Editor UDO ZÖLZER

DAFX

Digital Audio Effects

Second Edition

WILEY

DAFX: Digital Audio Effects

Second Edition

DAFX: Digital Audio Effects

Second Edition

Edited by

Udo Zölzer

Helmut Schmidt University – University of the Federal Armed Forces, Hamburg, Germany

WILEY

A John Wiley and Sons, Ltd., Publication

This edition first published 2011

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloguing-in-Publication Data

Zölzer, Udo.
DAFX : digital audio effects / Udo Zölzer. – 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-66599-2 (hardback)
1. Computer sound processing. 2. Sound–Recording and reproducing–Digital techniques.
3. Signal processing–Digital techniques. I. Title.
TK5105.8863.Z65 2011
006.5 – dc22
2010051411

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

Print ISBN: 978-0-470-66599-2 [HB]
e-PDF ISBN: 978-1-119-99130-4
o-Book ISBN: 978-1-119-99129-8
e-Pub ISBN: 978-0-470-97967-9

Typeset in 9/11pt Times by Laserwords Private Limited, Chennai, India

Contents

Preface

DAFX is a synonym for digital audio effects. It is also the name for a European research project for co-operation and scientific transfer, namely EU-COST-G6 “Digital Audio Effects” (1997–2001). It was initiated by Daniel Arfib (CNRS, Marseille). In the past couple of years we have had four EU-sponsored international workshops/conferences on DAFX, namely, in Barcelona (DAFX-98), Trondheim (DAFX-99), Verona (DAFX-00) and Limerick (DAFX-01). A variety of DAFX topics have been presented by international participants at these conferences. The papers can be found on the corresponding web sites.

This book not only reflects these conferences and workshops, it is intended as a profound collection and presentation of the main fields of digital audio effects. The contents and structure of the book were prepared by a special book work group and discussed in several workshops over the past years sponsored by the EU-COST-G6 project. However, the single chapters are the individual work of the respective authors.

Chapter 1 gives an introduction to digital signal processing and shows software implementations with the MATLAB® programming tool. Chapter 2 discusses digital filters for shaping the audio spectrum and focuses on the main building blocks for this application. Chapter 3 introduces basic structures for delays and delay-based audio effects. In Chapter 4 modulators and demodulators are introduced and their applications to digital audio effects are demonstrated. The topic of nonlinear processing is the focus of Chapter 5. First, we discuss fundamentals of dynamics processing such as limiters, compressors/expanders and noise gates, and then we introduce the basics of nonlinear processors for valve simulation, distortion, harmonic generators and exciters. Chapter 6 covers the wide field of spatial effects starting with basic effects, 3D for headphones and loudspeakers, reverberation and spatial enhancements. Chapter 7 deals with time-segment processing and introduces techniques for variable speed replay, time stretching, pitch shifting, shuffling and granulation. In Chapter 8 we extend the time-domain processing of Chapters 2–7. We introduce the fundamental techniques for time-frequency processing, demonstrate several implementation schemes and illustrate the variety of effects possible in the 2D time-frequency domain. Chapter 9 covers the field of source-filter processing, where the audio signal is modeled as a source signal and a filter. We introduce three techniques for source-filter separation and show source-filter transformations leading to audio effects such as cross-synthesis, formant changing, spectral interpolation and pitch shifting with formant preservation. The end of this chapter covers feature extraction techniques. Chapter 10 deals with spectral processing, where the audio signal is represented by spectral models such as sinusoids plus a residual signal. Techniques for analysis, higher-level feature analysis and synthesis are introduced, and a variety of new audio effects based on these spectral models are discussed. Effect applications range from pitch transposition, vibrato, spectral shape shift and gender change to harmonizer and morphing effects. Chapter 11 deals with fundamental principles of time and frequency warping techniques for deforming the time and/or the frequency axis. Applications of these techniques are presented for pitch-shifting inharmonic sounds, the inharmonizer, extraction of excitation signals, morphing and classical effects. Chapter 12 deals with the control of effect processors ranging from general control techniques to control based on sound features and gestural interfaces. Finally, Chapter 13 illustrates new challenges of bitstream signal representations, shows the fundamental basics and introduces filtering concepts for bitstream signal processing. MATLAB implementations in several chapters of the book illustrate software implementations of DAFX algorithms. The MATLAB files can be found on the web site http://www.dafx.de.

I hope the reader will enjoy the presentation of the basic principles of DAFX in this book and will be motivated to explore DAFX with the help of our software implementations. The creativity of a DAFX designer can only grow or emerge if intuition and experimentation are combined with profound knowledge of physical and musical fundamentals. The implementation of DAFX in software needs some knowledge of digital signal processing and this is where this book may serve as a source of ideas and implementation details.

I would like to thank the authors for their contributions to the chapters and also the EU-Cost-G6 delegates from all over Europe for their contributions during several meetings, especially Nicola Bernadini, Javier Casajús, Markus Erne, Mikael Fernström, Eric Feremans, Emmanuel Favreau, Alois Melka, Jøran Rudi and Jan Tro. The book cover is based on a mapping of a time-frequency representation of a musical piece onto the globe by Jøran Rudi. Thanks to Catja Schümann for her assistance in preparing drawings and LaTeX formatting, Christopher Duxbury for proof-reading and Vincent Verfaille for comments and cleaning up the code lines of Chapters 8 to 10. I also express my gratitude to my staff members Udo Ahlvers, Manfred Chrobak, Florian Keiler, Harald Schorr and Jörg Zeller for providing assistance during the course of writing this book. Finally, I would like to thank Birgit Gruber, Ann-Marie Halligan, Laura Kempster, Susan Dunsmore and Zoë Pinnock from John Wiley & Sons, Ltd for their patience and assistance.

My special thanks are directed to my wife Elke and our daughter Franziska.

Hamburg, March 2002 Udo Zölzer

Preface 2nd Edition

This second edition is the result of an ongoing DAFX conference series over the past years. Each chapter has new contributing co-authors who have gained experience in the related fields over the years. New emerging research fields are introduced by four new Chapters on Adaptive-DAFX, Virtual Analog Effects, Automatic Mixing and Sound Source Separation. The main focus of the book is still the audio effects side of audio research. The book offers a variety of proven effects and shows directions for new audio effects. The MATLAB files can be found on the web site http://www.dafx.de.

I would like to thank the co-authors for their contributions and effort, Derry FitzGerald and Nuno Fonseca for their contributions to the book and finally, thanks go to Nicky Skinner, Alex King, and Georgia Pinteau from John Wiley & Sons, Ltd for their assistance.

Hamburg, September 2010 Udo Zölzer

List of Contributors

Jonathan S. Abel is a Consulting Professor at the Center for Computer Research in Music and Acoustics (CCRMA) in the Music Department at Stanford University, where his research interests include audio and music applications of signal and array processing, parameter estimation and acoustics. From 1999 to 2007, Abel was a co-founder and chief technology officer of the Grammy Award-winning Universal Audio, Inc. He was a researcher at NASA/Ames Research Center, exploring topics in room acoustics and spatial hearing on a grant through the San Jose State University Foundation. Abel was also chief scientist of Crystal River Engineering, Inc., where he developed their positional audio technology, and a lecturer in the Department of Electrical Engineering at Yale University. As an industry consultant, Abel has worked with Apple, FDNY, LSI Logic, NRL, SAIC and Sennheiser, on projects in professional audio, GPS, medical imaging, passive sonar and fire department resource allocation. He holds PhD and MS degrees from Stanford University, and an SB from MIT, all in electrical engineering. Abel is a Fellow of the Audio Engineering Society.

Xavier Amatriain is Researcher in Telefonica R&D Barcelona which he joined in June 2007. His current focus of research is on recommender systems and other web science-related topics. He is also associate Professor at Universitat Pompeu Fabra, where he teaches software engineering and information retrieval. He has authored more than 50 publications, including several book chapters and patents. Previous to this, Dr. Amatriain worked at the University of California Santa Barbara as Research Director, supervising research on areas that included multimedia and immersive systems, virtual reality and 3D audio and video. Among others, he was Technical Director of the Allosphere project and he lectured in the media arts and technology program. During his PhD at the UPF (Barcelona), he was a researcher in the Music Technology Group and he worked on music signal processing and systems. At that time he initiated and co-ordinated the award-winning CLAM open source project for audio and music processing.

Daniel Arfib (1949–) received his diploma as “ingénieur ECP” from the Ecole Centrale of Paris in 1971 and is a “docteur-ingénieur” (1977) and “docteur es sciences” (1983) from the Université of Marseille II. After a few years in education or industry jobs, he has devoted his work to research, joining the CNRS (National Center for Scientific Research) in 1978 at the Laboratory of Mechanics and Acoustics (LMA) in Marseille (France). His main concern is to provide a combination of scientific and musical points of view on synthesis, transformation and interpretation of sounds using the computer as a tool, both as a researcher and a composer. As the chairman of the COST-G6 action named “Digital Audio Effects” he has been in the middle of a galaxy of researchers working on this subject. He also has a strong interest in the gesture and sound relationship, especially concerning creativity in musical systems. Since 2008, he is working in the field of sonic interaction design at the Laboratory of Informatics (LIG) in Grenoble, France.

David Berners is a Consulting Professor at the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University, where he has taught courses in signal processing and audio effects since 2004. He is also Chief Scientist at Universal Audio, Inc., a hardware and software manufacturer for the professional audio market. At UA, Dr Berners leads research and development efforts in audio effects processing, including dynamic range compression, equalization, distortion and delay effects, and specializing in modeling of vintage analog equipment. Dr Berners has previously held positions at the Lawrence Berkeley Laboratory, NASA Jet Propulsion Laboratory and Allied Signal. He received his PhD from Stanford University, MS from the California Institute of Technology, and his SB from Massachusetts Institute of Technology, all in electrical engineering.

Stefan Bilbao received his BA in Physics at Harvard University (1992), then spent two years at the Institut de Recherche et Coordination Acoustique Musicale (IRCAM) under a fellowship awarded by Harvard and the Ecole Normale Superieure. He then completed the MSc and PhD degrees in Electrical Engineering at Stanford University (1996 and 2001, respectively), while working at the Center for Computer Research in Music and Acoustics (CCRMA). He was subsequently a postdoctoral researcher at the Stanford Space Telecommunications and Radioscience Laboratory, and a lecturer at the Sonic Arts Research Centre at the Queen's University Belfast. He is currently a senior lecturer in music at the University of Edinburgh.

Jordi Bonada (1973–) received an MSc degree in electrical engineering from the Universitat Politècnica de Catalunya (Barcelona, Spain) in 1997, and a PhD degree in computer science and digital communications from the Universitat Pompeu Fabra (Barcelona, Spain) in 2009. Since 1996 he has been a researcher at the Music Technology Group of the same university, while leading several collaboration projects with Yamaha Corp. He is mostly interested in the field of spectral-domain audio signal processing, with focus on time scaling and singing-voice modeling and synthesis.

Giovanni De Poli is an Associate Professor of computer science at the Department of Electronics and Informatics of the University of Padua, where he teaches "Data Structures and Algorithms" and "Processing Systems for Music". He is the Director of the Centro di Sonologia Computazionale (CSC) of the University of Padua. He is a member of the Executive Committee (ExCom) of the IEEE Computer Society Technical Committee on Computer Generated Music, a member of the board of directors of AIMI (Associazione Italiana di Informatica Musicale), a member of the board of directors of CIARM (Centro Interuniversitario di Acustica e Ricerca Musicale), a member of the Scientific Committee of ACROE (Institut National Politechnique Grenoble), and Associate Editor of the *International Journal of New Music Research*. His main research interests are in algorithms for sound synthesis and analysis, models for expressiveness in music, multimedia systems and human–computer interaction, and the preservation and restoration of audio documents. He is the author of several scientific international publications, and has served in the Scientific Committees of international conferences. He is co-editor of the books *Representations of Music Signals*, MIT Press 1991, and *Musical Signal Processing*, Swets & Zeitlinger, 1996. Systems and research developed in his lab have been exploited in collaboration with digital musical instruments industry (GeneralMusic). He is the owner of patents on digital music instruments.

Kristjan Dempwolf was born in Osterode am Harz, Germany, in 1978. After finishing an apprenticeship as an electronic technician in 2002 he studied electrical engineering at the Technical University Hamburg-Harburg (TUHH). He spent one semester at the Norwegian University of Science and Technology (NTNU) in 2006 and obtained his Diplom-Ingenieur degree in 2008. He is currently working on a doctoral degree at the Helmut Schmidt University – University of the Federal Armed Forces, Hamburg, Germany. His main research interests are real-time modeling and nonlinear audio systems.

Sascha Disch received his Diplom-Ingenieur degree in electrical engineering from the Technische Universität Hamburg-Harburg (TUHH), Germany in 1999. From 1999 to 2007 he was with the Fraunhofer Institut für Integrierte Schaltungen (FhG-IIS), Erlangen, Germany. At Fraunhofer, he worked in research and development in the field of perceptual audio coding and audio processing, including the MPEG standardization of parametric coding of multi-channel sound (MPEG Surround). From 2007 to 2010 he was a researcher at the Laboratorium für Informationstechnologie, Leibniz Universität Hannover (LUH), Germany and is also a PhD candidate. Currently, he is again with Fraunhofer and is involved with research and development in perceptual audio coding. His research interests include audio signal processing/coding and digital audio effects, primarily pitch shifting and time stretching.

Pierre Dutilleux graduated in thermal engineering from the Ecole Nationale Supérieure des Techniques Industrielles et des Mines de Douai (ENSTIMD) in 1983 and in information processing from the Ecole Nationale Supérieure d'Electronique et de Radioélectricité de Grenoble (ENSERG) in 1985. From 1985 to 1991, he developed audio and musical applications for the Syter real-time audio processing system designed at INA-GRM by J.-F.Allouis. After developing a set of audio-processing algorithms as well as implementing the first wavelet analyser on a digital signal processor, he got a PhD in acoustics and computer music from the university of Aix-Marseille II in 1991 under the direction of J.-C.Risset. From 1991 through to 2000 he worked as a research and development engineer at the ZKM (Center for Art and Media Technology) in Karlsruhe where he planned computer and digital audio networks for a large digital-audio studio complex, and he introduced live electronics and physical modeling as tools for musical production. He contributed to multimedia works with composers such as K. Furukawa and M. Maiguashca. He designed and realised the AML (Architecture and Music Laboratory) as an interactive museum installation. He has been a German delegate of the Digital Audio Effects (DAFX) project. In 2000 he changed his professional focus from music and signal processing to wind energy. He applies his highly differentiated listening skills to the characterisation of the noise from wind turbines. He has been Head of Acoustics at DEWI, the German Wind-Energy Institute. By performing diligent reviews of the acoustic issues of wind farm projects before construction, he can identify at an early stage the acoustic risks which might impair the acceptance of the future wind farm projects by neighbours.

Gianpaolo Evangelista is Professor in Sound Technology at the Linköping University, Sweden, where he has headed the Sound and Video Technology research group since 2005. He received the Laurea in physics (summa cum laude) from "Federico II" University of Naples, Italy, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of California, Irvine. He has previously held positions at the Centre d'Etudes de Mathématique et Acoustique Musicale (CEMAMu/CNET), Paris, France; the Microgravity Advanced Research and Support (MARS) Center, Naples, Italy; the University of Naples Federico II and the Laboratory for Audiovisual Communications, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. He is the author or co-author of about 100 journal or conference papers and book chapters. He is a senior member of the IEEE and an active member of the DAFX (Digital Audio Effects) Scientific Committee. His interests are centered in audio signal representations, sound synthesis by physical models, digital audio effects, spatial audio, audio coding, wavelets and multirate signal processing.

Martin Holters was born in Hamburg, Germany, in 1979. He received the Master of Science degree from Chalmers Tekniska Högskola, Göteborg, Sweden, in 2003 and the Diplom-Ingenieur degree in computer engineering from the Technical University Hamburg-Harburg, Germany, in 2004. He then joined the Helmut-Schmidt-University – University of the Federal Armed Forces, Hamburg, Germany where he received the Dr-Ingenieur degree in 2009. The topic of his dissertation was delay-free audio coding based on adaptive differential pulse code modulation (ADPCM) with adaptive pre- and post-filtering. Since 2009 he has been chief scientist in the department of signal processing and communications. He is active in various fields of audio signal processing research with his main focus still on audio coding and transmission.

Florian Keiler was born in Hamburg, Germany, in 1972. He received the Diplom-Ingenieur degree in electrical engineering from the Technical University Hamburg-Harburg (TUHH) in 1999 and the Dr.-Ingenieur degree from the Helmut-Schmidt-University – University of the Federal Armed Forces, Hamburg, Germany in 2006. The topic of his dissertation was low-delay audio coding based on linear predictive coding (LPC) in subbands. Since 2005 he has been working in the audio and acoustics research laboratory of Technicolor (formerly Thomson) located in Hanover, Germany. He is currently working in the field of spatial audio.

Tapio Lokki was born in Helsinki, Finland, in 1971. He has studied acoustics, audio signal processing, and computer science at the Helsinki University of Technology (TKK) and received an MSc degree in electrical engineering in 1997 and a DSc (Tech.) degree in computer science and engineering in 2002. At present Dr. Lokki is an Academy Research Fellow with the Department of Media Technology at Aalto University. In addition, he is an adjunct professor at the Department of Signal Processing and Acoustics at Aalto. Dr. Lokki leads his virtual acoustics team which aims to create novel objective and subjective ways to evaluate concert hall acoustics. In addition, the team develops physically based room acoustics modeling methods to obtain authentic auralization. Furthermore, the team studies augmented reality audio and eyes-free user interfaces. The team is funded by the Academy of Finland and by Dr Lokki's starting grant from the European Research Council (ERC). Dr. Lokki is a member of the editorial board of Acta Acustica united with Acustica. Dr. Lokki is a member of the Audio Engineering Society, the IEEE Computer Society, and Siggraph. In addition, he is the president of the Acoustical Society of Finland.

Alex Loscos received BS and MS degrees in signal processing engineering in 1997. In 1998 he joined the Music Technology Group of the Universitat Pompeu Fabra of Barcelona. After a few years as a researcher, lecturer, developer and project manager he co-founded Barcelona Music & Audio Technologies in 2006, a spin-off company of the research lab. In 2007 he gained a PhD in computer science and immediately started as Chief Strategy Officer at BMAT. A year and a half later he took over the position of Chief Executive Officer which he currently holds. Alex is also passionate about music, an accomplished composer and a member of international distribution bands.

Sylvain Marchand has been an associate professor in the image and sound research team of the LaBRI (Computer Science Laboratory), University of Bordeaux 1, since 2001. He is also a member of the "Studio de Création et de Recherche en Informatique et Musique Électroacoustique" (SCRIME). Regarding the international DAFX (Digital Audio Effects) conference, he has been a member of the Scientific Committee since 2006, Chair of the 2007 conference held in Bordeaux and has attended all DAFX conferences since the first one in 1998–where he gave his first presentation, as a Ph.D. student. Now, he is involved in several international conferences on musical audio, and he is also associate editor of the *IEEE Transactions on Audio, Speech, and Language Processing*. Dr Marchand is particularly involved in musical sound analysis, transformation, and synthesis. He focuses on spectral representations, taking perception into account. Among his main research topics are sinusoidal models, analysis/synthesis of deterministic and stochastic sounds, sound localization/spatialization ("3D sound"), separation of sound entities (sources) present in polyphonic music, or "active listening" (enabling the user to interact with the musical sound while it is played).

Jyri Pakarinen (1979–) received MSc and DSc (Tech.) degrees in acoustics and audio signal processing from the Helsinki University of Technology, Espoo, Finland, in 2004 and 2008, respectively. He is currently working as a post-doctoral researcher and a lecturer in the Department of Signal Processing and Acoustics, Aalto University School of Science and Technology. His main research interests are digital emulation of electric audio circuits, sound synthesis through physical modeling, and vibro- and electroacoustic measurements. As a semiprofessional guitar player, he is also interested and involved in music activities.

Enrique Perez Gonzalez was born in 1978 in Mexico City. He studied engineering communications and electronics at the ITESM University in Mexico City, where he graduated in 2002. During his engineering studies he did a one-year internship at RMIT in Melbourne, Australia where he specialized in Audio. From 1999 to 2005 he worked at the audio rental company SAIM, one of the biggest audio companies in Mexico, where he worked as a technology manager and audio system engineer for many international concerts. He graduated with distinction with an MSc in music technology at the University of York in 2006, where he worked on delta sigma modulation systems. He completed his PhD in 2010 on Advanced Tools for Automatic Mixing at the Centre for Digital Music in Queen Mary, University of London.

Mark Plumbley has investigated audio and music signal analysis, including beat tracking, music transcription, source separation and object coding, using techniques such as neural networks, independent component analysis, sparse representations and Bayesian modeling. Professor Plumbley joined Queen Mary, University of London (QMUL) in 2002, he holds an EPSRC Leadership Fellowship on Machine Listening using Sparse Representations, and in September 2010 became Director of the Centre for Digital Music at QMUL. He is chair of the International Independent Component Analysis (ICA) Steering Committee, a member of the IEEE Machine Learning in Signal Processing Technical Committee, and an Associate Editor for *IEEE Transactions on Neural Networks*.

Ville Pulkki received his MSc and DSc (Tech.) degrees from Helsinki University of Technology in 1994 and 2001, respectively. He majored in acoustics, audio signal processing and information sciences. Between 1994 and 1997 he was a full time student at the Department of Musical Education at the Sibelius Academy. In his doctoral dissertation he developed vector base amplitude panning (VBAP), which is a method for positioning virtual sources to any loudspeaker configuration. In addition, he studied the performance of VBAP with psychoacoustic listening tests and with modeling of auditory localization mechanisms. The VBAP method is now widely used in multichannel virtual auditory environments and in computer music installations. Later, he developed with his group, a method for spatial sound reproduction and coding, directional audio coding (DirAC). DirAC takes coincident first-order microphone signals as input, and processes output to arbitrary loudspeaker layouts or to headphones. The method is currently being commercialized. Currently, he is also developing a computational functional model of the brain organs devoted to binaural hearing, based on knowledge from neurophysiology, neuroanatomy, and from psychoacoustics. He is leading a research group in Aalto University (earlier: Helsinki University of Technology, TKK or HUT), which consists of 10 researchers. The group also conducts research on new methods to measure head-related transfer functions, and conducts psychoacoustical experiments to better understand the spatial sound perception by humans. Dr. Pulkki enjoys being with his family (wife and two children), playing various musical instruments, and building his summer place. He is the Northern Region Vice President of AES and the co-chair of the AES Technical Committee on Spatial Audio.

Josh Reiss is a senior lecturer with the Centre for Digital Music at Queen Mary, University of London. He received his PhD in physics from Georgia Tech. He made the transition to audio and musical signal processing through his work on sigma delta modulators, which led to patents and a nomination for a best paper award from the IEEE. He has investigated music retrieval systems, time scaling and pitch-shifting techniques, polyphonic music transcription, loudspeaker design, automatic mixing for live sound and digital audio effects. Dr. Reiss has published over 80 scientific papers and serves on several steering and technical committees. As coordinator of the EASAIER project, he led an international consortium of seven partners working to improve access to sound archives in museums, libraries and cultural heritage institutions. His primary focus of research, which ties together many of the above topics, is on state-of-the-art signal processing techniques for professional sound engineering.

Davide Rocchesso received the PhD degree from the University of Padua, Italy, in 1996. Between 1998 and 2006 he was with the Computer Science Department at the University of Verona, Italy, as an Assistant and Associate Professor. Since 2006 he has been with the Department of Art and Industrial Design of the IUAV University of Venice, as Associate Professor. He has been the coordinator of EU project SOb (the Sounding Object) and local coordinator of the EU project CLOSED (Closing the Loop Of Sound Evaluation and Design) and of the Coordination Action $S2S^2$ (Sound-to-Sense; Sense-to-Sound). He has been chairing the COST Action IC-0601 SID (Sonic Interaction Design). Davide Rocchesso authored or co-authored over one hundred publications in scientific journals, books, and conferences. His main research interests are sound modelling for interaction design, sound synthesis by physical modelling, and design and evaluation of interactions.

Xavier Serra is Associate Professor of the Department of Information and Communication Technologies and Director of the Music Technology Group at the Universitat Pompeu Fabra in Barcelona. After a multidisciplinary academic education he obtained a PhD in computer music from Stanford University in 1989 with a dissertation on the spectral processing of musical sounds that is considered a key reference in the field. His research interests cover the understanding, modeling and generation of musical signals by computational means, with a balance between basic and applied research and approaches from both scientific/technological and humanistic/artistic disciplines.

Julius O. Smith teaches a music signal-processing course sequence and supervises related research at the Center for Computer Research in Music and Acoustics (CCRMA). He is formally a Professor of music and Associate Professor (by courtesy) of electrical engineering at Stanford University. In 1975, he received his BS/EE degree from Rice University, where he got a solid start in the field of digital signal processing and modeling for control. In 1983, he received the PhD/EE degree from Stanford University, specializing in techniques for digital filter design and system identification, with application to violin modeling. His work history includes the Signal Processing Department at Electromagnetic Systems Laboratories, Inc., working on systems for digital communications; the Adaptive Systems Department at Systems Control Technology, Inc., working on research problems in adaptive filtering and spectral estimation, and NeXT Computer, Inc., where he was responsible for sound, music, and signal processing software for the NeXT computer workstation. Professor Smith is a Fellow of the Audio Engineering Society and the Acoustical Society of America. He is the author of four online books and numerous research publications in his field.

Vesa Välimäki (1968–) is Professor of Audio Signal Processing at the Aalto University, Department of Signal Processing and Acoustics, Espoo, Finland. He received the Doctor of Science in technology degree from Helsinki University of Technology (TKK), Espoo, Finland, in 1995. He has published more than 200 papers in international journals and conferences. He has organized several special issues in scientific journals on topics related to musical signal processing. He was the chairman of the 11th International Conference on Digital Audio Effects (DAFX-08), which was held in Espoo in 2008. During the academic year 2008–2009 he was on sabbatical leave under a grant from the Academy of Finland and spent part of the year as a Visiting Scholar at the

Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, CA. He currently serves as an Associate Editor of the *IEEE Transactions on Audio, Speech and Language Processing*. His research interests are sound synthesis, audio effects processing, digital filters, and musical instrument acoustics.

Vincent Verfaille (1974–) studied applied mathematics at INSA (Toulouse, France) to become an engineer in 1997. He then adapted to a carrier change, where he studied music technology (DEA-ATIAM, Université Paris VI, France, 2000; PhD in music technology at CNRS-LMA and Université Aix-Marseille II, France, 2003) and adaptive audio effects. He then spent a few years (2003–2009) as a post-doctoral researcher and then as a research associate in both the Sound Processing and Control Lab (SPCL) and the Input Device for Musical Interaction Lab (IDMIL) at the Schulich School of Music (McGill University, CIRMMT), where he worked on sound synthesis and control. He also taught digital audio effects and sound transformation at ENSEIRB and Université Bordeaux I (Bordeaux, France, 2002–2006), signal processing at McGill University (Montreal, Canada, 2006) and musical acoustics at University of Montréal (Montréal, Canada, 2008). He is now doing another carrier change, far away from computers and music.

Emmanuel Vincent received the BSc degree in mathematics from École Normale Supérieure in 2001 and the PhD degree in acoustics, signal processing and computer science applied to music from Université Pierre et Marie Curie, Paris, France, in 2004. After working as a research assistant with the Center for Digital Music at Queen Mary College, London, UK, he joined the French National Research Institute for Computer Science and Control (INRIA) in 2006 as a research scientist. His research focuses on probabilistic modeling of audio signals applied to source separation, information retrieval and coding. He is the founding chair of the annual Signal Separation Evaluation Campaign (SiSEC) and a co-author of the toolboxes BSS Eval and BSS Oracle for the evaluation of source separation systems.

Adrian von dem Knesebeck (1982–) received his Diplom-Ingenieur degree in electrical engineering from the Technical University Hamburg-Harburg (TUHH), Germany in 2008. Since 2009 he has been working as a research assistant in the Department of Signal Processing and Communications at the Helmut Schmidt University – University of the Federal Armed Forces in Hamburg, Germany. He was involved in several audio research projects and collaboration projects with external companies so far and is currently working on his PhD thesis.

Udo Zölzer (1958–) received the Diplom-Ingenieur degree in electrical engineering from the University of Paderborn in 1985, the Dr.-Ingenieur degree from the Technical University Hamburg-Harburg (TUHH) in 1989 and completed a Habilitation in communications engineering at the TUHH in 1997. Since 1999 he has been a Professor and Head of the Department of Signal Processing and Communications at the Helmut Schmidt University – University of the Federal Armed Forces in Hamburg, Germany. His research interests are audio and video signal processing and communication. He is a member of the AES and the IEEE.

1

Introduction

V. Verfaille, M. Holters and U. Zölzer

1.1 Digital audio effects DAFX with MATLAB®

Audio effects are used by all individuals involved in the generation of musical signals and start with special playing techniques by musicians, merge to the use of special microphone techniques and migrate to effect processors for synthesizing, recording, production and broadcasting of musical signals. This book will cover several categories of sound or audio effects and their impact on sound modifications. Digital audio effects – as an acronym we use DAFX – are boxes or software tools with input audio signals or sounds which are modified according to some sound control parameters and deliver output signals or sounds (see Figure 1.1). The input and output signals are monitored by loudspeakers or headphones and some kind of visual representation of the signal, such as the time signal, the signal level and its spectrum. According to acoustical criteria the sound engineer or musician sets his control parameters for the sound effect he would like to achieve. Both input and output signals are in digital format and represent analog audio signals. Modification of the sound characteristic of the input signal is the main goal of digital audio effects. The settings of the control parameters are often done by sound engineers, musicians (performers, composers, or digital instrument makers) or simply the music listener, but can also be part of one specific level in the signal processing chain of the digital audio effect.

The aim of this book is the description of digital audio effects with regard to:

- *Physical and acoustical effect:* we take a short look at the physical background and explanation. We describe analog means or devices which generate the sound effect.
- *Digital signal processing:* we give a formal description of the underlying algorithm and show some implementation examples.
- *Musical applications:* we point out some applications and give references to sound examples available on CD or on the web.

DAFX: Digital Audio Effects, Second Edition. Edited by Udo Zölzer.

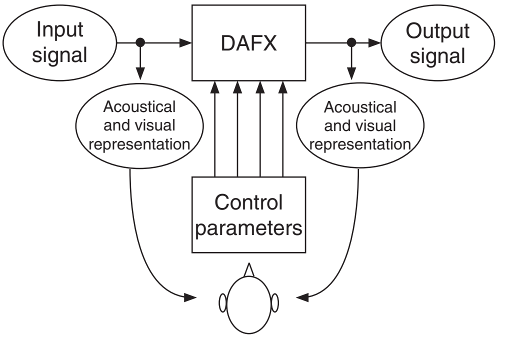

Figure 1.1 Digital audio effect and its control [Arf99].

The physical and acoustical phenomena of digital audio effects will be presented at the beginning of each effect description, followed by an explanation of the signal processing techniques to achieve the effect, some musical applications and the control of effect parameters.

In this introductory chapter we next introduce some vocabulary clarifications, and then present an overview of classifications of digital audio effects. We then explain some simple basics of digital signal processing and show how to write simulation software for audio effects processing with the **MATLAB**[1] simulation tool or freeware simulation tools[2]. **MATLAB** implementations of digital audio effects are a long way from running in real time on a personal computer or allowing real-time control of its parameters. Nevertheless the programming of signal processing algorithms and in particular sound-effect algorithms with **MATLAB** is very easy and can be learned very quickly.

Sound effect, audio effect and sound transformation

As soon as the word "effect" is used, the viewpoint that stands behind is the one of the subject who is observing a phenomenon. Indeed, "effect" denotes an impression produced in the mind of a person, a change in perception resulting from a cause. Two uses of this word denote related, but slightly different aspects: "sound effects" and "audio effects." Note that in this book, we discuss the latter exclusively. The expression – "sound effects" – is often used to depict sorts of earcones (icons for the ear), special sounds which in production mode have a strong signature and which therefore are very easily identifiable. Databases of sound effects provide natural (recorded) and processed sounds (resulting from sound synthesis and from audio effects) that produce specific effects on perception used to simulate actions, interaction or emotions in various contexts. They are, for instance, used for movie soundtracks, for cartoons and for music pieces. On the other hand, the expression "audio effects" corresponds to the tool that is used to apply transformations to sounds in order to modify how they affect us. We can understand those two meanings as a shift of the meaning of "effect": from the perception of a change itself to the signal processing technique that is used to achieve this change of perception. This shift reflects a semantic confusion between the object (what is perceived) and the tool to make the object (the signal processing technique). "Sound effect" really deals with the subjective viewpoint, whereas "audio effect" uses a subject-related term (effect) to talk about an objective reality: the tool to produce the sound transformation.

Historically, it can arguably be said that audio effects appeared first, and sound transformations later, when this expression was tagged on refined sound models. Indeed, techniques that made use of an analysis/transformation/synthesis scheme embedded a transformation step performed on a refined model of the sound. This is the technical aspect that clearly distinguishes "audio effects"

[1] http://www.mathworks.com
[2] http://www.octave.org

and "sound transformations," the former using a simple representation of the sound (samples) to perform signal processing, whereas the latter uses complex techniques to perform enhanced signal processing. Audio effects originally denoted simple processing systems based on simple operations, e.g. chorus by random control of delay line modulation; echo by a delay line; distortion by non-linear processing. It was assumed that audio effects process sound at its *surface*, since sound is represented by the wave form samples (which is not a high-level sound model) and simply processed by delay lines, filters, gains, etc. By surface we do not mean how strongly the sound is modified (it in fact can be deeply modified; just think of distortion), but we mean how far we go in unfolding the sound representations to be accurate and refined in the data and model parameters we manipulate. Sound transformations, on the other hand, denoted complex processing systems based on analysis/transformation/synthesis models. We, for instance, think of the phase vocoder with fundamental frequency tracking, the source-filter model, or the sinusoidal plus residual additive model. They were considered to offer deeper modifications, such as high-quality pitch-shifting with formant preservation, timbre morphing, and time-scaling with attack, pitch and panning preservation. Such deep manipulation of control parameters allows in turn the sound modifications to be heard as very subtle.

Over time, however, practice blurred the boundaries between audio effects and sound transformations. Indeed, several analysis/transformation/synthesis schemes can simply perform various processing that we consider to be audio effects. On the other hand, usual audio effects such as filters have undergone tremendous development in terms of design, in order to achieve the ability to control the frequency range and the amplitude gain, while taking care to limit the phase modulation. Also, some usual audio effects considered as simple processing actually require complex processing. For instance, reverberation systems are usually considered as simple audio effects because they were originally developed using simple operations with delay lines, even though they apply complex sound transformations. For all those reasons, one may consider that the terms "audio effects," "sound transformations" and "musical sound processing" are all refering to the same idea, which is to apply signal processing techniques to sounds in order to modify how they will be perceived, or in other words, to transform a sound into another sound with a perceptually different quality. While the different terms are often used interchangeably, we use "audio effects" throughout the book for the sake of consistency.

1.2 Classifications of DAFX

Digital audio effects are mainly used by composers, performers and sound engineers, but they are generally described from the standpoint of the DSP engineers who designed them. Therefore, their classification and documentation, both in software documentation and textbooks, rely on the underlying techniques and technologies. If we observe what happens in different communities, there exist other classification schemes that are commonly used. These include signal processing classification [Orf96, PPPR96, Roa96, Moo90, Zöl02], control type classification [VWD06], perceptual classification [ABL$^+$03], and sound and music computing classification [CPR95], among others. Taking a closer look in order to compare these classifications, we observe strong differences. The reason is that each classification has been introduced in order to best meet the needs of a specific audience; it then relies on a series of features. Logically, such features are relevant for a given community, but may be meaningless or obscure for a different community. For instance, signal-processing techniques are rarely presented according to the perceptual features that are modified, but rather according to acoustical dimensions. Conversely, composers usually rely on perceptual or cognitive features rather than acoustical dimensions, and even less on signal-processing aspects.

An interdisciplinary approach to audio effect classification [VGT06] aims at facilitating the communication between researchers and creators that are working on or with audio effects.[3] Various

[3] e.g. DSP programmers, sound engineers, sound designers, electroacoustic music composers, performers using augmented or extended acoustic instruments or digital instruments, musicologists.

disciplines are then concerned: from acoustics and electrical engineering to psychoacoustics, music cognition and psycholinguistics. The next subsections present the various standpoints on digital audio effects through a description of the communication chain in music. From this viewpoint, three discipline-specific classifications are described: based on underlying techniques, control signals and perceptual attributes, then allowing the introduction of interdisciplinary classifications linking the different layers of domain-specific descriptors. It should be pointed out that the presented classifications are not classifications *stricto sensu*, since they are neither exhaustive nor mutually exclusive: one effect can be belong to more than one class, depending on other parameters such as the control type, the artefacts produced, the techniques used, etc.

Communication chain in music

Despite the variety of needs and standpoints, the technological terminology is predominantly employed by the actual users of audio effects: composers and performers. This technological classification might be the most rigorous and systematic one, but it unfortunately only refers to the techniques used, while ignoring our perception of the resulting audio effects, which seems more relevant in a musical context.

We consider the communication chain in music that essentially produces musical sounds [Rab, HMM04]. Such an application of the communication-chain concept to music has been adapted from linguistics and semiology [Nat75], based on Molino's work [Mol75]. This adaptation in a tripartite semiological scheme distinguishes three levels of musical communication between a composer (producer) and a listener (receiver) through a physical, neutral trace such as a sound. As depicted in Figure 1.2, we apply this scheme to a complete chain in order to investigate all possible standpoints on audio effects. In doing so, we include all actors intervening in the various processes of the conception, creation and perception of music, who are instrument-makers, composers, performers and listeners. The *poietic level* concerns the conception and creation of a *musical message* to which instrument-makers, composers and performers participate in different ways and at different stages. The *neutral level* is that of the physical "trace" (instruments, sounds or scores). The *aesthetic level* corresponds to the perception and reception of the *musical message* by a listener. In the case of audio effects, the instrument-maker is the signal-processing engineer who designs the effect and the performer is the user of the effect (musician, sound engineer). In the context of home studios and specific musical genres (such as mixed music creation), composers, performers and instrument-makers (music technologists) are usually distinct individuals who need to efficiently communicate with one another. But all actors in the chain are also listeners who can share descriptions of what they hear and how they interpret it. Therefore we will consider the perceptual and cognitive standpoints as the entrance point to the proposed interdisciplinary network of the various domain-specific classifications. We also consider the specific case of the home studio where a performer may also be his very own sound engineer, designs or sets his processing chain, and performs the mastering. Similarly, electroacoustic music composers often combine such tasks with additional programming and performance skills. They conceive their own processing system, control and perform on their instruments. Although all production tasks are performed by a single multidisciplinary artist in these two cases, a transverse classification is still helpful to achieve a

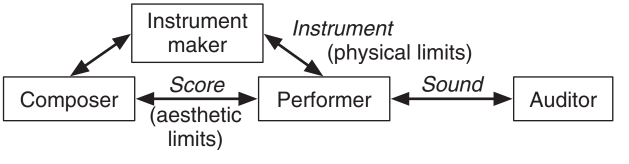

Figure 1.2 Communication chain in music: the composer, performer and instrument maker are also listeners, but in a different context than the auditor.

better awareness of the relations, between the different description levels of an audio effect, from technical to perceptual standpoints.

1.2.1 Classification based on underlying techniques

Using the standpoint of the "instrument-maker" (DSP engineer or software engineer), this first classification focuses on the underlying techniques that are used in order to implement the audio effects. Many digital implementations of audio effects are in fact emulations of their analog ancestors. Similarly, some analog audio effects implemented with one technique were emulating audio effects that already existed with another analog technique. Of course, at some point analog and/or digital techniques were also creatively used so as to provide new effects. We can distinguish the following analog technologies, in chronological order:

- Mechanics/acoustics (e.g., musical instruments and effects due to room acoustics)
- Electromechanics (e.g., using vinyls)
- Electromagnetics (e.g., flanging and time-scaling with magnetic tapes)
- Electronics (e.g., filters, vocoder, ring modulators).

With mechanical means, such as designing or choosing a specific room for its acoustical properties, music was modified and shaped to the wills of composers and performers. With electromechanical means, vinyls could be used to time-scale and pitch-shift a sound by changing disk rotation speed.[4] With electromagnetic means, flanging was originally obtained when pressing the thumb on the flange of a magnetophone wheel[5] and is now emulated with digital comb filters with varying delays. Another example of electromagnetic means is the time-scaling effect without pitch-shifting (i.e., with "not-too-bad" timbre preservation) performed by the composer and engineer Pierre Schaeffer back in the early 1950s. Electronic means include ring modulation, which refers to the multiplication of two signals and borrows its name from the analog ring-shaped circuit of diodes originally used to implement this effect.

Digital effects emulating acoustical or perceptual properties of electromechanic, electric or electronic effects include filtering, the wah-wah effect,[6] the vocoder effect, reverberation, echo and the Leslie effect. More recently, electronic and digital sound processing and synthesis allowed for the creation of new unprecedented effects, such as robotization, spectral panoramization, prosody change by adaptive time-scaling and pitch-shifting, and so on. Of course, the boundaries between imitation and creative use of technology is not clear cut. The vocoding effect, for example, was first developed to encode voice by controlling the spectral envelope with a filter bank, but was later used for musical purposes, specifically to add a vocalic aspect to a musical sound. A digital synthesis counterpart results from a creative use (LPC, phase vocoder) of a system allowing for the imitation of acoustical properties. Digital audio effects can be organized on the basis of implementation techniques, as it is proposed in this book:

- Filters and delays (resampling)
- Modulators and demodulators

[4] Such practice was usual in the first cinemas with sound, where the person in charge of the projection was synchronizing the sound to the image, as explained with a lot of humor by the awarded filmmaker Peter Brook in his autobiography: Threads of Time: Recollections, 1998.

[5] It is considered that flanging was first performed by George Martin and the Beatles, when John Lennon was asking for a technical way to replace dubbing.

[6] It seems that the term wah-wah was first coined by Miles Davis in the 1950s to describe how he manipulated sound with his trumpet's mute.

- Non-linear processing
- Spatial effects
- Time-segment processing
- Time-frequency processing
- Source-filter processing
- Adaptive effects processing
- Spectral processing
- Time and frequency warping
- Virtual analog effects
- Automatic mixing
- Source separation.

Another classification of digital audio effects is based on the domain where the signal processing is applied (namely time, frequency and time-frequency), together with the indication whether the processing is performed sample-by-sample or block-by-block:

- Time domain:
  - block processing using overlap-add (OLA) techniques (e.g., basic OLA, synchronized OLA, pitch synchronized OLA)
  - sample processing (filters, using delay lines, gain, non-linear processing, resampling and interpolation)
- Frequency domain (with block processing):
  - frequency-domain synthesis with inverse Fourier transform (e.g., phase vocoder with or without phase unwrapping)
  - time-domain synthesis (using oscillator bank)
- Time and frequency domain (e.g., phase vocoder plus LPC).

The advantage of such kinds of classification based on the underlying techniques is that the software developer can easily see the technical and implementation similarities of various effects, thus simplifying both the understanding and the implementation of multi-effect systems, which is depicted in the diagram in Figure 1.3. It also provides a good overview of technical domains and signal-processing techniques involved in effects. However, several audio effects appear in two places in the diagram (illustrating once again how these diagrams are not real classifications), belonging to more than a single class, because they can be performed with techniques from various domains. For instance, time-scaling can be performed with time-segment processing as well as with time-frequency processing. One step further, adaptive time-scaling with time-synchronization [VZA06] can be performed with SOLA using either block-by-block or time-domain processing, but also with the phase vocoder using a block-by-block frequency-domain analysis with IFFT synthesis.

Depending on the user expertise (DSP programmer, electroacoustic composer), this classification may not be the easiest to understand, even more since this type of classification does not explicitly handle perceptual features, which are the common vocabulary of all listeners. Another reason for introducing the perceptual attributes of sound in a classification is that when users can choose between various implementations of an effect, they also make their choice depending on