THE KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE

APPLICATIONS OF DIGITAL SIGNAL PROCESSING TO AUDIO AND ACOUSTICS

edited by

Mark Kahrs
Rutgers University
Piscataway, New Jersey, USA

Karlheinz Brandenburg
Fraunhofer Institut Integrierte Schaltungen
Erlangen, Germany

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-3064-7042-X
Print ISBN 0-7923-8130-0

Created in the United States of America

Visit Kluwer Online at: http://www.kluweronline.com
and Kluwer's eBookstore at: http://www.ebooks.kluweronline.com

Contents

List of Figures

List of Tables

Acknowledgments

Mark Kahrs would like to acknowledge the support of J.L. Flanagan. He would also like to acknowledge the the assistance of Howard Trickey and S.J. Orfanidis. Jean Laroche has helped out with the production and served as a valuable forcing function. The patience of Diane Litrnan has been tested numerous times and she has offered valuable advice.

Karlheinz Brandenburg would like to thank Mark for his patience while he was always late in delivering his parts.

Both editors would like to acknowledge the patience of Bob Holland, our editor at Kluwer.

Contributing Authors

John G. Beerends was born in Millicent, Australia, in 1954. He received a degree in electrical engineering from the HTS (Polytechnic Institute) of The Hague, The Netherlands, in 1975. After working in industry for three years he studied physis and mathematics at the University of Leiden where he received the degree of M.Sc. in 1984. In 1983 he was awarded a prize of DF1 45000,- by Job Creation, for an innovative idea in the field of electro-acoustics. During the period 1984 to 1989 he worked at the Institute for Perception Research where he received a Ph.D. from the Technical University of Eindhoven in 1989. The main part of his Ph.D. work, which deals with pitch perception, was patented by the NV. Philips Gloeilampenfabriek. In 1989 he joined the audio group of the KPN research lab in Leidschendam where he works on audio quality assessment. Currently he is also involved in the development of an objective video quality measure.

Karlheinz Brandenburg received M.S. (Diplom) degrees in Electrical Engineering in 1980 and in Mathematics in 1982 from Erlangen University. In 1989 he earned his Ph.D. in Electrical Engineering, also from Erlangen University, for work on digital audio coding and perceptual measurement techniques. From 1989 to 1990 he was with AT&T Bell Laboratories in Murray Hill, NJ, USA. In 1990 he returned to Erlangen University to continue the research on audio coding and to teach a course on digital audio technology. Since 1993 he is the head of the Audio/Multimedia department at the Fraunhofer Institute for Integrated Circuits (FhG-IIS). Dr. Brandenburg is a member of the technical committee on Audio and Electroacoustics of the IEEE Signal Processing Society. In 1994 he received the ASE Fellowship Award for his work on perceptual audio coding and psychoacoustics.

Olivier Cappé was born in Villeurbanne, France, in 1968. He received the M.Sc. degree in electrical engineering from the Ecole Supérieure d'Electricité (ESE), Paris in 1990, and the Ph.D. degree in signal processing from the Ecole Nationale Supérieure des Télécommunications (ENST), Paris, in 1993. His Ph.D. tesis dealt with noise-reduction for degraded audio recordings. He is currently with the Centre National de la Recherche Scientifique (CNRS) at ENST, Signal department. His research interests are in statistical signal processing for telecomunications and speech/audio processing. Dr. Cappé received the IEE Signal Processing Society's Young Author Best Paper Award in 1995.

Bill Gardner was born in 1960 in Meriden, CT, and grew up in the Boston area. He received a bachelor's degree in computer science from MIT in 1982 and shortly thereafter joined Kurzweil Music Systems as a software engineer. For the next seven years, he helped develop software and signal processing algorithms for Kurzweil synthesizers. He left Kurzweil in 1990 to enter graduate school at the MIT Media Lab, where he recently completed his Ph.D. on the topic of 3-D audio using loudspeakers. He was awarded a Motorola Fellowship at the Media Lab, and was recipient of the 1997 Audio Engineering Society Publications Award. He is currently an independent consultant working in the Boston area. His research interests are spatial audio, reverberation, sound synthesis, realtime signal processing, and psychoacoustics.

Simon Godsill studied for the B.A. in Electrical and Information Sciences at the University of Cambridge from 1985-88. Following graduation he led the technical development team at the newly-formed CEDAR Audio Ltd., researching and developing DSP algorithms for restoration of degraded sound recordings. In 1990 he took up a post as Research Associate in the Signal Processing Group of the Engineering Department at Cambridge and in 1993 he completed his doctoral thesis: The Restoration of Degraded Audio Signals. In 1994 he was appointed as a Research Fellow at Corpus Christi College, Cambridge and in 1996 as University Lecturer in Signal Processing at the Engineering Department in Cambridge. Current research topics include: Bayesian and statistical methods in signal processing, modelling and enhancement of speech and audio signals, source signal separation, non-linear and non-Gaussian techniques, blind estimation of communications channels and image sequence analysis.

Mark Kahrs was born in Rome, Italy in 1952. He received an A.B. from Revelle College, University of California, San Diego in 1974. He worked intermittently for Tymshare, Inc. as a Systems Programmer from 1968 to 1974. During the summer of 1975 he was a Research Intern at Xerox PARC and then from 1975 to 1977 was a Research Programmer at the Center for Computer Research in Music and

Acoustics (CCRMA) at Stanford University. He was a chercheur at the Institut de Recherche et Coordination Acoustique Musique (IRCAM) in Paris during the summer of 1977. He received a PhD. in Computer Science from the University of Rochester in 1984. He worked and consulted for Bell Laboratories from 1984 to 1996. He has been an Assistant Professor at Rutgers University from 1988 to the present where he taught courses in Computer Architecture, Digital Signal Processing and Audio Engineering. In 1993 he was General Chair of the *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* ("Mohonk Workshop"). Since 1993 he has chaired the Technical Committee on Audio And Electroacoustics in the Signal Processing Society of the IEEE.

James M. Kates was born in Brookline, Massachusetts, in 1948. He received the degrees of BSEE and MSEE from the Massachusetts Institute of Technology in 1971 and the professional degree of Electrical Engineer from MIT in 1972. He is currently Senior Scientist at AudioLogic in Boulder, Colorado, where he is developing signal processing for a new digital hearing aid. Prior to joining AudioLogic, he was with the Center for Research in Speech and Hearing Sciences of the City University of New York. His research interests at CUNY included directional microphone arrays for hearing aids, feedback cancellation strategies, signal processing for hearing aid test and evaluation, procedures for measuring sound quality in hearing aids, speech enhancement algorithms for the hearing-impaired, new procedures for fitting hearing aids, and modeling normal and impaired cochlear function. He also held an appointment as an Adjunt Assistant Professor in the Doctoral Program in Speech and Hearing Sciences at CUNY, where he taught a course in modeling auditory physiology and perception. Previously, he has worked on applied research for hearing aids (Siemens Hearing Instruments), signal processing for radar, speech, and hearing applications (SIGNATRON, Inc.), and loudspeaker design and signal processing for audio applications (Acoustic Research and CBS Laboratories). He has over three dozen published papers and holds eight patents.

Jean Laroche was born in Bordeaux, France, in 1963 He earned a degree in Mathematics and Sciences from the Ecole Polytechnique in 1986, and a Ph.D. degree in Digital Signal Processing from the Ecole Nationale des Télécommunications in 1989. He was a post-doc student at the Center for Music Experiment at UCSD in 1990, and came back to the Ecole Nationale des Télécommunications in 1991 where he taught audio DSP, and acoustics. Since 1996 he has been a researcher in audio/music DSP at the Joint Emu/Creative Technology Center in Scotts Valley, CA.

Robert J. McAulay was born in Toronto, Ontario, Canada on October 23, 1939. He received the B.A.Sc. degree in Engineering Physics with honors from the University of Toronto, in 1962; the M.Sc. degree in Electrical Engineering from the University of Illinois, Urbana in 1963; and the Ph.D. degree in Electrical Engineering from the University of California, Berkeley, in 1967. He joined the Radar Signal Processing Group of the Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, MA, where he worked on problems in estimation theory and signal/filter design using optimal control techniques. From 1970 until 1975, he was a member of the Air Traffic Control Division at Lincoln Laboratory, and worked on the development of aircraft tracking algorithms, optimal MTI digital signal processing and on problems of aircraft direction finding for the Discrete Address Beacon System. On a leave of absence from Lincoln Laboratory during the winter and spring of 1974, he was a Visiting Associate Professor at McGill University, Montreal, P.Q., Canada. From 1975 until 1996, he was a member of the Speech Systems Technology Group at Lincoln Laboratory, where he was involved in the development of robust narrowband speech vocoders. In 1986 he served on the National Research Council panel that reviewed the problem of the removal of noise from speech. In 1987 he was appointed to the position of Lincoln Laboratory Senior Staff. On retiring from Lincoln Laboratory in 1996, he accepted the position of Senior Scientist at Voxware to develop high-quality speech products for the Internet. In 1978 he received the M. Barry Carlton Award for the best paper published in the IEEE Transactions on Aerospace and Electronic Systems for the paper “Interferometer Design for Elevation Angle Estimation”. In 1990 he received the IEEE Signal Processing Society’s Senior Award for the paper “Speech Analysis/Synthesis Based on a Sinusoidal Representation”, published in the IEEE Transactions on Acoustics, Speech and Signal Processing.

Dana C. Massie studied electronic music synthesis and composition at Virginia Commonwealth University in Richmond Virginia, and electrical engineering at Virginia Polytechnic Institute and State University in Blacksburg, VA. He worked in professional analog recording console and digital telecom systems design at Datatronix, Inc., in Reston, VA from 1981 through 1983. He then moved to E-mu Systems, Inc., in California, to design DSP algorithms and architectures for electronic music. After brief stints at NeXT Computer, Inc. and WaveFrame, Inc., developing MultiMedia DSP applications, he returned to E-mu Systems to work in digital filter design, digital reverberation design, and advanced music synthesis algorithms. He is now the Director of the Joint E-mu/Creative Technology Center, in Scotts Valley, California. The “Tech Center” develops advanced audio technologies for both E-mu Systems and Creative Technology, Limited in Singapore, including VLSI designs, advanced music synthesis algorithms, 3D audio algorithms, and software tools.

Thomas F. Quatieri was born in Somerville, Massachusetts on January 31, 1952. He received the B.S. degree from Tufts University, Medford, Massachusetts in 1973, and the SM., E.E., and Sc.D. degrees from the Massachusetts Institute of Technology (M.I.T.), Cambridge, Massachusetts in 1975, 1977, and 1979, respectively. He is currently a senior research staff member at M.I.T. Lincoln Laboratory, Lexington, Massachusetts. In 1980, he joined the Sensor Processing Technology Group of M.I.T., Lincoln Laboratory, Lexington, Massachusetts where he worked on problems in multi-dimensional digital signal processing and image processing. Since 1983 he has been a member of the Speech Systems Technology Group at Lincoln Laboratory where he has been involved in digital signal processing for speech and audio applications, underwater sound enhancement, and data communications. He has contributed many publications to journals and conference proceedings, written several patents, and co-authored chapters in numerous edited books including: *Advanced Topics in Signal Processing* (Prentice Hall, 1987), *Advances in Speech Signal Processing* (Marcel Dekker, 1991), and *Speech Coding and Synthesis* (Elsevier, 1995). He holds the position of Lecturer at MIT where he has developed the graduate course *Digital Speech Processing,* and is active in advising graduate students on the MIT campus. Dr. Quatieri is the recipient of the 1982 Paper Award of the IEEE Acoustics, Speech and Signal Processing Society for the paper, “Implementation of 2-D Digital Filters by Iterative Methods”. In 1990, he received the IEEE Signal Processing Society’s Senior Award for the paper, “Speech Analysis/Synthesis Based on a Sinusoidal Representation”, published in the *IEEE Transactions on Acoustics, Speech and Signal Processing,* and in 1994 won this same award for the paper “Energy Separation in Signal Modulations with Application to Speech Analysis” which was also selected for the 1995 IEEE W.R.G. Baker Prize Award. He was a member of the IEEE Digital Signal Processing Technical Committee, from 1983 to 1992 served on the steering committee for the bi-annual Digital Signal Processing Workshop, and was Associate Editor for the *IEEE Transactions on Signal Processing* in the area of nonlinear systems.

Peter J.W. Rayner received the M.A. degree from Cambridge University, U.K., in 1968 and the Ph. D. degree from Aston University in 1969. Since 1968 he has been with the Department of Engineering at Cambridge University and is Head of the Signal Processing and Communications Research Group. In 1990 he was appointed to an ad-hominem Readership in Information Engineering. He teaches course in random signal theory, digital signal processing, image processing and communication systems. His current research interests include image sequence restoration, audio restoration, non-linear estimation and detection and time series modelling and classification.

Julius O. Smith received the B.S.E.E. degree from Rice University, Houston, TX, in 1975. He received the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1978 and 1983, respectively. His Ph.D. research involved the application of digital signal processing and system identification techniques to the modeling and synthesis of the violin, clarinet, reverberant spaces, and other musical systems. From 1975 to 1977 he worked in the Signal Processing Department at ESL in Sunnyvale, CA, on systems for digital communications. From 1982 to 1986 he was with the Adaptive Systems Department at Systems Control Technology in Palo Alto, CA, where he worked in the areas of adaptive filtering and spectral estimation. From 1986 to 1991 he was employed at NeXT Computer, Inc., responsible for sound, music, and signal processing software for the NeXT computer workstation. Since then he has been an Associate Professor at the Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, teaching courses in signal processing and music technology, and pursuing research in signal processing techniques applied to musical instrument modeling, audio spectral modeling, and related topics.

INTRODUCTION

Karlheinz Brandenburg and Mark Kahrs

With the advent of multimedia, digital signal processing (DSP) of sound has emerged from the shadow of bandwidth-limited speech processing. Today, the main applications of audio DSP are high quality audio coding and the digital generation and manipulation of music signals. They share common research topics including perceptual measurement techniques and analysis/synthesis methods. Smaller but nonetheless very important topics are hearing aids using signal processing technology and hardware architectures for digital signal processing of audio. In all these areas the last decade has seen a significant amount of application oriented research.

The topics covered here coincide with the topics covered in the biannual workshop on “Applications of Signal Processing to Audio and Acoustics”. This event is sponsored by the IEEE Signal Processing Society (Technical Committee on Audio and Electroacoustics) and takes place at Mohonk Mountain House in New Paltz, New York.

A short overview of each chapter will illustrate the wide variety of technical material presented in the chapters of this book.

John Beerends: Perceptual Measurement Techniques. The advent of perceptual measurement techniques is a byproduct of the advent of digital coding for both speech and high quality audio signals. Traditional measurement schemes are bad estimates for the subjective quality after digital coding/decoding. Listening tests are subject to statistical uncertainties and the basic question of repeatability in a different environment. John Beerends explains the reasons for the development of perceptual measurement techniques, the psychoacoustic fundamentals which apply to both perceptual measurement and perceptual coding and explains some of the more advanced techniques which have been developed in the last few years. Completed and ongoing standardization efforts concludes his chapter. This is recommended reading not only to people interested in perceptual coding and measurement but to anyone who wants to know more about the psychoacoustic fundamentals of digital processing of sound signals.

Karlheinz Brandenburg: Perceptual Coding of High Quality Digital Audio. High quality audio coding is rapidly progressing from a research topic to widespread applications. Research in this field has been driven by a standardization process within the Motion Picture Experts Group (MPEG). The chapter gives a detailed introduction of the basic techniques including a study of filter banks and perceptual models. As the main example, MPEG Audio is described in full detail. This includes a description of the new MPEG-2 Advanced Audio Coding (AAC) standard and the current work on MPEG-4 Audio.

William G. Gardner: Reverberation Algorithms. This chapter is the first in a number of chapters devoted to the digital manipulation of music signals. Digitally generated reverb was one of the first application areas of digital signal processing to high quality audio signals. Bill Gardner gives an in depth introduction to the physical and perceptual aspects of reverberation. The remainder of the chapter treats the different types of artificial reverberators known today. The main quest in this topic is to generate natural sounding reverb with low cost. Important milestones in the research, various historic and current types of reverberators are explained in detail.

Simon Godsill, Peter Rayner and Olivier Cappé: Digital Audio Restoration. Digital signal processing of high quality audio does not stop with the synthesis or manipulation of new material: One of the early applications of DSP was the manipulation of sounds from the past in order to restore them for recording on new or different media. The chapter presents the different methods for removing clicks, noise and other artifacts from old recordings or film material.

Mark Kahrs: Digital Audio System Architecture. An often overlooked part of the processing of high quality audio is the system architecture. Mark Kahrs introduces current technologies both for the conversion between analog and digital world and the processing technologies. Over the years there is a clear path from specialized hardware architectures to general purpose computing engines. The chapter covers specialized hardware architectures as well as the use of generally available DSP chips. The emphasis is on high throughput digital signal processing architectures for music synthesis applications.

James M. Kates: Signal Processing for Hearing Aids. A not so obvious application area for audio signal processing is the field of hearing aids. Nonetheless this field has seen continuous research activities for a number of years and is another field where widespread application of digital technologies is under preparation today. The chapter contains an in-depth treatise of the basics of signal processing for hearing aids including the description of different types of hearing loss, simpler amplification and compression techniques and current research on multi-microphone techniques and cochlear implants.

Jean Laroche: Time and Pitch Scale Modification of Audio Signals. One of the conceptionally simplest problems of the manipulation of audio signals is difficult enough to warrant ongoing research for a number of years: Jean Laroche explains the basics of time and pitch scale modification of audio signals for both speech and musical signals. He discusses both time domain and frequency domain methods including methods specially suited for speech signals.

Dana C. Massie: Wavetable Sampling Synthesis. The most prominent example today of the application of high quality digital audio processing is wavetable sampling synthesis. Tens of millions of computer owners have sound cards incorporating wavetable sampling synthesis. Dana Massie explains the basics and modern technologies employed in sampling synthesis.

T.F. Quatieri and R.J. McAulay: Audio Signal Processing Based on Sinusoidal Analysis/Synthesis. One of the basic paradigms of digital audio analysis, coding (i.e. analysis/synthesis) and synthesis systems is the sinusoidal model. It has been used for many systems from speech coding to music synthesis. The chapter contains the unified view of both the basics of sinusoidal analysis/synthesis and some of the applications.

Julius O. Smith III: Principles of Digital Waveguide Models of Musical Instruments. This chapter describes a recent research topic in the synthesis of music instruments: Digital waveguide models are one method of physical modeling. As in the case of the Vocoder for speech, a model of an existing or hypothetical instrument is used for the sound generation. In the tutorial the vibrating string is taken as the principle illustrative example. Another example using the same underlying principles is the acoustic tube. Complicated instruments are derived by adding signal scattering and reed-bore or bow-string interactions.

Summary This book was written to serve both as a text book for an advanced graduate course on digital signal processing for audio or as a reference book for the practicing engineer. We hope that this book will stimulate further research and interest in this fascinating and exciting field.