

---

# **NEURAL NETWORKS AND SPEECH PROCESSING**

---

# THE KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE

## VLSI, COMPUTER ARCHITECTURE AND DIGITAL SIGNAL PROCESSING

*Consulting Editor*  
**Jonathan Allen**

- Unified Methods for VLSI Simulation and Test Generation**, K.T. Cheng,  
V.D. Agrawal, ISBN: 0-7923-9025-3
- ASIC System Design With VHDL: A Paradigm**, S.S. Leung, M.A. Shanblatt,  
ISBN: 0-7923-9032-6
- BiCMOS Technology and Applications**, A. R. Alvarez, Editor  
ISBN: 0-7923-9033-4
- Analog VLSI Implementation of Neural Systems**, C. Mead, M. Ismail (Editors),  
ISBN: 0-7923-9040-7
- The MIPS-X RISC Microprocessor**, P. Chow. ISBN: 0-7923-9045-8
- Nonlinear Digital Filters: Principles and Applications**, I. Pitas, A.N.  
Venetsanopoulos, ISBN: 0-7923-9049-0
- Algorithmic and Register-Transfer Level Synthesis: The System Architect's  
Workbench**, D.E. Thomas, E.D. Lagnese, R.A. Walker, J.A. Nestor, J.V. Ragan,  
R.L. Blackburn, ISBN: 0-7923-9053-9
- VLSI Design for Manufacturing: Yield Enhancement**, S.W. Director, W. Maly,  
A.J. Strojwas, ISBN: 0-7923-9053-7
- Testing and Reliable Design of CMOS Circuits**, N.K. Jha, S. Kundu,  
ISBN: 0-7923-9056-3
- Hierarchical Modeling for VLSI Circuit Testing**, D. Bhattacharya, J.P.  
Hayes, ISBN: 0-7923-9058-X
- Steady-State Methods for Simulating Analog and Microwave Circuits**,  
K. Kundert, A. Sangiovanni-Vincentelli, J. White,  
ISBN: 0-7923-9069-5
- Introduction to Analog VLSI Design Automation**, M. Ismail, J. Franca,  
ISBN: 0-7923-9102-0
- Gallium Arsenide Digital Circuits**, O. Wing, ISBN: 0-7923-9081-4
- Principles of VLSI System Planning**, A.M. Dewey ISBN: 0-7923-9102
- Mixed-Mode Simulation**, R. Saleh, A.R. Newton, ISBN: 0-7923-9107-1
- Automatic Programming Applied to VLSI CAD Software: A Case Study**,  
D. Setliff, R.A. Rutenbar, ISBN: 0-7923-9112-8
- Models for Large Integrated Circuits**, P. Dewilde, Z.Q. Ning  
ISBN: 0-7923-9115-2
- Hardware Design and Simulation in VAL/VHDL**, L.M. Augustin, D.C. Luckham,  
B.A. Gennart, Y. Huh, A.G. Stanculescu ISBN: 0-7923-9087-3
- Subband Image Coding**, J. Woods, editor, ISBN: 0-7923-9093-8
- Low-Noise Wide-Band Amplifiers in Bipolar and CMOS Technologies**, Z.Y. Chang,  
W.M.C. Sansen, ISBN: 0-7923-9096-2
- Iterative Identification and Restoration Images**, R. L. Lagendijk, J. Biemond  
ISBN: 0-7923-9097-0
- VLSI Design of Neural Networks**, U. Ramacher, U. Ruckert  
ISBN: 0-7923-9127-6
- Synchronization Design for Digital Systems**, T. H. Meng ISBN: 0-7923-9128-4
- Hardware Annealing in Analog VLSI Neurocomputing**, B. W. Lee, B. J. Sheu  
ISBN: 0-7923-9132-2

---

# NEURAL NETWORKS AND SPEECH PROCESSING

by

**David P. Morgan**  
**Christopher L. Scofield**

*Foreword by Leon N. Cooper*



Springer Science+Business Media, LLC

**Library of Congress Cataloging-in-Publication Data**

**Morgan, David P., 1961-**

Neural networks and speech processing / by David P. Morgan,  
Christopher L. Scofield ; foreword by Leon N. Cooper.

p. cm. — (The Kluwer international series in engineering and  
computer science. VLSI, computer architecture, and digital signal  
processing)

Includes bibliographical references and index.

ISBN 978-1-4613-6763-5 ISBN 978-1-4615-3950-6 (eBook)

DOI 10.1007/978-1-4615-3950-6

1. Neural networks (Computer science) 2. Speech processing  
systems. I. Scofield, Christopher L., 1957- II. Title.

III. Series.

QA76.87.M67 1991

006.3—dc20

90-27214

CIP

---

**Copyright** © 1991 by Springer Science+Business Media New York

Originally published by Kluwer Academic Publishers in 1991

Softcover reprint of the hardcover 1st edition 1991

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, mechanical, photo-copying, recording, or otherwise, without the prior written permission of the publisher, Springer Science + Business Media, LLC

*Printed on acid-free paper.*

*To our wives, Christine and Dale,  
for their support and encouragement.*

# Contents

<b>Preface</b>	<b>xiii</b>
<b>Foreword</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 A Few Words on Speech Recognition . . . . .	3
1.3 A Few Words on Neural Networks . . . . .	4
1.4 Contents . . . . .	6
<b>2 The Mammalian Auditory System</b>	<b>9</b>
2.1 Introduction to Auditory Processing . . . . .	9
2.2 The Anatomy and Physiology of Neurons . . . . .	9
2.3 Neuroanatomy of the Auditory System . . . . .	12
2.3.1 The Ear . . . . .	13
2.3.2 The Cochlea . . . . .	15
2.3.3 The Eighth Nerve . . . . .	18
2.3.4 The Cochlear Nucleus . . . . .	24
2.3.5 The Superior Olivary Complex . . . . .	28
2.3.6 The Inferior Colliculus . . . . .	31
2.3.7 The Medial Geniculate Nucleus . . . . .	32
2.3.8 The Auditory Cortex . . . . .	35
2.4 Recurrent Connectivity in the Auditory Pathway . . . . .	37
2.5 Summary . . . . .	39

<b>3</b>	<b>An Artificial Neural Network Primer</b>	<b>41</b>
3.1	A Neural Network Primer for Speech Scientists . . . . .	41
3.2	Elements of Artificial Neural Networks . . . . .	42
3.2.1	Similarity Measures and Activation Functions . . .	43
3.2.2	Networks and Mappings . . . . .	48
3.3	Learning in Neural Networks . . . . .	50
3.4	Supervised Learning . . . . .	51
3.4.1	The Perceptron and Gradient-Descent Learning . .	51
3.4.2	Associative Memories . . . . .	54
3.4.3	The Hopfield Network . . . . .	57
3.5	Multi-Layer Networks . . . . .	59
3.5.1	The Restricted Coulomb Energy Network . . . . .	63
3.5.2	The Backward Error Propagation Network . . . .	69
3.5.3	The Charge Clustering Network . . . . .	75
3.5.4	Recurrent Back Propagation . . . . .	78
3.6	Unsupervised Learning . . . . .	80
3.6.1	The BCM Network . . . . .	81
3.6.2	The Kohonen Feature Map . . . . .	84
3.7	Summary . . . . .	88
<b>4</b>	<b>A Speech Technology Primer</b>	<b>91</b>
4.1	A Speech Primer for Neural Scientists . . . . .	91
4.2	Human Speech Production/Perception . . . . .	92
4.2.1	Information in the Speech Signal . . . . .	96
4.3	ASR Technology . . . . .	102
4.3.1	A General Speech Recognition Model . . . . .	103
4.4	Signal Processing and Feature Extraction . . . . .	106
4.4.1	Linear Predictive Coding . . . . .	110
4.4.2	Feature Extraction and Modeling . . . . .	113
4.4.3	Vector Quantization . . . . .	114
4.5	Time Alignment and Pattern Matching . . . . .	115
4.5.1	Dynamic Time Warping . . . . .	117
4.5.2	Hidden Markov Models . . . . .	119
4.5.3	Pronunciation Network Word Models . . . . .	126
4.6	Language Models . . . . .	127
4.6.1	Parsers . . . . .	128
4.6.2	Statistical Models . . . . .	129

4.7	Summary . . . . .	130
<b>5</b>	<b>Methods in Neural Network Applications</b>	<b>131</b>
5.1	The Allure of Neural Networks for Speech Processing . . .	131
5.2	The Computational Properties of ANNs . . . . .	132
5.2.1	Computability and Network Size . . . . .	135
5.3	ANN Limitations: The Scaling Problem . . . . .	141
5.3.1	The Scaling of Learning . . . . .	142
5.3.2	The Scaling of Generalization . . . . .	143
5.4	Structured ANN Solutions . . . . .	145
5.4.1	Hierarchical Modules . . . . .	145
5.4.2	Hybrid Systems . . . . .	149
5.4.3	Multiple Neural Network Systems . . . . .	156
5.4.4	Integrating Neural Speech Modules . . . . .	158
5.5	Summary . . . . .	161
<b>6</b>	<b>Signal Processing and Feature Extraction</b>	<b>163</b>
6.1	The Importance of Signal Representations . . . . .	163
6.2	The Signal Processing Problem Domain . . . . .	163
6.3	Biologically Motivated Signal Processing . . . . .	166
6.3.1	Review of Speech Representation in the Auditory Nerve . . . . .	166
6.3.2	The Silicon Cochlea and Temporal-Place Representations for ASR . . . . .	169
6.3.3	The Role of Automatic Gain Control in Noisy Environments . . . . .	173
6.4	ANNs for Conventional Signal Processing . . . . .	179
6.4.1	Adaptive Filtering . . . . .	179
6.4.2	A Noise Reduction Network . . . . .	183
6.5	Feature Representations . . . . .	190
6.5.1	Unsupervised Feature Extraction for Phoneme Clas- sification . . . . .	191
6.5.2	Feature Maps . . . . .	196
6.6	Summary . . . . .	201
<b>7</b>	<b>Time Alignment and Pattern Matching</b>	<b>203</b>
7.1	Modeling Spectro-Temporal Structure . . . . .	203
7.2	Time Normalization Via Pre-Processing . . . . .	204
7.2.1	Interpolation and Decimation Techniques . . . . .	204



7.2.2	Feature-Set Transformations . . . . .	206
7.3	The Dynamic Programming Neural Network . . . . .	209
7.3.1	The DPNN Architecture . . . . .	210
7.3.2	The Time Warping Structure . . . . .	211
7.3.3	The DPNN Training Procedure . . . . .	213
7.3.4	Application to Speaker-Independent Digit Recognition . . . . .	214
7.4	HMM Motivated Networks . . . . .	215
7.4.1	The Viterbi Network . . . . .	216
7.4.2	The HMM Network . . . . .	220
7.5	Recurrent Networks for Temporal Modeling . . . . .	226
7.5.1	The Temporal Flow Model . . . . .	226
7.5.2	Temporal Flow Experiments . . . . .	228
7.6	The Time Delay Neural Network . . . . .	230
7.6.1	The TDNN Temporal Architecture . . . . .	231
7.6.2	TDNN Training . . . . .	234
7.6.3	Application to Phoneme Classification . . . . .	234
7.6.4	Interpreting the TDNN Spectro-Temporal Representation . . . . .	236
7.6.5	Phoneme Classification Summary . . . . .	240
7.6.6	TDNNs for Word Discrimination . . . . .	242
7.7	Summary . . . . .	243
<b>8</b>	<b>Natural Language Processing</b>	<b>245</b>
8.1	The Importance of Language Processing . . . . .	245
8.2	Syntactic Models . . . . .	248
8.2.1	NETgrams: An ANN Word Category Predictor . .	253
8.2.2	An ANN for Word Category Disambiguation . . .	260
8.2.3	Recurrent Networks and Formal Languages . . .	264
8.3	Semantic Models . . . . .	273
8.3.1	Pronoun Reference ANNs . . . . .	274
8.4	Knowledge Representation . . . . .	277
8.4.1	Knowledge Representation in a Hopfield Network .	279
8.5	Summary . . . . .	288
<b>9</b>	<b>ANN Keyword Recognition</b>	<b>289</b>
9.1	Keyword Spotting . . . . .	289
9.2	The Primary KWS System . . . . .	291
9.2.1	Experimental Data . . . . .	294

- 9.3 DUR Experiments . . . . . 296
  - 9.3.1 Selecting a Fixed-Length Feature Representation . 296
  - 9.3.2 Single and Multiple Networks . . . . . 297
  - 9.3.3 Experiments with Hybrid Systems . . . . . 302
- 9.4 Secondary Processing Experiments . . . . . 305
  - 9.4.1 The Pattern Matching Approach . . . . . 305
  - 9.4.2 An Investigation of Temporal Models . . . . . 313
- 9.5 Summary . . . . . 326
- 10 Neural Networks and Speech Processing . . . . . 329**
  - 10.1 Speech Processing Applications . . . . . 329
    - 10.1.1 Speech Synthesis . . . . . 331
    - 10.1.2 Speech Coding . . . . . 334
    - 10.1.3 Speaker Separation . . . . . 335
    - 10.1.4 Speech Enhancement . . . . . 336
    - 10.1.5 Speaker Verification/Identification . . . . . 337
    - 10.1.6 Language Identification . . . . . 338
    - 10.1.7 Keyword/Keyphrase Spotting . . . . . 338
  - 10.2 Summary of Efforts in ASR . . . . . 339
    - 10.2.1 The Past: Institutions Involved in ASR . . . . . 340
    - 10.2.2 The Current Status of ANNs in ASR . . . . . 342
    - 10.2.3 The Future: Challenges and Goals . . . . . 345
  - 10.3 Concluding Remarks . . . . . 348
- A Faster Back Propagation . . . . . 349**
  - A.1 Methods for the Successful Application of ANNs . . . . . 349
    - A.1.1 Gradient-Descent Heuristics . . . . . 349
    - A.1.2 Second-Order Methods . . . . . 353
- B Acronyms . . . . . 357**
- Bibliography . . . . . 359**
- Index . . . . . 383**

## Preface

We would like to take this opportunity to thank all of those individuals who helped us assemble this text, including the people of Lockheed Sanders and Nestor, Inc., whose encouragement and support were greatly appreciated. In addition, we would like to thank the members of the *Laboratory for Engineering Man-Machine Systems* (LEMS) and the *Center for Neural Science* at Brown University for their frequent and helpful discussions on a number of topics discussed in this text. Although we both attended Brown from 1983 to 1985, and had offices in the same building, it is surprising that we did not meet until 1988.

We also wish to thank Kluwer Academic Publishers for their professionalism and patience, and the reviewers for their constructive criticism. Thanks to John McCarthy for performing the final proof, and to John Adcock, Chip Bachmann, Deborah Farrow, Nathan Intrator, Michael Perrone, Ed Real, Lance Riek and Paul Zeman for their comments and assistance. We would also like to thank Khrisna Nathan, our most unbiased and critical reviewer, for his suggestions for improving the content and accuracy of this text. A special thanks goes to Steve Hoffman, who was instrumental in helping us perform the experiments described in Chapter 9.

David P. Morgan  
Christopher L. Scofield

## Foreword

Today there is a great deal of interest and excitement in the investigation of artificial neural networks. Yet, when things sort themselves out, neural networks will do less than their most fervent supporters in their most enthusiastic moments suggest. But they will do more than the most pessimistic estimates of their most adamant detractors.

We will not waste time pondering the foolish question as to whether neural networks are either inspired by or imitate the brain. Those of us working with neural networks know that they do not duplicate the brain; however, the brain is where the inspiration for neural networks originated. If nothing else, the fact that the brain exists, learns, remembers, and thinks, is an existence proof that shows us that our ultimate goal is achievable. But how realistic is this goal? This is the “sixty-four dollar” question.

The hope from the beginning has been that we might capture the capacity to learn – which is clearly possible given that it exists in our heads. The question is, what does it take to capture that most sought after ability? Does it require an understanding of how individual neurons learn at a cellular level, how and where memory is stored? Does it take an understanding of reasonably complex but manageable subsystems of the brain – an understanding that could, in principal, be duplicated by either a set of instructions in computer simulation or directly in hardware? Or, is this capacity exhibited only as a property of a very large system of neurons containing very complex subsystems, so that in effect to capture these properties, we will have to do something as complex as reconstruct the entire brain!

Many researchers working with neural networks are currently exploring manageable systems that are in fact very much less complicated than the brain. Hopefully, these systems will capture some of its properties, attacking problems that, if not impossible, have been exceedingly difficult to tackle by what we might call conventional means. Among these, speech recognition ranks among the most difficult. It is a problem that combines acoustic recognition, representation, expectation, context and understanding.

Speech recognition and vision are two areas where neural networks can be motivated by underlying neurobiological structures. Although this text covers a variety of speech processing applications and neural

network topologies, David Morgan and Chris Scofield continually return to this theme. Their text describes the basic structure of the mammalian auditory system and contains sufficient background material so that researchers in related disciplines can become acquainted with the use of neural networks for speech processing.

Every area has its holy grail – in the field of neural networks there are several, and one surely is the recognition/understanding of continuous speech. Our capacity to both speak and understand one another seems so uniquely human that to simulate it in a machine would indicate to most that we had, in fact, captured a significant portion of the processing capacity that is contained in our heads.

Leon N. Cooper

---

# NEURAL NETWORKS AND SPEECH PROCESSING