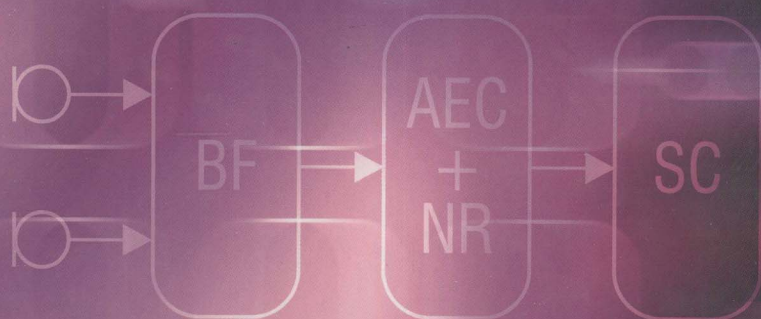


PETER VARY RAINER MARTIN

Digital Speech Transmission

Enhancement, Coding and Error Concealment



 WILEY

TN 912.3
V327

Digital Speech Transmission

Enhancement, Coding and Error Concealment

Peter Vary

*Institute of Communication Systems and Data Processing
RWTH Aachen University, Germany*

Rainer Martin

*Institute of Communication Acoustics
Ruhr-Universität Bochum, Germany*



JOHN WILEY & SONS, LTD



E200602360

Copyright © 2006

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk

Visit our Home Page on www.wiley.com

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, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Vary, Peter.

Digital speech transmission : enhancement, coding, and error concealment/

Peter Vary and Rainer Martin.

p. cm.

Includes bibliographical references and index.

ISBN 0-471-56018-9 (cloth : alk. paper)

1. Speech processing systems. 2. Signal processing—Digital techniques.

3. Error-correcting codes (Information theory) I. Martin, Rainer. II.

Title.

TK7882.S65.V37 2005

621.39'9—dc22

2005034991

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-471-56018-9

ISBN-10 0-471-56018-X

Typeset by the authors using LATEX software

Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire

This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Digital Speech Transmission

Preface

The digital processing, storage, and transmission of speech signals have gained great practical importance. The main application areas are digital mobile radio, acoustic human-machine communication, and digital hearing aids. In fact, these applications are the driving force behind many scientific and technological developments in this field. A specific characteristic of these application areas is that theory and practice are closely linked; there is a seamless transition from theory to system simulations using general-purpose computers and to system realizations with programmable processors.

This book has been written for electrical engineers, information technology engineers, as well as for engineering students. It summarizes recent developments in the broad field of digital speech transmission and is based to a large extent on joint research of the authors. This book is used in courses at RWTH Aachen University and Ruhr-Universität Bochum. Portions of this volume are translated and revised from the German edition of *Digitale Sprachsignalverarbeitung*, by P. Vary, U. Heute, and W. Hess, with kind permission of Teubner Verlag. The reader will find supplementary information, publications, programs, and audio samples on the following web sites:

<http://www.ind.rwth-aachen.de>

<http://www.rub.de/ika>

The scope of the individual subjects treated in the book often exceeds that of the lectures; recent research results, standards, problems of realization, and applications have been included, as well as many suggestions for further reading. To gain maximum benefit from the text, the reader should be familiar with the fundamentals of digital signal processing and statistical signal and system description. A summary of spectral analysis, digital filter banks, as well as stochastic signals and estimation is provided.

The authors are grateful to all members of staff and students who contributed to the book through research results, discussions, and editorial work. We thank Dr.-Ing. Tim Fingscheidt, Dr.-Ing. Peter Jax, and Dr.-Ing. Marc Adrat for contributions to Chapters 9 and 10. Horst Krott prepared most of the diagrams. Dipl.-Ing. David Bauer and Dipl.-Ing. Laurent Schmalen helped us with L^AT_EX editing, and Diplom-Anglistin Heike Hagena and Christina Storms, MA, supported us in translating the texts. These contributions are gratefully acknowledged.

We would like to express our thanks for the friendly assistance of the editors of John Wiley & Sons.

Finally, we would especially like to thank Dr.-Ing. Christiane Antweiler, whose tireless efforts resulted in completion, consistent layout, and content coordination of the entire manuscript.

Aachen and Bochum, January, 2006

Peter Vary and Rainer Martin

Reference List

Reproduced with permission of Elsevier:

Fig. 4.5, Fig. 11.17, Fig. 13.27

Reproduced with permission of IEEE:

Fig. 5.9, Fig. 5.10, Fig. 9.13, Fig. 9.16, Fig. 9.17, Fig. 11.12, Fig. 11.15, Fig. 11.16, Fig. 11.18, Fig. 12.16, Fig. B.4

Reproduced with permission of Schiele & Schön:

Fig. 8.29

Reproduced with permission of Springer Verlag:

Fig. 2.17, Fig. 2.18, Fig. 2.19, Fig. 2.20, Fig. 2.21, Fig. 2.22

Reproduced with permission of B. G. Teubner:

Fig. 5.2, Fig. 5.6, Fig 5.7, Fig. 5.8, Fig. 6.1, Fig. 6.6, Fig. 6.8, Fig. 6.12, Fig. 7.3, Fig. 7.4, Fig. 7.5, Fig 7.6, Fig. 7.8, Fig. 7.9, Fig. 7.11, Fig. 7.12, Fig. 7.16, Fig. 8.2, Fig. 8.4, Fig. 8.5, Fig 8.7, Fig. 8.8, Fig. 8.9, Fig. 8.10, Fig. 8.11, Fig. 8.17, Fig. 8.18, Fig. 8.19, Fig. 8.20, Fig 8.21, Fig. 8.22, Fig. 8.23, Fig. 8.25, Fig. 8.26, Fig. 8.30, Fig. 8.31, Fig. 8.36, Fig. 11.21, Fig 11.27, Fig. 13.2, Fig. 13.3, Fig. 13.5, Fig. 13.6, Fig. 13.7, Fig. 13.8, Fig. 13.9, Fig. 13.11, Fig 13.12, Fig. 13.21, Fig. 13.25, Fig. 13.26, Fig. 13.28, Fig. A.3-b, Fig. A.3-c, Fig. A.4, Fig. A.6-b, Fig. A.7, Fig. A.13

Contents

Preface	xv
1 Introduction	1
2 Models of Speech Production and Hearing	5
2.1 Organs of Speech Production	6
2.2 Characteristics of Speech Signals	8
2.3 Model of Speech Production	10
2.3.1 Acoustic Tube Model of the Vocal Tract	11
2.3.2 Digital All-Pole Model of the Vocal Tract	19
2.4 Anatomy of Hearing	25
2.5 Psychoacoustic Properties of the Auditory Organ	28
2.5.1 Hearing and Loudness	28
2.5.2 Spectral Resolution	30
2.5.3 Masking	32
Bibliography	33
3 Spectral Transformations	35
3.1 Fourier Transform of Continuous Signals	35
3.2 Fourier Transform of Discrete Signals	37
3.3 Linear Shift Invariant Systems	39
3.3.1 Frequency Response of LSI Systems	41
3.4 The z -transform	41
3.4.1 Relation to FT	43
3.4.2 Properties of the ROC	44
3.4.3 Inverse z -transform	44
3.4.4 z -transform Analysis of LSI Systems	46
3.5 The Discrete Fourier Transform	47
3.5.1 Linear and Cyclic Convolution	50

3.5.2	The DFT of Windowed Sequences	52
3.5.3	Spectral Resolution and Zero Padding	55
3.5.4	Fast Computation of the DFT: The FFT	56
3.5.5	Radix-2 Decimation-in-Time FFT	57
3.6	Fast Convolution	61
3.6.1	Fast Convolution of Long Sequences	61
3.6.2	Fast Convolution by Overlap-Add	61
3.6.3	Fast Convolution by Overlap-Save	62
3.7	Cepstral Analysis	65
3.7.1	Complex Cepstrum	65
3.7.2	Real Cepstrum	66
3.7.3	Applications of the Cepstrum	67
	Bibliography	70
4	Filter Banks for Spectral Analysis and Synthesis	73
4.1	Spectral Analysis Using Narrowband Filters	73
4.1.1	Short-Term Spectral Analyzer	78
4.1.2	Prototype Filter Design for the Analysis Filter Bank	82
4.1.3	Short-Term Spectral Synthesizer	84
4.1.4	Short-Term Spectral Analysis and Synthesis	86
4.1.5	Prototype Filter Design for the Analysis-Synthesis Filter Bank	88
4.1.6	Filter Bank Interpretation of the DFT	90
4.2	Polyphase Network Filter Banks	93
4.2.1	PPN Analysis Filter Bank	93
4.2.2	PPN Synthesis Filter Bank	101
4.3	Quadrature Mirror Filter Banks	105
4.3.1	Analysis-Synthesis Filter Bank	105
4.3.2	Compensation of Aliasing and Signal Reconstruction	107
4.3.3	Efficient Implementation	111
	Bibliography	115
5	Stochastic Signals and Estimation	119
5.1	Basic Concepts	119
5.1.1	Random Events and Probability	119
5.1.2	Conditional Probabilities	121
5.1.3	Random Variables	121
5.1.4	Probability Distributions and Probability Density Functions	122

5.1.5 Conditional PDFs 123

5.2 Expectations and Moments 124

5.2.1 Conditional Expectations and Moments 125

5.2.2 Examples 125

5.2.3 Transformation of a Random Variable 128

5.2.4 Relative Frequencies and Histograms 129

5.3 Bivariate Statistics 130

5.3.1 Marginal Densities 130

5.3.2 Expectations and Moments 130

5.3.3 Uncorrelatedness and Statistical Independence 131

5.3.4 Examples of Bivariate PDFs 132

5.3.5 Functions of Two Random Variables 133

5.4 Probability and Information 135

5.4.1 Entropy 135

5.4.2 Kullback–Leibler Divergence 135

5.4.3 Mutual Information 136

5.5 Multivariate Statistics 136

5.5.1 Multivariate Gaussian Distribution 137

5.5.2 χ^2 -distribution 137

5.6 Stochastic Processes 138

5.6.1 Stationary Processes 138

5.6.2 Auto-correlation and Auto-covariance Functions 139

5.6.3 Cross-correlation and Cross-covariance Functions 140

5.6.4 Multivariate Stochastic Processes 140

5.7 Estimation of Statistical Quantities by Time Averages 142

5.7.1 Ergodic Processes 142

5.7.2 Short-Time Stationary Processes 143

5.8 Power Spectral Densities 144

5.8.1 White Noise 145

5.9 Estimation of the Power Spectral Density 145

5.9.1 The Periodogram 145

5.9.2 Smoothed Periodograms 147

5.10 Statistical Properties of Speech Signals 147

5.11 Statistical Properties of DFT Coefficients 148

5.11.1 Asymptotic Statistical Properties 149

5.11.2 Signal-plus-Noise Model 150

5.11.3 Statistical Properties of DFT Coefficients for Finite Frame Lengths 152

5.12 Optimal Estimation 154

5.12.1 MMSE Estimation 155

5.12.2 Optimal Linear Estimator 156

5.12.3 The Gaussian Case 157

5.12.4 Joint Detection and Estimation 158

Bibliography 160

6	Linear Prediction	163
6.1	Vocal Tract Models and Short-Term Prediction	164
6.2	Optimal Prediction Coefficients for Stationary Signals	171
6.2.1	Optimum Prediction	171
6.2.2	Spectral Flatness Measure	174
6.3	Predictor Adaptation	177
6.3.1	Block-Oriented Adaptation	177
6.3.2	Sequential Adaptation	188
6.4	Long-Term Prediction	192
	Bibliography	198
7	Quantization	201
7.1	Analog Samples and Digital Representation	201
7.2	Uniform Quantization	203
7.3	Non-uniform Quantization	211
7.4	Optimal Quantization	221
7.5	Adaptive Quantization	222
7.6	Vector Quantization	228
7.6.1	Principle	228
7.6.2	The Complexity Problem	230
7.6.3	Lattice Quantization	231
7.6.4	Design of Optimal Vector Code Books	232
7.6.5	Gain-Shape Vector Quantization	236
	Bibliography	237
8	Speech Coding	239
8.1	Classification of Speech Coding Algorithms	240
8.2	Model-Based Predictive Coding	243
8.3	Differential Waveform Coding	245
8.3.1	First-Order DPCM	245
8.3.2	Open-Loop and Closed-Loop Prediction	249
8.3.3	Quantization of the Residual Signal	250
8.3.4	Adaptive Differential Pulse Code Modulation	260
8.4	Parametric Coding	262
8.4.1	Vocoder Structures	262
8.4.2	LPC Vocoder	265
8.4.3	Quantization of the Predictor Coefficients	266
8.5	Hybrid Coding	273

8.5.1	Basic Codec Concepts	273
8.5.2	Residual Signal Coding: RELP	282
8.5.3	Analysis by Synthesis: CELP	290
8.5.4	Analysis by Synthesis: MPE, RPE	301
8.6	Adaptive Postfiltering	305
	Bibliography	309
9	Error Concealment and Soft Decision Source Decoding	315
9.1	Hard Decision Source Decoding	316
9.2	Conventional Error Concealment	317
9.3	Softbits and L -values	321
9.3.1	Binary Symmetric Channel (BSC)	321
9.3.2	Fading-AWGN Channel	329
9.3.3	Channel with Inner SISO Decoding	335
9.4	Soft Decision (SD) Source Decoding	336
9.4.1	Parameter Estimation	338
9.4.2	The A <i>Posteriori</i> Probabilities	340
9.5	Application to Model Parameters	345
9.5.1	Soft Decision Decoding without Channel Coding	346
9.5.2	Soft Decision Decoding with Channel Coding	348
9.6	Further Improvements	353
	Bibliography	355
10	Bandwidth Extension (BWE) of Speech Signals	361
10.1	Narrowband versus Wideband Telephony	362
10.2	Speech Coding with Integrated BWE	366
10.3	BWE without Auxiliary Transmission	369
10.3.1	Basic Approaches and Classification	369
10.3.2	Spectral Envelope Estimation	372
10.3.3	Extension of the Excitation Signal	375
10.3.4	Example BWE Algorithm	377
	Bibliography	382

11 Single and Dual Channel Noise Reduction	389
11.1 Introduction	390
11.2 Linear MMSE Estimators	392
11.2.1 Non-causal IIR Wiener filter	392
11.2.2 The FIR Wiener Filter	395
11.3 Speech Enhancement in the DFT Domain	396
11.3.1 The Wiener Filter Revisited	398
11.3.2 Spectral Subtraction	400
11.3.3 Estimation of the <i>A Priori</i> SNR	402
11.3.4 Musical Noise and Countermeasures	403
11.3.5 Aspects of Spectral Analysis/Synthesis	408
11.4 Optimal Non-linear Estimators	411
11.4.1 Maximum Likelihood Estimation	412
11.4.2 Maximum <i>A Posteriori</i> Estimation	414
11.4.3 MMSE Estimation	414
11.4.4 MMSE Estimation of Functions of the Spectral Magnitude	416
11.5 Joint Optimum Detection and Estimation of Speech	419
11.6 Computation of Likelihood Ratios	422
11.7 Estimation of the <i>A Priori</i> Probability of Speech Presence	423
11.7.1 A Hard-Decision Estimator Based on Conditional Probabilities	423
11.7.2 Soft-Decision Estimation	424
11.7.3 Estimation Based on the <i>A Posteriori</i> SNR	424
11.8 VAD and Noise Estimation Techniques	425
11.8.1 Voice Activity Detection	426
11.8.2 Noise Estimation Using a Soft-Decision Detector	432
11.8.3 Noise Power Estimation Based on Minimum Statistics	434
11.9 Dual Channel Systems	443
11.9.1 Noise Cancellation	449
11.9.2 Noise Reduction	452
11.9.3 Implementations of Dual Channel Noise Reduction Systems	453
11.9.4 Combined Single and Dual Channel Noise Reduction	454
Bibliography	456

12 Multi-channel Noise Reduction	467
12.1 Introduction	467
12.2 Sound Waves	468
12.3 Spatial Sampling of Sound Fields	470
12.3.1 The Farfield Model	472
12.3.2 The Uniform Linear Array	474
12.3.3 Phase Ambiguity and Coherence	475
12.3.4 Spatial Correlation Properties of Acoustic Signals	476
12.4 Beamforming	477
12.4.1 Delay-and-Sum Beamforming	477
12.4.2 Filter-and-Sum Beamforming	478
12.5 Performance Measures and Spatial Aliasing	481
12.5.1 Array Gain and Array Sensitivity	481
12.5.2 Directivity Pattern	482
12.5.3 Directivity and Directivity Index	484
12.5.4 Example: Differential Microphones	485
12.6 Design of Fixed Beamformers	488
12.6.1 Minimum Variance Distortionless Response Beamformer	488
12.6.2 MVDR Beamformer with Limited Susceptibility	491
12.7 Multi-channel Wiener Filter and Postfilter	493
12.8 Adaptive Beamformers	495
12.8.1 The Frost Beamformer	495
12.8.2 Generalized Side-Lobe Canceller	498
12.8.3 Generalized Side-lobe Canceller with Adaptive Blocking Matrix .	500
12.9 Optimal Non-linear Multi-channel Noise Reduction	501
Bibliography	501
 13 Acoustic Echo Control	 505
13.1 The Echo Control Problem	505
13.2 Evaluation Criteria	511
13.3 The Wiener Solution	513
13.4 The LMS and NLMS Algorithms	514
13.4.1 Derivation and Basic Properties	514
13.5 Convergence Analysis and Control of the LMS Algorithm	516
13.5.1 Convergence in the Absence of Interference	517
13.5.2 Convergence in the Presence of Interference	520
13.5.3 Filter Order of the Echo Canceller	523

13.5.4 Stepsize Parameter 524

13.6 Geometric Projection Interpretation of the NLMS Algorithm 527

13.7 The Affine Projection Algorithm 529

13.8 Least-Squares and Recursive Least-Squares Algorithms 531

13.8.1 The Weighted Least-Squares Algorithm 532

13.8.2 The RLS Algorithm 533

13.9 Block Processing and Frequency Domain Adaptive Filters 536

13.9.1 Block LMS Algorithm 537

13.9.2 The Exact Block NLMS Algorithm 537

13.9.3 Frequency Domain Adaptive Filter (FDAF) 539

13.9.4 Subband Acoustic Echo Cancellation 549

13.10 Additional Measures for Echo Control 550

13.10.1 Echo Canceller with Center Clipper 550

13.10.2 Echo Canceller with Voice-Controlled Switching 551

13.10.3 Echo Canceller with Adaptive Postfilter in the Time Domain . . . 553

13.10.4 Echo Canceller with Adaptive Postfilter in the Frequency Domain 554

13.10.5 Initialization with Perfect Sequences 555

13.11 Stereophonic Acoustic Echo Control 557

13.11.1 The Non-uniqueness Problem 559

13.11.2 Solutions to the Non-uniqueness Problem 559

Bibliography 561

Appendix A Codec Standards 569

A.1 Evaluation Criteria 570

A.2 ITU-T/G.726: Adaptive Differential Pulse Code Modulation (ADPCM) . 572

A.3 ITU-T/G.728: Low-Delay CELP Speech Coder 573

A.4 ITU-T/G.729: Conjugate-Structure Algebraic CELP Codec 576

A.5 ITU-T/G.722: 7 kHz Audio Coding within 64 kbit/s 579

A.6 ETSI-GSM 06.10: Full Rate Speech Transcoding 580

A.7 ETSI-GSM 06.20: Half Rate Speech Transcoding 582

A.8 ETSI-GSM 06.60: Enhanced Full Rate Speech Transcoding 584

A.9 ETSI-GSM 06.90: Adaptive Multi-Rate (AMR) Speech Transcoding . . . 586

A.10 ETSI/3GPP AMR Wideband Speech Transcoding 590

A.11 ETSI/3GPP Extended AMR Wideband Codec, AMR-WB⁺ 592

A.12 TIA IS-96: Speech Service Option Standard for
Wideband Spread-Spectrum Systems 594

A.13 INMARSAT: Improved Multi-Band Excitation Codec (IMBE) 595

Appendix B	Speech Quality Assessment	597
B.1	Auditive Speech Quality Measures	597
B.2	Instrumental Speech Quality Measures	602
	Bibliography	604
Index		607

1

Introduction

Language is the most essential means of human communication. It is used in two modes: as spoken language (*speech communication*) and as written language (*textual communication*). In our modern information society both modes are greatly enhanced by technical systems and devices. E-mail, short-messaging, and the world-wide web have revolutionized textual communication while

- digital mobile radio systems,
- acoustic human-machine communication, and
- digital hearing aids

have significantly expanded the possibilities and convenience of speech communication.

Digital processing of speech signals for the purpose of transmission (or storage) is a branch of information technology and an engineering science which draws on various other disciplines such as physiology, phonetics, linguistics, acoustics, and psychoacoustics. It is this multidisciplinary aspect which makes digital speech processing a challenging as well as rewarding task.

The goal of this book is a comprehensive discussion of fundamental issues, standards, and recent trends in speech communication technology. Speech communi-

Digital Speech Transmission: Enhancement, Coding and Error Concealment

Peter Vary and Rainer Martin

© 2006 John Wiley & Sons, Ltd