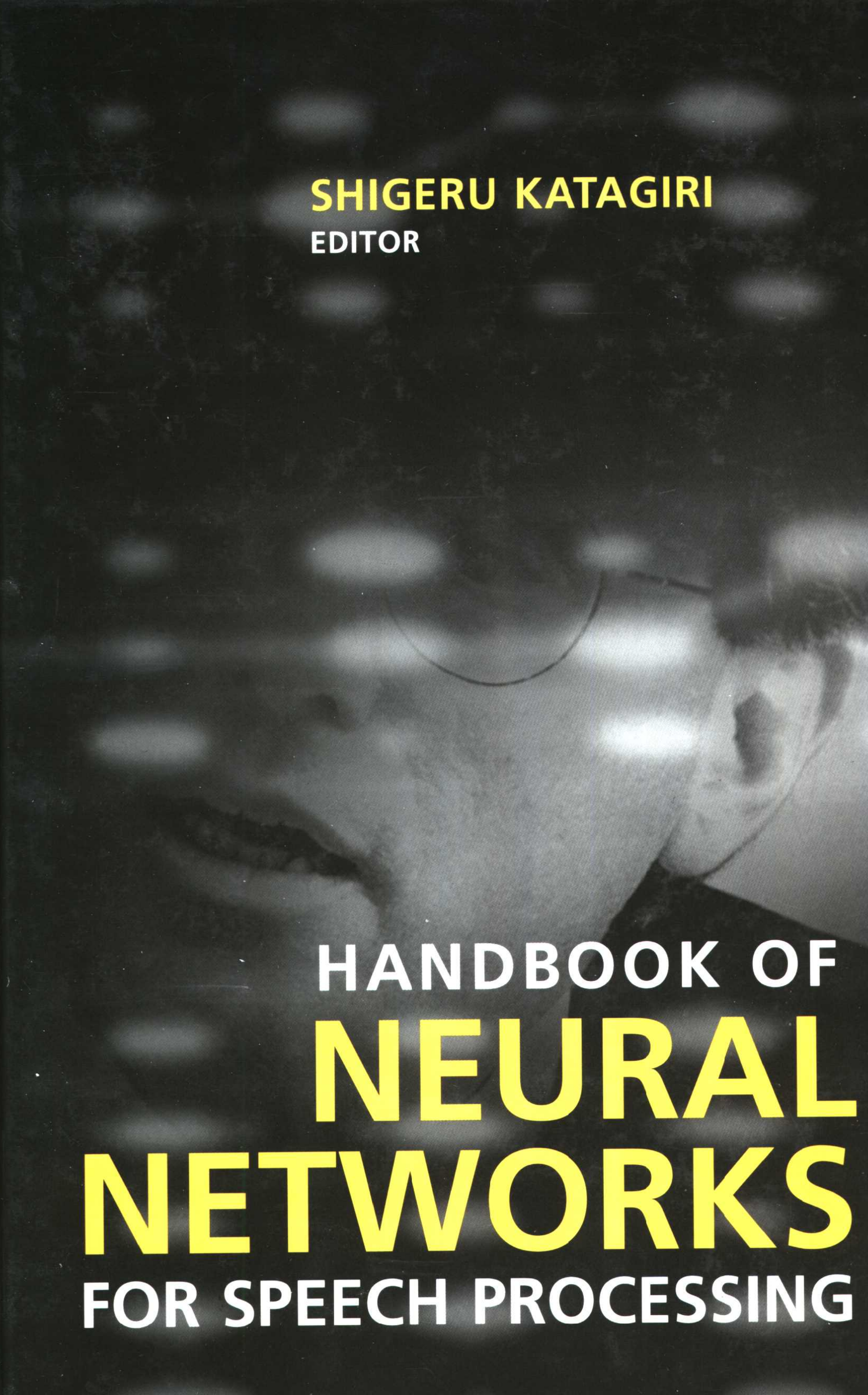

SHIGERU KATAGIRI
EDITOR

HANDBOOK OF NEURAL NETWORKS FOR SPEECH PROCESSING

Handbook of Neural Networks for Speech Processing

Shigeru Katagiri

Editor

Artech House
Boston • London
www.artechhouse.com

Library of Congress Cataloging-in-Publication Data
Handbook of neural networks for speech processing / Shigeru Katagiri, editor.
p. cm. — (Artech House signal processing library)
Includes bibliographical references and index.
ISBN 0-89006-954-9 (alk. paper)
1. Automatic speech recognition—Handbooks, manuals, etc. 2. Neural networks (Computer science)—Handbooks, manuals, etc. 3. Speech processing systems—Handbooks, manuals, etc. I. Katagiri, Shigeru. II. Series.

TK7895.S65 H36 2000 00-055883
006.4'54—dc21 CIP

British Library Cataloguing in Publication Data
Handbook of neural networks for speech processing
(Artech House signal processing library)
1. Neural networks (Computer science) 2. Speech processing systems 3. Speech perception
I. Katagiri, Shigeru
006.5'4

ISBN 0-89006-954-9

Cover design by Igor Valdman

© 2000 ARTECH HOUSE, INC.
685 Canton Street
Norwood, MA 02062

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

International Standard Book Number: 0-89006-954-9
Library of Congress Catalog Card Number: 00-055883

10 9 8 7 6 5 4 3 2 1

Handbook of Neural Networks for Speech Processing

For a listing of recent titles in the *Artech House Signal Processing Library*, turn to the back of this book.

此为试读，需要完整PDF请访问：www.ertongbook.com

Preface

Speech is one of the most important means of human communication. Human beings communicate with one another by speaking and hearing, and they have made great efforts to achieve freer and more convenient communication by developing speech telecommunications technology such as telephony. In recent years, such technology has greatly advanced. The dream of communication—speaking with anybody, anywhere, anytime—is now becoming a reality. The dream, however, continues to expand. A more economic telecommunications system is desired. Even language barriers are expected to be overcome. Morever, work is being carried out to develop human-machine interfaces that use speech as a communications medium. Indeed, satisfying these demands is of great significance. Global and borderless speech communications assist daily human life, promote mutual understanding among nations, and greatly contribute to the general progress of human society. In pursuing these challenges, a key goal is to develop a useful engineering model of the human speech processing mechanism. This handbook seeks to provide a comprehensive introduction to one of the most important approaches to this technological challenge, speech processing using artificial neural networks (ANNs).

ANNs and speech processing are both transdisciplinary research fields, each having a history of several decades and involving many disciplines such as physiology, physics, statistics, psychology, linguistics, and engineering. Accordingly, a single book cannot be expected to fully cover the entire subject of speech processing using ANNs.

Fortunately, many excellent textbooks have been published in both the ANNs and speech processing fields. Readers can find detailed expertise in such books. In this handbook, we shall concentrate on the results of emerging and challenging studies on using ANNs for speech processing. We shall strive to convey the excitement of developments on the research front. The handbook will guide its readers through the unfinished but promising technological strides made in this field.

The book's 13 chapters are divided into three parts. Part I, entitled "Fundamentals," consists of 4 chapters that introduce basic information about speech processing and ANN technologies. In this introductory section, however, our focus is on speech processing technologies; the fundamentals of ANNs are discussed in Parts II and III. Part II, "Current Issues in Speech Recognition," emphasizes the considerable vigor with which ANNs have been investigated in their application to speech recognition. This second part consists of Chapters 5 through 9. Finally, Part III, "Current Issues in Speech Signal Processing," discusses speech-related topics such as speaker recognition, voice conversion, speech coding, and speech enhancement.

The handbook is aimed at researchers, engineers, and graduate-level students who wish to study the fundamentals and practical applications of neural-network-based speech processing. For these readers, Part I provides the necessary basis for proceeding to the later chapters on applications; Parts II and III provide a comprehensive introduction to the research front of neural networks for speech processing, which can be useful for further study of topics elaborated in related technical books and journals. Despite the primary aim of providing an introductory text, the book can also be used by researchers who have experience in ANN and speech processing technologies. The two latter parts provide good archives of results of the ongoing research. Each chapter was written with the aim of being as self-contained as possible. The book can therefore be used for various types of study by various readers. However, we, the chapter contributors, have a common motivation: we hope that the handbook becomes a vehicle to stimulate research by scientists and engineers on the human mechanism of speech processing and its engineering embodiment, i.e., neural networks for speech processing.

The inspiration for this handbook came from the series editor, Professor Alexsander D. Poularikas, University of Alabama in Huntsville. I wish to express my sincere thanks to him for his careful planning and for asking me to serve as editor of this text. The value of the book derives from the considerable efforts of the contributors, each of whom brought extensive experience to the task of writing the individual chapters. I would like to express my considerable gratitude to these authors. Finally, it would not have been possible to produce the book without the patience and professional assistance provided by the crew at Artech House, especially Mark Walsh, Alexia Rosoff, Traci Beane, and Barbara Lovenvirth. I wish to thank them for help in all phases of making this book a reality.

—Shigeru Katagiri

Recent Titles in the Artech House Signal Processing Library

Computer Speech Technology, Robert D. Rodman

Handbook of Neural Networks for Speech Processing, Shigeru Katagiri, editor

Hilbert Transforms in Signal Processing, Stefan L. Hahn

Phase and Phase-Difference Modulation in Digital Communications, Yuri Okunev

Signals, Oscillations, and Waves: A Modern Approach, David Vakman

Statistical Signal Characterization, Herbert L. Hirsch

Statistical Signal Characterization Algorithms and Analysis Programs, Herbert L. Hirsch

Voice Recognition, Richard L. Klevans and Robert D. Rodman

For further information on these and other Artech House titles, including previously considered out-of-print books now available through our In-Print-Forever® (IPF®) program, contact:

Artech House
685 Canton Street
Norwood, MA 02062
Phone: 781-769-9750
Fax: 781-769-6334
e-mail: artech@artechhouse.com

Artech House
46 Gillingham Street
London SW1V 1AH UK
Phone: +44 (0)171-973-8077
Fax: +44 (0)171-630-0166
e-mail: artech-uk@artechhouse.com

Find us on the World Wide Web at:
www.artechhouse.com

Contents

Preface xxi

Part I Fundamentals

1 Introduction 3

1.1 Speech Processing 3

1.2 Neural Networks 6

1.2.1 Fundamentals 6

1.2.2 Taxonomy of Neural Networks 8

1.2.2.1 Overview 8

1.2.2.2 Structure 9

1.2.2.3 Measurement 11

1.2.2.4 Objective Function 13

1.2.2.5 Optimization 14

1.3 Neural Networks for Speech Processing 15

1.4 Handbook Overview 15

1.4.1 Part I: Fundamentals 15

1.4.2 Part II: Current Issues in Speech Recognition 16

1.4.3 Part III: Current Issues in Speech Signal Processing 17

References 17

2 The Speech Signal and Its Production Model 19

2.1 Introduction 19
2.2 Information Conveyed by Speech 21
2.2.1 Linguistic Information 21
2.2.1.1 Segmental Features 21
2.2.1.2 Suprasegmental Features 22
2.2.2 Paralinguistic Information 23
2.2.3 Nonlinguistic Information 27
2.2.3.1 Idiosyncratic Factors 27
2.2.3.2 Emotional Factors 29
2.2.4 Hierarchical Speech Production Processes 30
2.2.5 Summary 31
2.3 Physical and Physiological Processes in Speech Production 32
2.3.1 Respiration System 32
2.3.1.1 Normal Breathing Without Speech Production 32
2.3.1.2 Expiration in Speech Production 34
2.3.2 Phonatory System 34
2.3.2.1 Framework of the Larynx 34
2.3.2.2 Abduction versus Adduction 34
2.3.2.3 F0 Control During Speech 37
2.3.3 Articulatory System 39
2.3.3.1 Morphology of Articulators 39
2.3.3.2 Vowel Production 40
2.3.3.3 Consonant Production 41
2.3.4 Summary 46
2.4 Models and Theories of Speech Production 49
2.4.1 Laryngeal System 49
2.4.1.1 Vocal Fold Vibration 49
2.4.1.2 F0 Control in Running Speech Production 50
2.4.1.3 Vertical Movements of the Larynx 51
2.4.2 Dynamic Characteristic of Articulators 52

此为试读，需要完整PDF请访问：www.ertongbook.com

2.4.2.1 Models of Individual Articulators 52
2.4.2.2 Articulatory Models of Speech Production 54
2.4.3 Summary 56
2.5 Conclusion 57
Acknowledgment 58
References 58

3 **Speech Recognition** **63**

3.1 Introduction 63
3.2 Overview 64
3.2.1 Hearing and Machine Recognition 64
3.2.2 Recognition-Oriented Speech Feature Representation 66
3.2.2.1 Sound Spectrogram: Time-Frequency-Energy Representation 66
3.2.2.2 Acoustic Feature Vector 69
3.2.2.3 Static Versus Dynamic Nature 72
3.2.3 Variety of Recognition Tasks 73
3.2.4 Recognition Mechanism 73
3.2.4.1 Example Task Setting 73
3.2.4.2 Distance-Based Recognition 76
3.2.4.3 Distance Computation Based on Dynamic Time Warping 77
3.2.4.4 Remarks 80
3.3 Bayes Decision Theory 80
3.3.1 Overview 80
3.3.2 Maximum Likelihood Estimation Approach 83
3.3.3 Bayesian Approach 85
3.3.4 Discriminant Function Approach 86
3.3.4.1 Example Task and Decision Rule 86
3.3.4.2 Loss 88
3.3.4.3 Optimization 88
3.3.4.4 Design of Linear Discriminant Function Classifier 90

3.3.4.5 Remarks 91
3.4 Acoustic Feature Extraction 91
3.4.1 Filter-Bank 91
3.4.1.1 Artificial Cochlea Filter 93
3.4.1.2 Fourier-Transform-Based Filter 94
3.4.2 Autoregressive Modeling 94
3.4.3 Cepstrum Modeling 98
3.4.4 Dynamic Feature Modeling 101
3.5 Probabilistic Acoustic Modeling Based on Hidden Markov Model 103
3.5.1 Principles of Hidden Markov Model 103
3.5.2 Selection of Output Probability Function 106
3.5.2.1 Discrete Model 106
3.5.2.2 Continuous Model 106
3.5.3 MLE-Based Design Method 109
3.5.3.1 Forward-Backward Method 109
3.5.4 Trellis Algorithm and Viterbi Algorithm 111
3.5.5 Discriminative Design Methods 112
3.6 Language Modeling 112
3.6.1 Role of Language Modeling 112
3.6.2 *N*-Gram Language Modeling 114
3.7 Concluding Remarks 115
3.7.1 Overview 115
3.7.2 Selection of Model Units 115
3.7.3 Open-Vocabulary Recognition 116
3.7.4 Bibliographical Remarks 117
References 117

4 Speech Coding 121

4.1 Introduction 121
4.2 Attributes of Speech Coders 122
4.3 Basic Principles of Speech Coders 123
4.4 Quantization 126

4.4.1 Scalar Quantization 126
4.4.2 Vector Quantization 126
4.5 Linear Prediction 128
4.5.1 Linear Prediction Principles 128
4.5.2 Speech Coding Based on Linear Prediction 129
4.5.3 The Analysis-by-Synthesis Principle 131
4.5.4 Perceptual Filtering 135
4.5.5 Quantization of the Linear Prediction Coefficients 136
4.6 Sinusoidal Coding 141
4.7 Waveform Interpolation Methods 142
4.8 Subband Coding 143
4.9 Variable-Rate Coding 144
4.9.1 Basics 144
4.9.2 Phonetic Segmentation 145
4.9.3 Variable Rate Coders for ATM Networks 145
4.9.4 Voice over IP 146
4.10 Wideband Coders 146
4.11 Measuring Speech Coder Performance 147
4.12 Speech Coding over Noisy Channels 150
4.13 Speech Coding Standards 151
4.14 Conclusions 152
References 153

Part II Current Issues in Speech Recognition

5 Discriminative Prototype-Based Methods for Speech Recognition 159

5.1 Introduction 159
5.2 Bayes Decision Theory 161
5.2.1 The Bayes Decision Rule 161
5.2.2 Discriminant Functions 161
5.2.3 Discriminant Functions for Prototype-Based Methods 163

5.3 Example-Based Methods 165
5.3.1 Density Estimation 166
5.3.2 Estimation of Posterior Probabilities 167
5.3.3 The *k*-Nearest-Neighbor Method 168
5.3.3.1 The Nearest-Neighbor Rule 168
5.3.3.2 Error Bounds for *k*-Nearest-Neighbor Classification 168
5.3.4 Parzen Windows 168
5.3.4.1 An Example 170
5.3.5 Advantages and Limitations of Example-Based Methods 171
5.3.6 Smoothing 172
5.3.7 Applications to Speech Recognition 173
5.4 Prototype-Based Methods for Speech Recognition 173
5.5 Prototype-Based Classifier Design Using Minimum Classification Error 175
5.5.1 Definition of Discriminant Function 175
5.5.2 Definition of Misclassification Measure 176
5.5.3 Definition of Local Loss Function 176
5.5.4 Overall Loss Function and Optimization 177
5.5.5 Modified Newton's Method: The Quickprop Algorithm 178
5.5.6 Relation of MCE Loss to the Bayes Error 181
5.5.7 Choice of Smoothing Parameters for MCE-Based Optimization 182
5.6 Learning Vector Quantization 182
5.6.1 Shift-Tolerant LVQ for Speech Recognition 185
5.6.1.1 HMM Interpretation of STLVQ 187
5.6.1.2 Limitations and Strengths of STLVQ Architecture and Training 187
5.6.2 Expanding the Scope of LVQ for Speech Recognition: Incorporation into Hidden Markov Modeling 188
5.6.2.1 LVQ-HMM 189

5.6.2.2 HMM-LVQ 190

5.6.3 Minimum Classification Error Interpretation of LVQ 192

5.6.4 Smoothness of MCE Loss 193

5.6.5 LVQ Summary 195

5.7 Prototype-Based Methods Using Dynamic Programming 195

5.7.1 MCE-Trained Prototypes for DTW-Based Speech Recognition 196

5.7.1.1 Practical Implementation of MCE/GPD 198

5.7.1.2 MCE-DTW Results 199

5.7.2 Prototype-Based Minimum Error Classifier 199

5.7.2.1 PBMEC State Distance and Discriminant Function 200

5.7.2.2 MCE/GPD in the Context of Speech Recognition Using Phoneme Models 202

5.7.2.3 PBMEC Results 202

5.7.3 Summary of Prototype-Based Methods Using DP 203

5.8 Hidden Markov Model Design Based on MCE 204

5.8.1 HMM State Likelihood and Discriminant Function 206

5.8.2 MCE Misclassification Measure and Loss 207

5.8.3 Calculation of MCE Gradient for HMMs 207

5.8.3.1 Derivative of Loss with Respect to Misclassification Measure 207

5.8.3.2 Derivative of Misclassification Measure with Respect to Discriminant Functions 208

5.8.3.3 Derivative of Discriminant Function with Respect to Observation Probability Density Function 208

5.8.3.4 Derivative of Observation Probability with Respect to Mixing Weights 209

5.8.3.5 Derivative of Observation Probability with Respect to Mean Vectors 209

5.8.3.6 Derivative of Observation Probability with Respect to Covariances 210
5.8.3.7 Application of the Chain Rule 210
5.8.4 MCE-HMM Results 211
5.9 Conclusion 212
References 213

6 Recurrent Neural Networks for Speech Recognition 217

6.1 Introduction 217
6.1.1 Background and Motivation 217
6.1.2 Chapter Overview 219
6.2 Speech Recognition Theory 220
6.3 Basics of Neural Networks 223
6.3.1 Parameter Estimation by Maximum Likelihood 224
6.3.2 Problem Classification 225
6.3.2.1 Regression 225
6.3.2.2 Classification 226
6.3.3 Neural Network Training 226
6.3.3.1 Gradient Descent Training 227
6.3.3.2 RPROP Training 228
6.3.3.3 ARPROP Training 229
6.3.4 Neural Network Architectures 230
6.3.4.1 Multilayer Perceptrons 232
6.3.4.2 Time-Delay Neural Networks 232
6.4 Recurrent Neural Networks 232
6.4.1 Unidirectional Recurrent Neural Network 233
6.4.1.1 RNN Architecture 233
6.4.1.2 RNN Training 235
6.4.2 Bidirectional Recurrent Neural Network 235
6.4.2.1 BRNN Architecture 235
6.4.2.2 BRNN Training 236
6.5 Modeling Phonetic Context 237

此为试读，需要完整PDF请访问：www.ertongbook.com