

Preface

Human beings use speech as a primary mode of communication for conveying messages. A speech signal carries multiple cues related to intended message, speaker and language identities, behavioural and emotional mood of the speaker and characteristics of background environment. Human beings exploit all these cues for performing various speech tasks. Now a days in all areas of applications, machines are extensively used for performing the desired tasks automatically. For interacting with machines, humans feel speech interface to be the most convenient one. Speech interface to machine should take care of basic tasks such as speech understanding and speech generation.

Most of the present speech systems support the basic speech tasks for the neutral speech recorded in a clean environment. But, in general human speech is mostly embedded with emotions to convey the intended message. Processing emotional speech by a machine is a complex task. In real life, emotions are highly dynamic and depend on several factors such as speaker, language, culture, society and geographical regions. The basic objective of this book is to explore various features and models for characterising and discriminating the emotions. In addition to conventional speech features, emotion-specific spectral and prosodic features are introduced in this book for performing robust emotion recognition.

This book primarily focuses on spectral and prosodic features to discriminate emotions. In addition to conventional spectral features extracted from the entire speech through block processing, sub-syllabic regions such as consonants, vowels and CV (Consonant-to-Vowel) transition regions are explored for computing spectral features. Spectral features extracted through pitch synchronous analysis are explored for reducing the ambiguity in confusable emotions. Prosodic features extracted from words and syllables present at different positions (initial, middle and final) are investigated to analyse the emotion-specific knowledge. Various nonlinear models, such as auto-associative neural networks, support vector machines and Gaussian mixture models, are explored to capture the emotion-specific information from the features mentioned above. In this book, emotional database Indian Institute of Technology Kharagpur Simulated Emotion Speech Corpus (IITKGP-SESC) in an Indian language Telugu, Berlin emotional speech

database (Emo-DB) and Hindi movie database are used for analysing the emotion recognition performance.

This book is mainly intended for researchers working on emotion recognition from speech. The book is also useful for young researchers, who want to pursue research in speech processing using basic excitation source, vocal tract and prosodic features. Hence, this may be recommended as the text or reference book at the postgraduate level advanced speech processing course. The book has been organised as follows:

[Chapter 1](#) introduces automatic emotion recognition from speech as one of the thrust areas of research in speech processing. Psychological and engineering aspects of emotional speech signal have been discussed. [Chapter 2](#) discusses the non-conventional spectral features extracted from pitch synchronous analysis and sub-syllabic regions for discriminating the emotions. [Chapter 3](#) discusses the global and local prosodic features extracted from words and syllables for classifying the emotions. [Chapter 4](#) exploits the complementary information provided by various features by combining their evidences using appropriate fusion techniques. [Chapter 5](#) introduces multi-stage emotion classification using speaking rate features. [Chapter 6](#) performs real-life emotion classification using the proposed robust features. [Chapter 7](#) summarises the contents of the book, highlights the contributions of the chapters and discusses the scope for future work.

Many people have helped us during the course of preparation of this book. 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 organisation of this book. Special thanks to our colleagues at Indian Institute of Technology, Kharagpur, India and Graphic Era University, Dehradun, India for their co-operation and coordination to carry out the work. We preferably acknowledge the efforts of Anurag Barthwal, Swati Devliyal, Kritik Sharma, Shan-e-Fatima and Deepika Rastogi in preparing the draft of [Chap. 6](#). 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
Shashidhar G. Koolagudi

Robust Emotion Recognition using Spectral and
Prosodic Features

Rao, K.S.; Koolagudi, S.G.

2013, XII, 118 p. 37 illus., 15 illus. in color., Softcover

ISBN: 978-1-4614-6359-7