

**2000 Neural Networks for
Signal Processing X
(V.A)**

TN 911.7-53
N494
2000
V.A

NEURAL NETWORKS FOR SIGNAL PROCESSING X

PROCEEDINGS
OF THE 2000 IEEE
SIGNAL PROCESSING
SOCIETY WORKSHOP

Tenth in a Series of Workshops
Organized by the IEEE Signal Processing Society
Neural Networks Technical Committee

Edited by

Bernard Widrow
Ling Guan
Kuldip Paliwa
Tulay Adali
Jan Larsen
Elizabeth Wilson
Scott Douglas



Published under the sponsorship of the IEEE Signal Processing Society
in cooperation with the IEEE Neural Networks Council
and the Motorola Australian Research Centre

The Institute of Electrical and Electronic Engineers, Inc.
New York, New York



E200200228

Neural Networks for Signal Processing X—Proceedings of the 2000 IEEE Signal Processing Society Workshop

Copyright © 2000 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved.

Copyright and Reprints Permission

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or reproduction requests should be addressed to:
IEEE Copyrights Manager, IEEE Operations Center, 445 Hoes Lane, P.O. Box 1331,
 Piscataway, NJ 08855-1331

IEEE Catalog Number	00TH8501
ISBN	0-7803-6278-0
ISSN	1089-3555

Additional copies of this publication are available from

IEEE Operations Center
P.O. Box 1331
445 Hoes Lane
Piscataway, NJ 08855-1331 USA

800 678 IEEE
732 981 1393
732 981 0600
732 981 9667 (FAX)
Email: customer.service@ieee.org

PREFACE

This book contains refereed papers presented at the Tenth IEEE Workshop on Neural Networks for Signal Processing (NNSP2000) held at University of Sydney, Sydney, Australia, December 11-13, 2000.

The Neural Networks for Signal Processing Technical Committee of the IEEE Signal Processing Society sponsored the NNSP2000 Workshop with co-operation from the IEEE Neural Network Council.

In recent years, the field of neural networks has matured considerably in both methodology and real-world application domains, going far beyond the “standard” backpropagation network and academic examples. Contemporary neural networks for signal processing research shares many ideas of adaptive signal/image processing, machine learning, and advanced statistics. We consider the field of neural networks for signal processing as an open-minded playground that combines many approaches for solving complex real-world signal processing problems.

Continuing the tradition of paperless and easy communication, many of the details of the NNSP2000 Workshop were handled electronically through a re-designed workshop webpage (<http://eivind.imm.dtu.dk/nns2000>) that, among other features, included web-based submission, review, and registration.

This year, the workshop features a special session organized by Shigeru Katagiri on Intelligent Multimedia Processing, an increasingly important application area for neural networks.

Our sincere thanks goes to Professor Tamas Roska of the Hungary Academy of Science, Dr. David Fogel of Natural Selection, and Dr. Brian Ferguson of Defence Science and Technology Organisation, Australia for accepting to give plenary speeches at the workshop. Further, we would like to thank Professor Sun-Yuan Kung of Princeton University for agreeing to present a joint keynote speech on “Adaptive Techniques for Intelligent Internet Multimedia Communication” to the attendees of both the NNSP2000 Workshop and the 1st IEEE Pacific-Rim Conference on Multimedia. Also, we would like to thank the Technical Committee for taking the time to provide quality reviews. Special thanks go to the folks at University of Sydney Signal and Multimedia Processing Laboratory for the careful handling of workshop registration and local arrangements. Above all, thanks to all of the authors who made this workshop possible and successful.

Bernard Widrow, *Stanford University*

Ling Guan, *University of Sydney*

Kuldip Paliwa, *Griffith University*

Tulay Adali, *University of Maryland, Baltimore County*

Jan Larsen, *Technical University of Denmark*

Elizabeth Wilson, *Raytheon Company*

Scott Douglas, *Southern Methodist University*

NNSP 2000 Technical Committee

Amir Assadi	University of Wisconsin at Madison, USA
Yianni Attikiouzel	University of Western Australia, Australia
John Asenstorfer	University of South Australia, Australia
Andrew Back	Katestone Scientific, Brisbane, Australia
Geoff Barton	University of Sydney, Australia
Hervé Bourlard	IDIAP, Switzerland
Andy Chalmers	Manukau Polytechnic, New Zealand
Zheru Chi	Hong Kong Polytechnic University, Hong Kong
Andrzej Cichocki	RIKEN, Japan
Tharam Dillon	La Trobe University, Australia
Tom Downs	University of Queensland, Australia
Hsin Chia Fu	National Chiao-Tung University, Taiwan, ROC
Suresh Hangenahally	Griffith University, Australia
Marwan Jabri	University of Sydney, Australia
Haosong Kong	Motorola Australian Research Centre, Australia
Shigeru Katagiri	NTT and ATR, Japan
Anthony Kuh	University of Hawaii at Manoa, USA
Yi Liu	CSIRO Minerals Division, Australia
Fa-Long Luo	ReSound Corporation, USA
David Miller	Pennsylvania State University, USA
Christophe Molina	The Sanger Centre, Cambridge, UK
M Mohammadian	University of Canberra, Australia
Erkki Oja	Helsinki University of Technology, Finland
Soo-Chang Pei	National Taiwan University, Taiwan, ROC
Jose Principe	University of Florida, USA
Ponnuthurai Suganthan	Nanyang Technological University, Singapore
Ah Chung Tsoi	University of Wollongong, Australia
Marc Van Hulle	K.U. Leuven, Belgium
A.N. Venetsanopoulos	University of Toronto, Canada
Yue Wang,	The Catholic University of America, USA
Wilson Wen	Telstra Research Laboratories, Australia

CONTENTS

INTELLIGENT MULTIMEDIA PROCESSING

Audio-Visual Speech Recognition using Minimum Classification Error Training	3
C. Miyajima, K. Tokuda, T. Kitamura	
Realtime Face Analysis and Synthesis using Neural Network	13
S. Morshima	
Estimation and Generalization of Multimodal Speech Production	23
E. Vatikiotis-Bateson, H.C. Yehia	
HMM-Based Transmodal Mapping from Audio Speech to Talking Faces	33
S. Nakamura	

THEORY AND ALGORITHMS

A Tunable Approximately Piecewise Linear Model Derived from the Modified Probabilistic Neural Network	45
A. Zaknich, Y. Attikiouzel	
Two Constructive Algorithms for Improved Time Series Processing with Recurrent Neural Networks	55
R. Boné, M. Crucianu, J.P. Asselin de Beauville	
Separable Nonlinear Least-Squares Methods for On-Line Estimation of Neural nets Hammerstein Models	65
L.S.H. Ngia	
A Fast On-Line Algorithm for Computing Reduced-Rank Wiener Filters	75
M. Nikpour, H. Ali, J.H. Manton, Y. Hua	
Analysis of Nonstationary Time Series by Mixtures of Self-Organizing Predictors	85
J. Kohlmorgen, S. Lemm, G. Rätsch, K.R. Müller	
Robust Adaptive Identification of Nonlinear System using Neural Network	95
Q. Song, L. Yin, Y.C. Soh	

Sequential Order Selection for Real Time Signal Processing H. Ni, T. Adali	105
A Hierarchical Bayesian Nonlinear Time Series Prediction Weighted by Marginal Likelihoods M. Saito, M. Asano, T. Matsumoto	115
Bayesian Methods for Autoregressive Models W.D. Penny, S.J. Roberts	125
Variational Bayes for Non-Gaussian Autoregressive Models W.D. Penny, S.J. Roberts	135
Optimal Model Inference for Bayesian Mixture of Experts N. Ueda, Z. Ghahramani	145
Bayesian MCMC Nonlinear Time Series Prediction: Predictive Mean and Error Bar Y. Nakada, T. Matsumoto	155
Initialising Neural Networks with a-priori Problem Knowledge R. Chaplin, S. Gunitileke, R.M. Hodgson	165
Neural Network Frames on the Sphere H.N. Mhaskar, F.J. Narcowich, J.D. Ward	175
The Use of Möbius Transformations in Neural Networks and Signal Processing D. Mandic	185
On-Line Learning by Active Sampling using Orthogonal Decision Support Vectors J.M. Park	195
Regularisation by Convolution in Probability Density Estimation is Equivalent to Jittering C.G. Molina, J. Zerubia	204
Quantum Neurocomputation and Signal Processing H. Eghbalnia, A. Assadi	211
On Comparison of Adaptive Regularization Methods S. Sigurdsson, J. Larsen, L.K. Hansen	221
Fully Complex Backpropagation for Constant Envelope Signal Processing T. Kim, T. Adali	231

Monitoring the Formation of Kernel-based Topographic Maps M.M. Van Hulle	241
Hierarchical Density-Based Clustering in High-Dimensional Spaces using Topographic Maps T. Gautama, M.M. Van Hulle	251
Clustering Method using Self-Organizing Maps M. Endo, M. Ueno, T. Tanabe, M. Yamamoto	261
K-tree: A Height Balanced Tree Structured Vector Quantiser S. Geva	271
Competitive Learning using Gradient and Reinitialization Methods for Adaptive Vector Quantization S. Kurogi, T. Nishida	281
Robust Principal Component Analysis M. Partridge, M. Jabri	289
A Fast On-Line Algorithm for PCA and its Convergence Characteristics Y.N. Rao, J.C. Principe	299
Adaptive Transform Coding as Constrained Vector Quantization C. Archer, T.K. Leen	308
Bullet-Hole Image Classification with Support Vector Machines W.F. Xie, D.J. Hou, Q. Song	318
Quantum Neural Networks Versus Conventional Feedforward Neural Networks: An Experimental Study R. Kretzschmar, R. Bueeler, N.B. Karayiannis, F. Eggimann	328
Using Minimum Classification Error Training in Dimensionality Reduction X. Wang, K.K. Paliwal	338
An Analysis of a Multi-Layered Competitive Net for Invariant Pattern Recognition T. Nishida, S. Kurogi	346
DirectSVM: A Fast and Simple Support Vector Machine Perceptron D. Roobaert	356
Attributed Relational Graph Matching by Neural-Gas Networks P.N. Suganthan	366

Visualizing Class Structure in Data using Mutual Information K. Torkkola	376
On-Line Probabilistic Classification with Particle Filters P. Højten-Sørensen, N. de Freitas, T. Fog	386
An Adaptive Spline Non-Linear Function for Blind Signal Processing M. Solazzi, F. Piazza, A. Uncini	396
Blind Separation of Nonstationary and Temporally Correlated Sources from Noisy Mixtures S. Choi, A. Cichocki	405
The BID for blind equalization of FIR MIMO channels Y. Hua, Y. Xiang	415
Kalman Filter and State Space Approach to Blind Deconvolution L.Q. Zhang, A. Cichocki, S. Amari	425
An Ensemble Learning Approach to Independent Component Analysis R. Choudrey, W.D. Penny, S.J. Roberts	435
Maximum Likelihood Estimation of ICA Model for Wide Class of Source Distributions J. Karvanen, J. Eriksson, V. Koivunen	445
A Blind Extraction of Temporally Correlated but Statistically Dependent Acoustic Signals A. Cichocki, T. Rutkowski, A. Kardec Barros, S.H. Oh	455
Canonical Correlation Analysis: A Blind Source Separation using Non-Circularity J. Galy, C. Adnet	465

APPLICATIONS

Neural-Network-Based Real-Time Human Body Posture Estimation K. Takahashi, T. Uemura, J. Ohya	477
Edge Detection from Noisy Images using a Neural Edge Detector K. Suzuki, I. Haribo,, N. Sugie	487
Low-Complexity Fusion of Intensity, Motion, Texture and Edge for Image Sequence Segmentation: A Neural Network Approach J. Kim, T. Chen	497

Face Detection and Eye Localization by Neural Network based Color Segmentation	507
H.C. Fu, P.S. Lai, R.S. Lou, H.T. Pao	
Fingerprint Image Compression	517
G.A. Khuwaja, A.S. Tolba	
Application of Flexible Neuronal Matching for Face Recognition	527
X. Zeng, S.W. Foo	
Training under Achievement Quotient Criterion	537
K. Suzuki, I. Haribo, N. Sugie	
Multiresolution Digital Watermarking for Still Image	547
X.M. Niu, S.H. Sun	
Image Processing using Cellular Neural Networks based on Multi-Valued and Universal Binary Neurons	557
I. Aizenberg, N. Aizenberg, T. Bregin, C. Butakov, E. Farberov	
A Recursive Approach to Joint Image Restoration and Compensated Blur Identification	567
K.H. Yap, L. Guan	
Measuring the Complexity of the Ocean Floor	576
T. Sams, B. Stage, F.T. Agerkvist, T. Christensen	
Face Recognition using a New Distance Metric	584
M. Partridge, M. Jabri	
Echocardiographic Image Sequence Segmentation using Self-Organizing Maps	594
M.L. Siqueira, C.V. Gasperin, J. Scharcanski, P. Zielinsky, P.O.A. Navaux	
An Edge Detection Scheme using Radial Basis Function Networks	604
C.R. De Silva, L.C. De Silva, S. Ranganath	
Learning-based Approach for License Plate Recognition	614
K.K. Kim, K.I. Kim, J.B. Kim, H.J. Kim	
Color Image Segmentation using Color Space Analysis and Fuzzy Clustering	624
D.X. Zhong, H. Yan	
Support Vector Machine-based Text Detection in Digital Video	634
C.S. Shin, K.I. Kim, M.H. Park, H.J. Kim	

Feature Grid Neural Networks for Curve Partitioning M. Tan, Q. Gao	642
Two Dimensional Phase Retrieval using Neural Networks A. Burian, P. Kuosmanen, J. Saarinen, C. Rusu	652
Fuzzy Logic Based Microcalcification Detection N. Pandey, Z. Salcic, J. Sivaswamy	662
Stable Fuzzy State Space Controllers for an AGV K.R.S. Kodagoda, W.S. Wijesoma, E.K. Teoh	672
The Use of Problem Knowledge to Improve the Robustness of a Fuzzy Neural Network S. Gunetileke, R.I. Chaplin, R.M. Hodgson	682
Agglomerative Learning for General Fuzzy Min-Max Neural Network B. Gabrys	692
An Empirical Investigation of the User-Parameters and Performance of Continuous PBIL Algorithms M. Gallagher	702
A Novel Method for CFAR Data Fusion W. Liu, Y. Lu, J.S. Fu	711
Neural Computation Approach for the Maximum-Likelihood Sequence Estimation of Communications Signal Y. Tan	721
Adaptive Multidimensional Spline Neural Network for Digital Equalization M. Solazzi, A. Uncini	729
A Neural Predictor for Blind Equalization of Digital Communication Systems: Is It Plausible? C.C. Cavalcante, J.R. Montalvão Filho, B. Dorizzi, J.C. M. Mota	736
Multiuser Demodulators using Adaptive Polynomial Perceptrons in CDMA Systems S.Y. Choi, T. Choi, J. Rho, K. Ko, D. Hong	746
Networks for Networks: Internet Analysis using Graphical Statistical Models M. Coates, R. Nowak	755

Implementation of Channel Repacking Algorithms on Hopfield Neural Networks Cellular Systems	765
D. Tissainayagam, M. Palaniswami	
Support Vector Machines for Speaker Verification and Identification	775
V. Wan, W.M. Campbell	
Lateral Inhibition Mechanism in Computational Auditory Model and Its Application in Robust Speech Recognition	785
L. Xugang, L. Gang, W. Lipo	
Using Polynomial Networks for Speech Recognition	795
W.M. Campbell, C.C. Broun	
Estimation of Coupled Hidden Markov Models with Application to Biosignal Interaction Modelling	804
I. Rezek, S.J. Roberts	
Alertness Monitor using Neural Networks for EEG Analysis	814
B.J. Wilson, T.D. Bracewell	
EEG Classification by Autocorrelation-Pulse in Left and Right Motor Imaginary Data	821
I.V. Mayer, H. Takahashi and K. Sakamoto	
Use of Neural Networks for Feature Based Recognition of Liver Region on Ct Images	831
S.A. Husain, E. Shigeru	
Optimal Feature Selection using Information Maximisation: Case of Biomedical Data	841
A. Al-Ani, M. Deriche	
Using Statistical and Neural Network Methods to Explore the Relationship Between Systematic Risk and Firm's Long Term Investment Activities	851
K. Wang, H.T. Pao, H.C. Fu	
Off-Line Signature Verification using HMMs and Cross-Validation	851
A. El-Yacoubi, E.J.R. Justino, R. Sabourin, F. Bortolozzi	
Recognition of Defects in High Voltage Transmission Lines using The Acoustic Signal of Corona Effect	86
R. Tadeusiewicz, T. Wszolek, A. Izvorski, W. Wszolek	
A Hopfield Neural Network Approach for the Reconstruction of Wide-Bandwidth Sonar Data	87
S.W. Perry, R.J. Wyber	

Automatic Evaluation of Adaptive Algorithms over the Internet S. Lucas, P. Beattie, J. Coy	886
Effect of Parallel Ensembles to Self-Generating Neural Networks for Chaotic Time Series Prediction H. Inoue, H. Narihisa	896
A Modular Signal Processing Model for Permeability Prediction in Petroleum Reservoir K.W. Wong, T. Gedeon	906
Nonlinear Active Noise Control using Lyapunov Theory and RBF Network S.K. Phooi, M. Zhihong, H.R. Wu	916
Fusion of Multiple Handwritten Word Recognition Techniques B. Verma, P. Gader	926
Arithmetic-Unit and Processor Design for Neural Networks A.R. Omondi	935
Author Index	945

THEORY AND ALGORITHMS

AUDIO-VISUAL SPEECH RECOGNITION USING MINIMUM CLASSIFICATION ERROR TRAINING

Chiyoumi Miyajima, Keiichi Tokuda, and Tadashi Kitamura
Department of Computer Science, Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan
Phone: +81 52 735 5404 Fax: +81 52 735 5477
E-mail: {chiyoumi, tokuda, kitamura}@ics.nitech.ac.jp

Abstract. This paper presents a framework for designing a hidden Markov model (HMM)-based audio-visual automatic speech recognition system based on minimum classification error (MCE) training. Audio/visual HMMs are optimized with MCE training based on the generalized probabilistic descent (GPD) method, and their likelihoods are combined using model-dependent stream weights which are also estimated with the GPD method. Experimental results of speaker independent isolated word recognition show that the GPD optimization of the audio/visual HMMs and the use of GPD-based model-dependent stream weights provide significant improvement in system performance leading to 47% – 81% error reduction over a conventional system which consists of HMMs trained based on the maximum likelihood criterion and globally-tied stream weights estimated with the GPD method.

INTRODUCTION

Human speech perception is enhanced by seeing speaker's lip movements as well as hearing speaker's voice. Consequently, an audio-visual (bimodal) aspect of automatic speech recognition (ASR) has been widely studied and many works have shown that the robustness and accuracy of ASR can be improved by the use of labial information in addition to the acoustic speech signal, especially in noisy conditions [1, 2, 3, 4].

In audio-visual as well as auditory speech recognition, hidden Markov models (HMMs) [5] are most currently used for modeling acoustic and labial features. Such systems consists of HMM-based audio/visual classifiers and integration modules in which the outputs of the HMMs are combined using stream weights. Most widely used HMM training algorithms are based on the maximum likelihood (ML) estimation principle, which aims at approximating the assumed source distribution of each class. The ML estimation, however, is inappropriate for the training of stream weights. Instead, minimum classification error (MCE) training based on the generalized probabilistic descent (GPD) method [6] has been successfully applied to the stream weight estimation [4]. Furthermore, the utility of GPD is not

limited to stream weight estimation. The GPD training aims at maximizing the separation between classes and has been successfully applied to classifier designs in the wide range of pattern recognition domains [7], including HMM parameter training for speech recognition [8].

In this study, audio/visual HMM parameters as well as stream weights are estimated using the GPD method, and their likelihoods are combined using GPD-trained model-dependent stream weights [9]. To evaluate this framework, speaker-independent isolated word recognition experiments are conducted using the M2VTS database [10]. The GPD optimization of HMM parameters provides higher recognition abilities in audio- and visual-only ASR, yielding significant improvement in audio-visual ASR performance. A further significant gain is obtained by using different values of stream weights for each word class.

HMM-BASED AUDIO-VISUAL ASR SYSTEM

There are essentially two approaches to integrating audio and visual information: early integration, in which acoustic and labial feature vectors are combined by synchronizing their frame rates using a frame interpolation technique before being processed in a classifier, and late integration, in which separate classifiers are used for audio and visual data streams and a recognition score is computed by combining their outputs [2]. Our system is based on the latter approach.

As illustrated in Figure 1, each word w ($w = 1, 2, \dots, W$) is modeled by two separate HMMs $\{\lambda_{Aw}, \lambda_{Vw}\}$, and their likelihoods are combined being weighted by their respective stream weights $\{\gamma_{Aw}, \gamma_{Vw}\}$. Let each spoken word be represented by a set of audio and visual feature observation sequences:

$$X = (O_A, O_V), \quad (1)$$

$$O_A = (o_{A1}, o_{A2}, \dots, o_{AT_A}), \quad (2)$$

$$O_V = (o_{V1}, o_{V2}, \dots, o_{VT_V}), \quad (3)$$

with T_A and T_V being the number of frames in the audio and visual sequences, respectively. We define the audio-visual recognition score of the w -th word model as

$$S_w(X) = P(O_A, q_A | \lambda_{Aw})^{\gamma_{Aw}} P(O_V, q_V | \lambda_{Vw})^{\gamma_{Vw}}, \quad (4)$$

where q_A and q_V denote the optimal state sequences (Viterbi paths) of O_A and O_V in λ_{Aw} and λ_{Vw} , respectively:

$$q_A = (q_{A0}, q_{A1}, \dots, q_{AT_A}), \quad (5)$$

$$q_V = (q_{V0}, q_{V1}, \dots, q_{VT_V}), \quad (6)$$

and $P(O, q | \lambda)$ is the likelihood of O along a Viterbi path q in an HMM λ :

$$P(O, q | \lambda) = \prod_{t=1}^T a_{q_{t-1}q_t} b_{q_t}(o_t), \quad (7)$$