

SpringerBriefs in Electrical and Computer Engineering

For further volumes:
www.springer.com/series/10059

Sid-Ahmed Selouani

Speech Processing and Soft Computing



Springer

Sid-Ahmed Selouani
Université de Moncton
Shippagan Campus
218, Boul. J-D Gauthier
Moncton
Canada
selouani@umcs.ca

ISSN 2191-8112 e-ISSN 2191-8120
ISBN 978-1-4419-9684-8 e-ISBN 978-1-4419-9685-5
DOI 10.1007/978-1-4419-9685-5
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011936520

© Springer Science+Business Media, LLC 2011

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Soft Computing (SC) techniques have been recognized nowadays as attractive solutions for modeling highly nonlinear or partially defined complex systems and processes. These techniques resemble biological processes more closely than conventional (more formal) techniques. However, despite its increasing popularity, soft computing lacks a precise definition because it is continuously evolving by including new concepts and techniques. Generally speaking, SC techniques encompass two main concepts: approximate reasoning and function approximation and/or optimization. They constitute a powerful tool that can perfectly complement the well-established formal approaches when certain aspects of the problem to solve require dealing with uncertainty, approximation and partial truth. Many real-life problems related to sociology, economy, science and engineering can be solved most effectively by using SC techniques in combination with formal modeling. This book advocates the effectiveness of this combination in the field of speech technology which has provided systems that have become increasingly visible in a wide range of applications.

Speech is a very complex phenomenon involving biological information processing system that enables humans to accomplish very sophisticated communication tasks. These tasks use both logical and intuitive processing. Conventional ‘hard computing’ approaches have achieved prodigious progress, but their capabilities are still far behind that of human beings, particularly when called upon to cope with unexpected changes encountered in the real world.

Therefore, bridging the gap between the SC concepts and speech technology is the main purpose of this book. It aims at covering some important advantages that speech technology can draw from bio-inspired soft computing methods. Through practical cases, we will explore, dissect and examine how soft computing complement conventional techniques in speech enhancement and speech recognition in order to provide more robust systems.

This book is a result of my research, since 2000, at INRS-EMT Research Institute (Montreal, Canada) and LARIHS Laboratory in Moncton University (New Brunswick, Canada). Its content is structured so that principles and theory are

often followed by applications and supplemented by experiments. My goal is to provide a cohesive vision on the effective use of soft computing methods in speech enhancement and speech recognition approaches.

The book is divided into two parts. Each part contains four chapters. Part I is entitled *Soft Computing and Speech Enhancement*. It looks at conventional techniques of speech enhancement and their evaluation methods, advocates the usefulness of hybridizing hierarchical connectionist structure with subspace decomposition methods, as well as the effectiveness of a new criterion to optimize the process of the subspace-based noise reduction. It also shows the relevance of evolutionary-based techniques in speech enhancement. Part II, *Soft Computing and Speech Recognition*, addresses the speech recognition robustness problem, and suggests ways that can make performance improvements in adverse conditions and unexpected speaker changes. Solutions involving Autoregressive Time-Delayed Neural Networks (AR-TDNN), genetic algorithms and Karhunen Loève transforms are explained and experimentally evaluated.

It is my hope that this contribution will both inspire and succeed in passing on to the reader my continued fascination with speech processing and soft computing.

Shippagan (NB), Canada

Sid-Ahmed Selouani

Contents

- 1 Introduction 1**
 - 1.1 Soft Computing Paradigm 1
 - 1.2 Soft Computing in Speech Processing 2
 - 1.3 Organization of the Book 2
 - 1.4 Note to the Reader 4
- Part I Soft Computing and Speech Enhancement**
- 2 Speech Enhancement Paradigm..... 7**
 - 2.1 Speech Enhancement Usefulness..... 7
 - 2.2 Noise Characteristics and Estimation 8
 - 2.2.1 Noise Characteristics 8
 - 2.2.2 Noise Estimation..... 9
 - 2.3 Overview of Speech Enhancement Methods 10
 - 2.3.1 Spectral Subtractive Techniques 10
 - 2.3.2 Statistical-model-based Techniques..... 10
 - 2.3.3 Subspace Decomposition Techniques 11
 - 2.3.4 Perceptual-based Techniques..... 12
 - 2.4 Evaluation of Speech Enhancement Algorithms 12
 - 2.4.1 Time-Domain Measures 13
 - 2.4.2 Spectral Domain Measures 13
 - 2.4.3 Perceptual Domain Measures 13
 - 2.5 Summary 14
- 3 Connectionist Subspace Decomposition for Speech Enhancement 15**
 - 3.1 Method Overview 15
 - 3.2 Definitions 16
 - 3.3 Eigenvalue Decomposition 16
 - 3.4 Singular Value Decomposition 18
 - 3.5 KLT Model Identification in the Mel-scaled Cepstrum 19

3.6	Two-Stage Noise Removal Technique	21
3.7	Experiments	22
3.8	Summary	24
4	Variance of the Reconstruction Error Technique	25
4.1	General Principle	25
4.2	KLT Speech Enhancement using VRE Criterion	26
4.2.1	Optimized VRE	27
4.2.2	Signal Reconstruction	28
4.3	Evaluation of the KLT-VRE Enhancement Method	29
4.3.1	Speech Material	29
4.3.2	Baseline Systems and Comparison Results	29
4.4	Summary	32
5	Evolutionary Techniques for Speech Enhancement	33
5.1	Principle of the Method	33
5.2	Global Framework of Evolutionary Subspace Filtering Method	34
5.3	Hybrid KLT-GA Enhancement	34
5.3.1	Solution Representation	35
5.3.2	Selection Function	35
5.3.3	Crossover and Mutation	36
5.4	Objective Function and Termination	37
5.5	Experiments	37
5.5.1	Speech Databases	38
5.5.2	Experimental Setup	38
5.5.3	Performance Evaluation	39
5.6	Summary	40
 Part II Soft Computing and Automatic Speech Recognition		
6	Robustness of Automatic Speech Recognition	43
6.1	Evolution of Speech Recognition Systems	43
6.2	Speech Recognition Problem	44
6.3	Robust Representation of Speech Signals	46
6.3.1	Cepstral Acoustic Features	46
6.3.2	Robust Auditory-Based Phonetic Features	47
6.4	ASR Robustness	52
6.4.1	Signal compensation techniques	53
6.4.2	Feature Space Techniques	53
6.4.3	Model Space Techniques	54
6.5	Speech Recognition and Human-Computer Dialog	57
6.5.1	Dialog Management Systems	58
6.5.2	Dynamic Pattern Matching Dialog Application	59
6.6	ASR Robustness and Soft Computing Paradigm	61
6.7	Summary	62

7 Artificial Neural Networks and Speech Recognition	63
7.1 Related Work	63
7.2 Hybrid HMM/ANN Systems	64
7.3 Autoregressive Time-Delay Neural Networks	65
7.4 AR-TDNN vs. TDNN	67
7.5 HMM/AR-TDNN Hybrid Structure.....	68
7.6 Experiment and results	69
7.6.1 Speech Material and Tools	70
7.6.2 Setup of the Classification Task	71
7.6.3 Discussion	72
7.7 Summary	73
8 Evolutionary Algorithms and Speech Recognition	75
8.1 Expected Advantages	75
8.2 Problem Statement	76
8.3 Multi-Stream Statistical Framework	77
8.4 Hybrid KLT-VRE-GA-based Front-End Optimization.....	78
8.5 Evolutionary Subspace Decomposition using Variance of Reconstruction Error.....	79
8.5.1 Individuals' Representation and Initialization.....	79
8.5.2 Selection Function	80
8.5.3 Objective Function.....	81
8.5.4 Genetic Operators and Termination Criterion	81
8.6 Experiments and Results.....	82
8.6.1 Speech Material.....	82
8.6.2 Recognition Platform	83
8.6.3 Tests & Results	83
8.7 Summary	85
9 Speaker Adaptation Using Evolutionary-based Approach	87
9.1 Speaker Adaptation Approaches	87
9.2 MPE-based Discriminative Linear Transforms for Speaker Adaptation	88
9.3 Evolutionary Linear Transformation Paradigm.....	90
9.3.1 Population Initialization	91
9.3.2 Objective Function.....	92
9.3.3 Selection Function	92
9.3.4 Recombination	93
9.3.5 Mutation.....	94
9.3.6 Termination	94
9.4 Experiments	95
9.4.1 Resources and Tools	95
9.4.2 Genetic Algorithm Parameters	95
9.4.3 Result Discussion	95
9.5 Summary	96
References.....	97

