signal. Evidences obtained from each level are combined to derive complete implicit features of excitation source for LID task.

In addition to implicit features of excitation source, LP residual signal has also been parameterized at sub-segmental, segmental, and supra-segmental levels to capture the language-specific phonotactic information. At sub-segmental level, the characteristics of a single glottal pulse have been modeled using glottal flow derivative (GFD) parameters of LP residual signal. Residual mel frequency cepstral coefficients (RMFCC) and mel power difference of spectrum in sub-band (MPDSS) features are explored to derive language-specific excitation source information at segmental level. At supra-segmental level, temporal variations of pitch, epoch strength, and epoch sharpness are explored for capturing language-specific supra-segmental level source information. Evidences from each level are combined to capture complete parametric representation of excitation source. Further, evidences obtained from implicit and parametric features are combined to acquire overall language-specific excitation source information.

The nonoverlapping language-specific information present in excitation source and vocal tract features has also been investigated in this book. The robustness of proposed excitation source features has been examined by varying (i) amount of

training data, (ii) length of test samples, and (iii) background noise characteristics. From experimental studies, it has been observed that the excitation source is more robust, compared to vocal tract features for LID task.

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 excitation source features. Hence, this may be recommended as a text or reference book for the postgraduate level advanced speech processing course. The book has been organized as follows:

Chapter 1 introduces the basic concept of language identification (LID) and the various features used in LID. The application of LID system has been demonstrated. Chapter 2 provides a review of the methods reported in prior works of LID. Chapter 3 discusses the implicit features of excitation source for language recognition task. Chapter 4 explores the parametric features of excitation source for language discrimination study. Chapter 5 investigates the robustness of excitation source features in the context of language identification. Chapter 6 provides a brief summary and conclusion of the book with a glimpse toward 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 well wishers.

K. Sreenivasa Rao
Dipanjan Nandi

Language Identification Using Excitation Source
Features

Rao, K.S.; Nandi, D.

2015, XII, 119 p. 19 illus., 3 illus. in color., Softcover

ISBN: 978-3-319-17724-3